Informative Web Content Guidelines: A practitioner model for online content effectiveness

Jessica Muirhead

Submitted in partial fulfilment of the requirements
of the Degree of Doctor of Philosophy, March 2022

SALFORD
BUSINESS
SCHOOL

Table of contents

List of tables

List of figures

Acknowledgements

Thank you to my supervisors Dr Maria Kutar and Dr Gordon Fletcher for their inspiration, assistance, and invaluable feedback throughout my PhD candidature. Furthermore, I would like to thank the Digital Business team at Salford including Dr David Kreps, Prof Andrew Basden, Dr Yun Chen, and Dr Marie Griffiths for their reassurance with the research project.

Thanks also to the participants and organisations who provided their feedback during the studies. In addition, I appreciate the support from my colleagues at both Wrexham Glyndŵr University and CEG Digital during my doctoral studies.

I am also grateful to my friend Valerie for the ongoing encouragement throughout my doctoral studies. Thank you also to my feline buddies Boo and Cleo for their beautiful presence, even though they sadly didn't make it to the end of the PhD. Last but definitely not least, I appreciate the endless companionship of my little canine friend Mickey whose paw prints are on every page of this thesis.

Abbreviations

The following abbreviations are used in this thesis:

AR	Action Research
CMS	Content Management System
CTA	Concurrent Think Aloud
DQ	Data Quality
HCI	Human-Computer Interaction
IA	Information Architecture
IR	Information Retrieval
IQ	Information Quality
IWCG	Informative Web Content Guidelines *(output of this thesis)*
RTA	Retrospective Think Aloud
TAM	Technology Acceptance Model
UGC	User-Generated Content
UX	User eXperience
WCAG	Web Content Accessibility Guidelines

Abstract

With the rise of the world wide web, many organisations publish large knowledge bases as online informative content, enabling access for their current and potential stakeholders, customers, and service users. Providing universal access to information is a key feature of many national laws, ensuring that content is accessible for the intended audience, however there is little focus on its informativeness. Whilst there are many prior academic and industry frameworks for assessing the success of information systems, many of these focus on facets of the system itself or task completion, rather than the quality of the content. Evolutions of the WCAG (Web Content Accessibility Guidelines) have guided practitioners towards accessibility, neglecting the other attributes of information quality.

This interpretive study identifies the key attributes that have the greatest impact on information quality, using four action case studies to examine the attributes and identify areas for content improvement. Each action case study employs observations using task scenarios and the concurrent think aloud protocol to elicit user perceptions and cognitive understanding of information within websites and their inherent attributes of quality. The insights discovered from users feed the development of a model for practitioners to refine their content based on a synthesis between existing generalised literature and focused studies within the online space.

The Informative Web Content Guidelines (IWCG) is proposed as a new practitioner model for developing and assessing web content by promoting information quality. The guidelines parallel existing industry standards mandated by many national governments to improve online accessibility. Based on results from the case studies, this model combines key attributes from prior literature with three new attributes identified through the case

studies: those of fallback, information usability, and interactivity. By combining existing academic information quality frameworks with focused data from the case studies, a specialised selection of attributes for online information quality is proposed.

This thesis narrates the study, including the identification of potential information quality attributes from prior literature, the development of a practitioner-focused model based on WCAG principles and validation through a final action case study.

COVID-19 Impact Statement

During the period of doctoral studies, the COVID-19 pandemic and its associated implications applied additional constraints to the PhD research contained within this thesis.

Prior to the pandemic, all think aloud studies were conducted in-person, with the researcher observing the use of a case study website in the same room as the participant, asking clarifying questions and prompting as methodologically appropriate.

Following the requirements for national lockdowns and social distancing, the final think aloud sessions for the validation case study (see Chapter 12) were changed to use remote video conferencing tools (including Microsoft Teams). This adjustment remains methodically appropriate, with further reflection on the implications discussed on p. 187.

No further adjustments have been made to research activities in response to the pandemic.

Chapter 1. Introduction

As people and organisations rely more heavily on the Internet to find, retrieve, and understand information, it is important for practitioners to have tools and models to assess the *informativeness* of the text-based content they provide. Textual information is often authored for particular audiences and then published more generally without due consideration for its suitability or appropriateness for the audience beyond compliance with presentational and accessibility criteria such as the WCAG guidelines.

This study explores approaches for assessing web-based content using qualitative concurrent think aloud methods to produce a new set of guidelines for analysing and improving the informativeness of online content. The initial development of the Informative Web Content Guidelines (IWCG) combines information quality (IQ) attributes from existing academic literature with user task scenarios of three case studies: a university website, a non-profit organisation, and a government e-health resource. Applying an action case approach allows the iterative development of key attribute sets, identifying the important aspects of informative web content. These are then combined during the model development stage into the practitioner instrument, and further validated through two additional action case studies.

1.1. Background

Since the first world wide web page was authored in 1991 by Sir Tim Berners-Lee, the web has become an informational resource for all. Organisations are expected to have a web presence to disseminate key content to their stakeholders, customers, and service users providing on-demand explanation of topics.

1.1.1. Growth of the web as an information source

The web has evolved far from the first web page to the easy-to-access, universal, and often commercialised space available today. Much of the approaches to the development of the web as an information source are due to the direction of the early founders and individuals who imagined a free and flexible environment accessible to everyone around the world. These early visionaries led the production of technologies that would become embedded into not just our electronic devices, also our everyday lives.

1.1.1.1. 'Vague, but exciting'

Even with the early Internet (or ARPANET) developed by the US military in the 1970s becoming available to academic institutions around the world in the late 1980s, many collaborations were hampered by information separation: data stored on one computer could not be accessed from another, often requiring bespoke programs for access and processing. Information management as a discipline often focused on physical systems or the storage and retrieval of individual records on large, expensive, and specialised business machines.

Recognising the need for sharing information between researchers, Tim Berners-Lee prepared a proposal for a hypertext system built to run on the Internet, providing an interactive and standardised interface for writing and reading custom documents (Berners-Lee, 1989). Berners-Lee's boss reviewed the proposal and commented that it was 'vague, but exciting' (World Wide Web Foundation, 2020), laying a foundation for a worldwide network that offered uncertain yet unending possibilities for improving communication across our everyday lives.

1.1.1.2. Importance of information

Around the same time as the birth of the web, the fields of *information science* and *information systems* were developing into distinct academic

fields. Buckland (1991, p. 351) recognises that *information* is an ambiguous term, with three key meanings explored within the literature:

1. **Information as a process**
   The use of information to inform some other activity through communication channels.

2. **Information as knowledge**
   Key facts about a subject that may increase certainty.

3. **Information as thing**
   Objects and collections of data that have been processed to carry meaning and can be treated discretely.

The final meaning recognises a shift towards information as a collection of *data* or *capta*, identified by Checkland & Holwell (1997) with the Soft Systems Methodology as a processed form carrying additional meaning (see Table 3.1 on p. 63). If we consider information as a step towards knowledge, there is a need for the content of web pages to transform raw facts into something that can be processed and understood by users to inform their understanding of a specific topic. Viewing the web as a series of connected discourses presenting information ready to inform our internal knowledge and decision-making capabilities enables the use of models and theories from several different academic fields to explain user behaviours.

1.1.1.3. Information systems

The field of information systems (IS) is an inter-disciplinary domain focused on the intersection between technical computer science and social science-based management studies (Boland & Hirschheim, 1985, p. vii). Due to the growth of computers in every area of study, the information systems field can be hard to separate from other domains (Markus, 1999, p. 176) with a lack of overarching theories to underpin research studies (Gregor, 2006, p. 612). Some of the theories available for this research include those related to

information systems success (e.g., TAM and DeLone-McLean) and information quality frameworks derived from the original work of Wang & Strong (1996).

1.1.1.4. Information retrieval

Much of the early information systems research closely followed the implementation of computing within large organisations, with the widespread study of quantitative retrieval processes alongside the development of transistor and microprocessor-based devices within the 1970s and 1980s (Ingwersen & Järvelin, 2006, p. 111). A shift to this approach became apparent in the 1990s with a growing focus on the use of interfaces by everyday users, including considerations of their cognitive tasks. This led to the renewed online focus on information retrieval with mechanised approaches to ensure access to information objects as demanded by the user (Baeza-Yates & Ribeiro-Neto, 1999). The focus of study often focused on the delivery of information rather than the end goal of the user, with a lack of consideration for the task being conducted.

Moving away from the idea of information retrieval for informativeness rather than the sake of access provides the opportunity to explore deeper understandings of user activity, identifying successful patterns for task completion rather than simply use of a website. This is reflected in the use of task scenarios for user research activities during this study.

1.1.1.5. Accessibility

During the mid-90s amongst the growth of the web, there was a recognition that some users could not access all content due to disabilities or differences in their abilities to use contemporaneous computer equipment. A growing community of information systems researchers recognised the need for simple guidelines to aid developers in producing accessible web pages, including Vanderheiden (1995) who produced the first such document providing guidance for creating simple HTML pages. The author collected

advice from various online strategies, suggesting alternative methods and future strategies for universal web design.

The W3C along with the US Government initiated the Web Accessibility Initiative (WAI) recognised in early 1997 the need to provide further authoritative advice to the growing number of web developers on how to build websites that were accessible to all (Dardailler, 2009).

The WAI's project evolved into standardised frameworks known as the Web Content Accessibility Guidelines (WCAG), still referred to today by many countries' legislation on universal web access.

1.1.1.6. Inclusive design

As an evolution of accessibility as an activity, there is a growing practitioner trend for *inclusive design*. This promotes the philosophy that systems should be designed for all as standard, providing improved interfaces for everyone not just those considered to have a disability. The practice began in 1989 prior to the launch of the web, with cross-disciplinary conferences held shortly after to increase the awareness of inclusive design within physical products (Clarkson & Coleman, 2015, p. 237).

Within the Human-Computer Interface (HCI) community, the principles of inclusive design developed into a notion of universal accessibility, promoting digital inclusion by removing unnecessary barriers for an emerging digital audience (Abascal & Nicolle, 2005, p. 490). This concept is further developed by Persson et al. (2015), synthesising inclusive design as an extension of accessibility despite the lack of standardised definition.

1.1.1.7. Summary

Over the past three decades, the web has grown from a specialised tool to an everyday necessity, challenging the communication of information from companies and governments alike. With users now expecting a particular level of content provided in an accessible manner, the study of inclusive

design of systems has never been more important. Whilst much of the research focuses on the design of the system itself (e.g., web browsers and elements of the page), there is also a need to identify the properties of the information contained with a website.

The next section discusses the concept of informative content, including prior academic theories of information quality and the link to positive user outcomes.

1.1.2. What is informative content?

When discussing web and information-related topics in either academic or practitioner spaces, terminology can often be complex referring to different notions and concepts using the same nomenclature. Identifying exactly what is meant by informative content requires an understanding of the underlying academic theory as explored in Chapters 2 and 3.

1.1.2.1. Content

Content is one of the most challenging terms to define within the literature. Many academic sources (see Table 2.1 on p. 37) identify the link between information and content, noting that the most important facet is the meaning conveyed by the message. This returns to Shannon's (1948) early work on the theory of communication, identifying that the author is a *transmitter* and the reader a *receiver*. Meaning can only be conveyed where the reader understands the message the author has created (Krippendorff, 1980 updated 2018).

Since the development of computers as electronic information retrieval devices, the definitions of content have diverged into two separate fields: a technical, with content being a collection of objects or documents (ISO 9241-151, 2008, p. 3; Rosenfeld et al., 2015, Chapter 2); and the experience of the user when interpreting the information (Barker, 2016, Chapter 1; Stallman, 1996).

Whilst the original definitions were only concerned with textual content, the nature of the modern web has led to interactive, multimedia-rich experiences that have evolved beyond reading written messages. The key differentiator for content is that it contains the *meaning being conveyed to the user* rather than being part of the system itself.

1.1.2.2. Informative

Building on the academic theories of the Technology Acceptance Model (TAM; Davis, 1989) and DeLone & McLean (1992), *informativeness* of a system can be a key contributor to positive user outcomes. These can include overall task achievement, reduction in task time and reduction in cognitive load (Castañeda et al., 2007). Improving informativeness is a key approach to improving the acceptance of an information system (Lederer et al., 2000).

1.1.2.3. Informative content

The study of informative content on the web has led to the creation of fields such as information retrieval and information seeking behaviour. Despite the focus in these areas, academic studies cannot rely on a single unified measure of user perceptions towards the content (Thielsch & Hirschfeld, 2019). Understanding the effectiveness of content authoring on the web would help improve communications from organisations to their audience, improving the utility of websites whilst reducing the cognitive effort required to understand topics.

1.1.2.4. Information Quality

Liu & Arnett (2000) established that the success of a website can depend on the quality of information it contains, initiating a wide range of research into the facets of effective systems. Many of the definitions of information quality (IQ) relate to fitness for purpose (Bovee et al., 2003; Wang & Strong, 1996, p. 6), implying that information success is dependent on initial data and the users' context. The former is fully controlled by the web author, whilst the latter can only be hypothesised rather than comprehended. This distinction is

reflected in the early IQ frameworks based on mechanical properties of data rather than non-technical issues (Madnick et al., 2009).

1.1.2.5. Need for 'good' quality information

Producing informative content is a necessity for many organisations publishing online, with a positive correlation between the improvement of informativeness and benefits to the user through their perceptions of the system (DeLone & McLean, 2003). Adding additional value and meeting needs of the audience can increase user perceptions of a system, demonstrating the importance of subjective measures over quantitative definitions (Mawhinney & Lederer, 1990).

Research from the information seeking behaviours community has demonstrated the value of providing users easy access to the materials they need since the 1960s, identifying typical usage patterns and satisfying the requirements of the audience through stepwise improvements to system design. The evolution of this activity to incorporate information quality in addition to system quality is a natural extension reflecting the depth of HCI research and a need to continue evolving online access to materials.

1.1.2.6. Summary

Within the academic literature, definitions of *content* and *informativeness* are varied between domains and traditions. Understanding the concepts of information quality within a web space will help to identify features of textual content that can be improved to increase access to online information. This research study transforms the academic theories and attributes of information quality into a practitioner framework accessible to content authors within industry.

The next section summarises the key academic theories related to online information quality.

1.1.3. Academic theory

Information systems theories related to online information quality are divided between several domains. This adds complexity to a complete synthesis of prior research as there is no clear research field. The key theories relate to information systems success, information quality and information seeking behaviours, each of which are outlined below.

1.1.3.1. Information systems success

The study of information systems often focuses on the selection of an appropriate academic theory for success. One of the main models is the Technology Acceptance Model (TAM), developed by (Davis, 1989). TAM focuses on the acceptance of a system by its users, with user behaviour governed by the *perceived ease of use* and *perceived usefulness*. This has been applied to many web systems, with information quality demonstrated as the highest predictor of usefulness (Lederer et al., 2000).

With its origins in the late 1980s, TAM was developed in the context of organisational systems and forced use, reflecting the lack of choice for users with early computer technology. Users were typically employed by the organisation, resulting in studies to improve individual outcomes rather than analyse the overall success of a system.

The DeLone-McLean model (1992, updated 2003) takes a different approach, identifying overall information system success as dependant on a combination of three aspects of quality: the system itself, the service, and the information contained within the system. As a measure of overall success, the model can be applied to individual elements of a system (Urbach et al., 2009, p. 321) and by practitioners (Rosemann & Vessey, 2005, 2008).

Under the model, systems are deemed 'successful' if they provide net benefits for users, based on usage behaviour and user satisfaction. This considers patterns where there is an element of choice between various

systems, reflecting the situation in the modern web. Information quality has been identified as a strong antecedent for positive user satisfaction and increased net benefits (Petter et al., 2008), expanding the success of a system.

1.1.3.2. Information quality frameworks

Many academic information quality frameworks are grounded in data quality theories such as the seminal paper by Wang & Strong (1996). These take a technical, quantitative approach to establishing quality attributes which can often be distilled to simplified metrics. Pipino et al. (2002) recognise these frameworks as *objective* and *positivistic* in nature, leading to quantitative methodologies. As an alternative, *subjective* frameworks are proposed.

Subjective information quality frameworks are based on user opinions through studies on perceptions via an interpretivistic paradigm (Caballero et al., 2007). These often incorporate the user satisfaction measures as considered by DeLone-McLean (1992) as important facets of the success of an information system.

Many information quality frameworks have been documented by researchers for generic and specific applications (see Chapter 4), with many based on the Plan-Do-Check-Act cycle of iterative development. This reflects content as a process rather than a single-use activity, where stepwise improvements can continue to increase user outcomes after the original copy has been published.

Following the work of Zmud (1978), information quality researchers define the various dimensions of a system in terms of *attributes*. The specific attributes chosen for analysis are often context dependent and are not always mutually exclusive, providing a challenging space for comparison by the researcher. The analysis and selection of salient attributes is described in Chapter 7.

1.1.3.3. Information seeking behaviours

Theories on information seeking behaviours often focus on the user as a *problem solver*, as demonstrated by Dervin & Nilan's (1986) sense-making approach, which identifies how users bridge the gap between their situation and desired outcome. This further demonstrates how information is meaningless without context, requiring the study of web content in-situ rather than as an isolated activity.

On the web the problem being solved is often unclear: there are many instances where a user may be browsing without understanding their question or have an unclear idea of their situation. Applying problem solving approaches to information quality would allow for individualised studies, though not necessarily the overall information system success demonstrated by models such as DeLone-McLean.

1.1.3.4. Summary

Whilst there is an abundance of information quality frameworks, these often suffer from key challenges such as fully positivistic approaches, domain-specific focus, or conflicting attributes. This study will provide an academic model for web information quality, identifying the key areas where practitioners can focus content improvement to increase user success within their websites.

1.1.4. Practitioner frameworks

The world of the web moves quickly, and this is reflected in the lack of formalised practitioner guidance for content development. The main exception to this is accessibility, where the Web Content Accessibility Guidelines (WCAG) have provided a structured approach to evaluating web pages for the past two decades.

1.1.4.1. WCAG

The Web Content Accessibility Guidelines (WCAG) have evolved from a checklist (1.0) to a full suite of evaluation outcomes (3.0), providing a formalised analysis for demonstrating and improving accessibility within web pages. This has helped to improve access for disabled people to online content, with several governments legislating a minimum standard of compliance for all public body websites (e.g., the UK's Public Sector Bodies (Accessibility) regulations introduced in 2018).

WCAG has been largely a success, although there are several caveats around its application. For example, the guidelines are applied to each individual page without accounting for the user's journey between a website (except a standardised flow such as an ecommerce checkout), and user activity is often considered as desired patterns of behaviour rather than considering real activity. The main challenge of the guidelines is in providing a standardised analysis using metrics, often insufficient with complex user environments (Sloan & Kelly, 2011).

1.1.4.2. Lack of alternative frameworks

The lack of alternative frameworks is due to a multi-faceted complex situation. The development of such guidelines often requires collaboration between many different web technology experts, with their employers focused on achieving a single common goal. Accessibility is clearly a laudable aim, as widening access to people with disabilities an organisational and legislative imperative. This has an impact on both the profitability and public perceptions of an organisation.

Beyond accessibility, there is no clear direction for the evolution of the web. Several different fields such as content authoring, content management, web development and information retrieval have emerged from practice rather than theory, with practitioners documenting existing industry processes instead of defining them.

The definition of additional frameworks for spaces such as information quality could still meet organisational objectives; for example, improving the quality of content would help users find information more quickly, increase perceptions of the organisation and decrease the reliance on alternative more costly forms of communication.

1.1.4.3. Summary

Whilst there is a general lack of practitioner frameworks and guidelines for industry practices, WCAG provides a useful model for the establishment of standardised and documented practices for web development. A single authoritative measure of content quality that is not based on metrics-based analysis would improve the information provided by organisations, delivering a device for practitioners to analyse their own content practices.

1.2. Problem

The study of information quality within online contexts has not been previously explored, with a need to combine existing IQ academic theory with an industry approach to content evaluation. This will result in a practitioner framework for application to other websites, providing the benefit of IQ attributes to additional problem spaces.

This section defines key elements of this study, including the research questions, aims, objectives, and scope.

1.2.1. Research questions

Defining research questions is the first step towards scoping the study, identifying the problems that need to be solved. Oates (2005, p. 165) notes that research questions in action research should be formulated to solve problems of people within the context grounded in concerns of everyday people. The questions below have been formulated with this consideration.

1.2.1.1. How can the information quality of online content be evaluated? (RQ1)

The first question focuses on the evaluation of content. As information quality of online web pages has been conducted by prior studies (see Kandari et al., 2011a), there is precedence for the field. Many of the information quality frameworks explored by researchers have been based on quantitative measures and metrics that approach the problem from a positivistic perspective.

With an interpretivistic paradigm informed by insights from the DeLone-McLean model that information systems success largely depends on user satisfaction, this question focuses on the development of an appropriate technique to evaluate qualitative insights for information quality. This area of study has not been explored within the prior literature.

1.2.1.2. How are content quality and user achievement of online information-based objectives related? (RQ2)

As noted above, the DeLone-McLean model associates user satisfaction with systems success as an overall correlation of perceptions. RQ2 develops the academic theory further by examining the relationship between content quality and user achievement of information seeking tasks, identifying how such attributes influence overall website informativeness.

1.2.1.3. How can practitioners apply the key attributes to improve informativeness of online content? (RQ3)

The final research question relates to the practitioner focus of this study. In addition to the contribution to academic theory, the model produced by this research will guide those working in industry to improve their own website content using a set of guidelines like the existing WCAG model. Ensuring that the results of this study are accessible to practitioners is an important outcome, and methods will need to be developed to ensure this is applied in an appropriate manner for the web context.

The research questions will underpin the study and are analysed in Section 14.2 to ensure that they are answered by this research.

1.2.2. Aims

The aims provide overarching purpose of the study through main goals that the research needs to achieve. For this study, there are two aims:

1.2.2.1. To identify which information quality attributes influence the informativeness of online content. (RA1)

The first aim is to identify which attributes (existing and previously unknown) impact on the informativeness of online content. This is an important aim for the study as application by practitioners requires an understanding of the possible attributes that may be inherent within web pages. Although there have been small-scale studies of information quality in domain-specific cases, there is a lack of prior evidence for a universal framework of web content attributes. Without knowledge of the key properties that information may contain, improving such properties would not be possible.

1.2.2.2. To provide a practitioner model for online information quality based on academic literature and user studies. (RA2)

The second aim is to provide a practitioner model for improving content. This is dependent on the attributes derived in RA1 and based on both existing literature and the user studies documented within this thesis. This requires evidence to understand the influence that individual information quality attributes may have when applied to online content.

This research began as a study to help practitioners develop the quality of online content within their websites by proposing a model for assessment and improvement. Many researchers have defined frameworks for information quality analysis, many of which are based on the works of Wang & Strong (1996). These have lacked an analysis of which attributes are important to the user that this study will provide.

1.2.3. Research objectives

Objectives define the 'researcher's clear sense of purpose and direction' (Saunders et al., 2003, p. 25), and have more detail than research questions or aims. These include explicit criteria to ensure the research has been effective: how will the outcomes be measured?

For this study, there are three research objectives, each of which are described below.

1.2.3.1. To identify information quality attributes from academic literature inherent in online content. (RO1)

The first objective focuses on existing studies of information quality attributes. Starting from prior evidence within the IS field allows for the identification of initial attributes that may be applicable to the online space, providing a base from which to explore online content.

1.2.3.2. To determine the key attributes that impact on the success of online content. (RO2)

The existing academic literature contains initial taxonomical considerations for information quality attributes; however, these are not contextualised for web-based content. Analysing and refining the set of attributes to focus on the ones most impacting information quality of online content will produce additional theoretical knowledge to inform the practitioner model in RO3.

Whilst the previous research objectives have focused on identifying the attributes of online content, these need to be evaluated to find those with the largest potential to have an impact on information quality. Determining these attributes will require the study of existing websites and an understanding of how users find information, leading to the documentation of those causing the largest positive or negative impact.

1.2.3.3. To create and validate a practitioner model to describe the relationship between key information quality attributes and online content. (RO3)

The final objective outlines the creation of a model to aid practitioner audiences in the analysis of online content. This will provide an accessible outline to aid in the determination of information quality attributes that are most likely to have the largest impact on online content based on contextual variables.

An additional purpose of the model is to provide validation using another case study. This demonstrates the practical applicability of the information quality taxonomy applied to an additional context.

1.2.4. Scope of this study

This study will produce a practitioner model focused on the informative content provided by websites. Although many websites deliver additional content, such as multimedia, e-commerce, marketing and applications, this thesis will focus on the informative content only. There are four case study websites researched during this study: those belonging to a university, a non-profit charity, a national health resource and an encyclopædic resource. These demonstrate the broad range of domains where information quality analysis can aid practitioners in the improvement of website content.

1.2.5. Summary

Defining appropriate research questions, aims and objectives is key to a successful study. The questions outlined above provide a framework for studying the problem space, including an incorporation of both existing academic theory and new contributions through research activity.

The methodological approach including philosophical paradigm, methodology and specific techniques are described in the next section.

1.3. Methodology

Discovering and analysing the important attributes of information quality online requires in-depth study working with participants to identify ways in which content can be made more effective.

The methodological choices made during this study are summarised below, with more detailed discussion later in Chapter 6.

1.3.1. Interpretivist paradigm

Interpretivist research centres of the meanings people assign to their own activity (Orlikowski & Baroudi, 1991), exploring the types of questions that may arise due to involvement in the system (Klein & Myers, 1999, p. 69). Often this requires study of actors within their own setting, such as whilst they are using a system within a realistic environment (Walsham, 1995, p. 376). Along with the increasing diversity of research within the Information Systems field, there has been a growth in the number of published studies applying an interpretive paradigm (Mingers, 2001, p. 240), reflecting a recognition that success of a system is not just due to technical issues; socio-technical considerations can play a large role in widespread acceptance.

An interpretive approach is particularly suitable for this study as the problem space is complex without concrete prior assumptions. Many of the information quality attributes explored in Chapter 4 require subjective interpretation, viewing the content through the lens of the user and their direct experiences of the system.

1.3.2. Soft Systems Methodology

Checkland's (1981) Soft Systems Methodology (SSM) is an approach for researchers to explore complex systems that may involve many components, actors, and organisational constraints. All systems problems can be simplified to three key elements: a framework of ideas, a methodology for evaluating them and an area of concern for study (see Figure 6.3 on p. 156). As the

framework of ideas is improved with every application of the methodology, Checkland's SSM leads to the concept of iterative action research, where the researcher has a role within the activity as an observer, and results can be refined over multiple applications of the methodology. The researcher reflects on each application, informing future stages of the study. To simplify the problem space, SSM assumes that all people involved in a system act purposefully (Checkland & Poulter, 2006).

1.3.3. Action research

Action research as a methodology was first used within the Education field in the 1940s, and has since grown to be applied within many social sciences contexts (Oates, 2005, pp. 154–155). Within Information Systems, action research began to be applied in the 1980s and 1990s with a focus on studies with an *involved researcher*, participating as well as observing in the activities (Walsham, 1995). The growing use of action research reflects a need for researchers to better align their work with industry practices (Baskerville & Myers, 2004).

Canonical action research is based on a five-stage cyclical approach: planning, intervention, evaluation, reflection, and diagnosis (Davison et al., 2004, p. 66). The same stages are repeated with multiple iterations of activity, with the results of the prior stages informing later study. This leads the researcher to refine and scope research questions during the study, an important consideration given that many interpretive studies begin without a clear singular hypothesis or prior assumptions.

Action research is useful at producing various forms of knowledge, including additional insights beyond the propositional ideas from positivistic approaches. For example, *experiential* knowledge can be gained from interpreting user sentiments that can be challenging to vocalise; and *practical* knowledge through the researcher participating in the activity (Heron & Reason, 1997, pp. 280–281). These additional forms of knowledge

can gain further insights beyond quantitative analysis, enhancing deep understanding of the problem space.

Applying action research is closely linked to the aims of Checkland's Soft Systems Methodology, with a cyclical approach and self-reflection enabling deeper insights and understanding of socio-technical contexts.

1.3.4. Action case

Braa & Vidgen (1999) explored information systems research methods, identifying the three dynamics of *prediction*, *understanding*, and *change*. All three dynamics are apparent in all IS studies, however the selection of core methodology can adjust the balance between perspectives. By analysing existing research methods, the authors identified a previously unexplored space termed *action case* research, a theoretical combination of action research with soft (qualitative) case studies.

Case studies are frequently applied within the IS field when researchers are focusing on a single instance (Oates, 2005, p. 141) with no prior knowledge of the research context (Benbasat et al., 1987, p. 370). In particular, explanatory case studies are in-depth analyses as a descriptive study, answering *how* and *why* questions: a close match with the principles of interpretive research.

Since case studies often explore only a single isolated context, action case studies introduce the concept of multiple linked cases. Each of these cases may contain an alternative context with the same problem, with the insights from one case informing the next, more akin to action research. Performing cycles of action case studies requires the researcher to be an observer, refining research questions throughout the overall study (Simonsen, 2009, p. 113). This results in generalisations of concepts, theories, implications, and insights (Walsham, 1995), further contributing to the research discourse.

A cyclical action case approach is suitable for this study as informativeness needs to be generalised across multiple domains and contextual situations,

with the researcher observing use of the system. Each context can form an individualised case study, with the attributes discovered in each case informing future stages of research.

1.3.5. Task scenarios

Within each action case, the research needs to identify differing viewpoints, usage patterns and relationships (Carroll, 1999, pp. 5–6), leading to the use of task scenarios to simulate typical user activities.

Task scenarios are commonly used within the usability industry for information retrieval studies (Kim, 2012, p. 300) with realistic and actionable activities (Nielsen Norman Group, 2014). These facilitate natural search behaviours (Borlund & Ingwersen, 1997, p. 226), eliciting participants' thought processes producing additional knowledge beyond observation as an isolated activity.

1.3.6. Think aloud protocol

During a participant activity, the researcher requires a technique for eliciting and recording feedback. One such technique is the think aloud protocol, used widely within the fields of cognitive psychology, and HCI research (Blandford, 2019).

The think aloud protocol can be contextually applied to methodological situations, with activity centring on the participant completing a task whilst being observed by the researcher. In addition to watching the activity, the researcher may ask questions or prompts to elicit verbalisations and explanations of the participant's thought process (see p. 179). The researcher is required to deepen their understanding of the context whilst maintaining neutral instructions that do not bias or guide the participant to predetermined conclusions (Ericsson & Simon, 1984).

Within the HCI field, two different types of think aloud study have emerged: *concurrent* (in real-time as the participant conducts an activity) and *retrospective* (based on post-activity reflection). The former can lead to more

accurate findings with the researcher able to tailor questions to more deeply explore cognitive patterns whilst avoiding issues with remembering the activity (Altuntaç, 2015, p. 5). Concurrent think aloud is suitable for this study as it focuses on observable properties of content rather than the user process of understanding the content, whilst providing results with a small number of participants. This enables multiple rounds of action case study as achievable within the scope of this doctoral research.

1.3.7. Ethical approval

The researcher obtained full ethical approval prior to commencement of the action case studies conducted during this project. Documentation related to the ethical approval process is included in Appendix B.

1.3.8. Summary

The research context and questions lend themselves towards an interpretivistic approach, with *how* and *why* questions exploring deeper insights beyond known questions. The research questions themselves may evolve during the study as knowledge is gained from the research activities.

The concept of action case studies as individual contextual situations informed by prior research is appropriate as multiple websites and domains can be studied to produce generalisable results across multiple different cases. With the large variety of websites and content patterns on the modern web, any model produced for practitioners needs to be applicable to as many contexts as possible.

The think aloud protocol, commonplace in industry-based user research, is a technique that can be used to both observe activity and explore user cognitive patterns, discovering deeper insights into understanding of the website. This is important within the overall study as information quality attributes are often subjective, needing to be explored within context and with a model of the user's understanding of the content within.

The next section describes the researcher's motivation to study this topic.

1.4. Researcher

The researcher's perspectives and values can have an impact on the approach to the research study. This section contains a description of the researcher's motivation and values, followed by an overview of the research journey.

1.4.1. Motivation and values

The idea for this research follows from the researcher's decade-long experience working within UK digital agencies. Through planning, designing, developing, and maintaining dozens of online sites and systems for a diverse range of clients in a variety of industries, the same challenges kept reoccurring. The improvement of online content is often an unknown area to practitioners, who follow guidelines and best practices, yet are unsure how to assess the actual performance with real users.

Within the literature on information systems, there are many research-based tools and techniques in use, however none of these are suitable for the researcher's purposes, as they:

1. Are based on academic theory and often not presented in industry-compatible formats. This leads to practitioners who may not have a full understanding of the underlying theories that may help to improve their content authoring activities.

2. Utilise fully quantitative methods without considering the qualitative aspects of user experience and information seeking behaviours. This is contrary to the researcher's epistemological approach that it is not possible to fully encapsulate the perceptions and behaviour of users in a numerical manner without considering other *softer* aspects of information systems use.

This research has been conducted to supplement the existing methods provided within the Information Systems field by providing a practitioner model to explore information quality attributes within online content.

1.4.2. Research journey

The development of this thesis spanned 6 years of part time candidature, from 2016–2022. During this time, the focus of the research including objectives, questions, and aims were refined to produce the final study documented in the thesis.

The diagram in Figure 1.1 below identifies the key stages of the research journey.

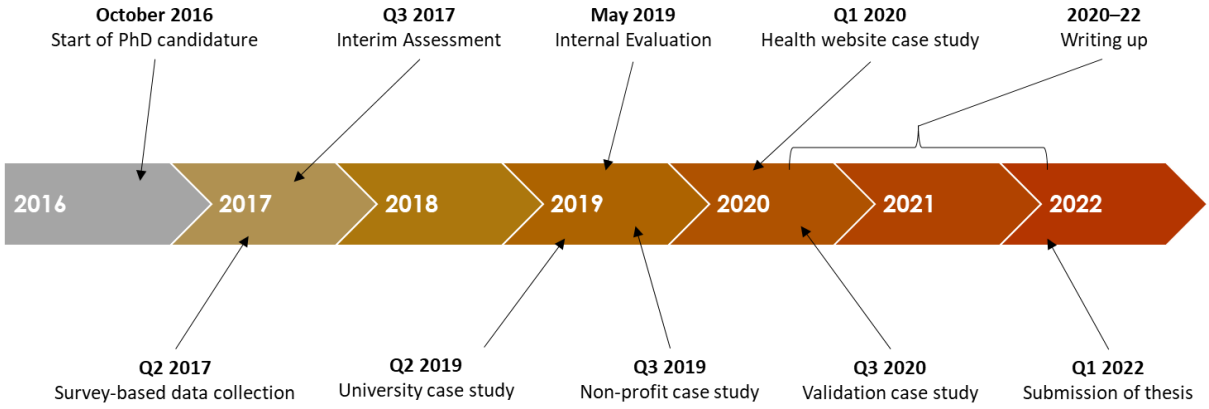


Figure 1.1: Research timeline identifying key stages in this study.

The research study began in October 2016 with the commencement of PhD candidature. Preliminary research directions focused on the use of user centred techniques to improve the informativeness of online content, including mixed methods such as a synthesis between user surveys (often used to collect feedback during website use) and Google Analytics tracking data. To identify the key facets of such an approach, the researcher presented an outline of the research questions at the Psychology of Programming Interest Group (PPIG) Doctoral Consortium, then designed and deployed an online survey to identify the extent to which digital agencies implement user research to study web content. The researcher presented

doctoral consortium submissions at the UK Academy for Information Systems (UKAIS) conference in 2016 and 2017, developing research perspectives within the Information Systems community.

As an evolution of the research and an outcome of the Interim Assessment in 2017, the focus changed to a purely interpretive approach identifying the key attributes of online content that impact on its informativeness due to a need to appropriately scope the study. This led to the development of the questions, objectives and aims as documented earlier in this chapter, and the identification of the action case study approach as particularly suitable for the nature of the planned research.

The researcher presented progress and direction of the project including preliminary data at the Internal Evaluation in May 2019. With the feedback from the examiners, the research was directed towards think aloud studies, eliciting responses from users during a task scenario. This follows standard user research practice, scoping the study to focus on several scenarios (as action case cycles) and providing insights beyond the verbalizations of participants.

This methodology led to doctoral consortium presentation at the BCS 2018 conference in Belfast, helping to develop the direction of the research within the Computer Science and Human-Computer Interaction (HCI) communities.

The researcher presented progress with the first case study at the ACM CHIIR 2019 conference in Glasgow, demonstrating the impact of qualitative task scenarios on the analysis of website informativeness. This provided a useful link with the online information seeking and behaviour communities with a heavier focus on quantitative methodologies.

During analysis of the case studies, the researcher identified a trend for content authors to misidentify key informative attributes within their websites. This demonstrated the need for a clear practitioner model, with the WCAG being the prevalent guidance within the industry for accessibility. With the

close link between accessibility and informativeness (explored further in Section 7.4.1 on p. 210), this is a natural approach for producing industry-focused guidelines.

Following the development of the model based on the first three case studies, the researcher conducted two further studies on a validation case study website both before and after the changes suggested by application of the guidance. This demonstrates the improvement to informativeness available through application of the research conducted during this study.

The practitioner model and information quality attributes examined during this study provide the potential for future research directions and outputs, as discussed in further detail in Section 13.6.

Throughout the period of PhD candidature, the researcher has held several academic roles including Graduate Teaching Assistant and Associate Lecturer, Digital Business, University of Salford; Lecturer and Programme Leader in Computer Science, Wrexham Glyndŵr University; and Head of Academic Practice at CEG Digital, an OPM (Online Programme Management) organisation delivering online degree courses with 9 UK partner universities. In addition to these roles, the researcher has been a Research Assistant on the Understanding Digital Events project (British Academy-funded); was awarded funds as the Principal Investigator for research projects identifying approaches for online age verification technologies (Alcohol Change UK and the Institute of Alcohol Studies); led a project to synthesise information provided within university Virtual Learning Environments (VLEs); was faculty representative on the Wrexham Glyndŵr University Research Ethics Sub-Committee; and was the student representative for the UK Academy for Information Systems (UKAIS).

1.4.3. Summary

Building on the researcher's prior industry experience, this study combines an interest in the study of informative online content with information systems

theory. The model produced by the research activities will be applicable to the researcher's professional roles, enabling the improvement of online content within educational environments. In addition, the contributions to theory discussed in the next section develop towards the researcher's ongoing interest in the interaction of users with online systems and the ways in which information can be provided in an easily accessible and understandable manner.

More details of the researcher's achievements during the period of doctoral studies are included in Appendix E on p. 473.

1.5. Contributions

Within this doctoral research, there are contributions to all three areas: theory, knowledge, and practice (see Figure 1.2). This section describes the contributions made by the study in each of these areas.

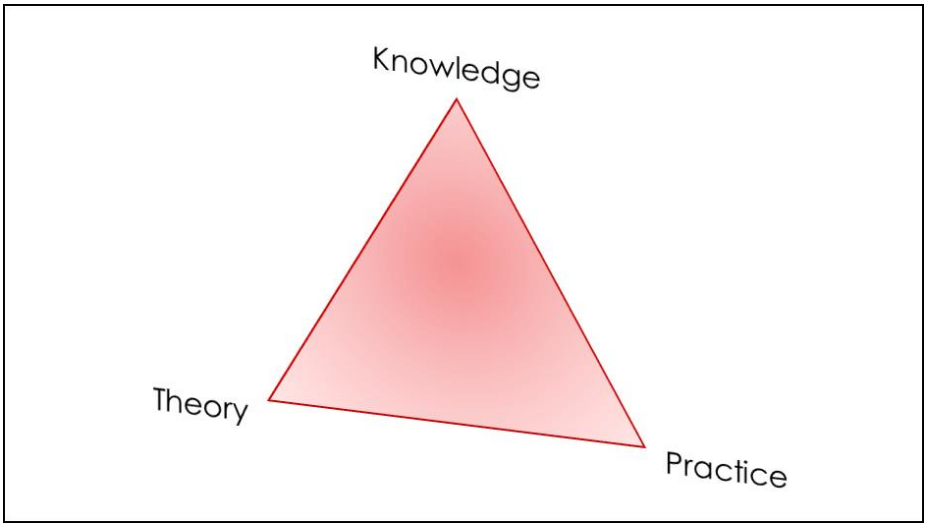


Figure 1.2: Three categories of contribution provided by research.

1.5.1. Contributions to theory

This research builds upon existing information systems theory, such as that provided by the DeLone-McLean model (1992 updated 2003). Information quality has been identified by Petter, DeLone, & McLean (2008) as one of the strongest influences on user satisfaction and net benefits of information

systems, however only a few studies have focused on applying the model within an online space (Schaupp et al., 2006). Applying the model to websites demonstrates the applicability to this category of information system whilst moving beyond the core attributes identified by Wang & Strong (1996) and used throughout the literature without consideration of their appropriateness to the context. This leads to the identification of the key attributes of quality in online content, resulting in a domain-specific taxonomy and model for evaluating the set of criteria for informative websites.

1.5.2. Contributions to knowledge

Prior works have focused on evaluating a shortlist of attributes inherent in content, however, there has been little focus on the appropriateness and viability of using these attributes within a web-based realm. By identifying the properties of online content through user-focused studies, this research contributes to the discourse with new and prioritised attributes that can be applied within future works.

1.5.3. Contributions to practice

This research leads to the identification of a set of information quality attributes that can be used to assess website content, providing a model for industry practitioners to further apply to their own websites. Linking the information quality attributes to the Informative Web Content Guidelines (IWCG) model following the WCAG framework allows for implementation by practitioners in accordance with existing web content authoring processes. A standardised approach ensures that the model as a contribution to practice is relevant and accessible within the industry, furthering the application of the research beyond this doctoral study.

1.5.4. Summary

This section has outlined several contributions expected to be generated by this research study. Table 1.1 below identifies each of the six intended contributions in this thesis mapped to research objectives and type.

#	Contribution	Research Objectives	Contribution type
C1	Identification of existing information quality attribute set relevant to online content	RO1, RO2	Theory
C2	Identification of new information quality attributes for online content	RO1, RO2	Knowledge
C3	Development of enhanced attribute set relevant to online content	RO1, RO2	Knowledge
C4	Methods for analysing online information quality	RO3	Practice
C5	Practitioner model for application of attributes to website content	RO3	Practice
C6	Guidance for practitioners to improve online content quality	RO3	Practice

Table 1.1: Research contributions incorporated into this study.

Contributions to knowledge, theory and practice are reviewed in Section 14.6 within the Conclusions chapter at the end of this thesis.

1.6. Organisation of the thesis

The organisation of this thesis follows the model described by Oates (2005, p. 12), who states that any research project in information systems to be composed of six aspects (see Figure 1.3). These aspects all need to be considered for a successful research project. Each element of the model is annotated with the corresponding chapters in this document.

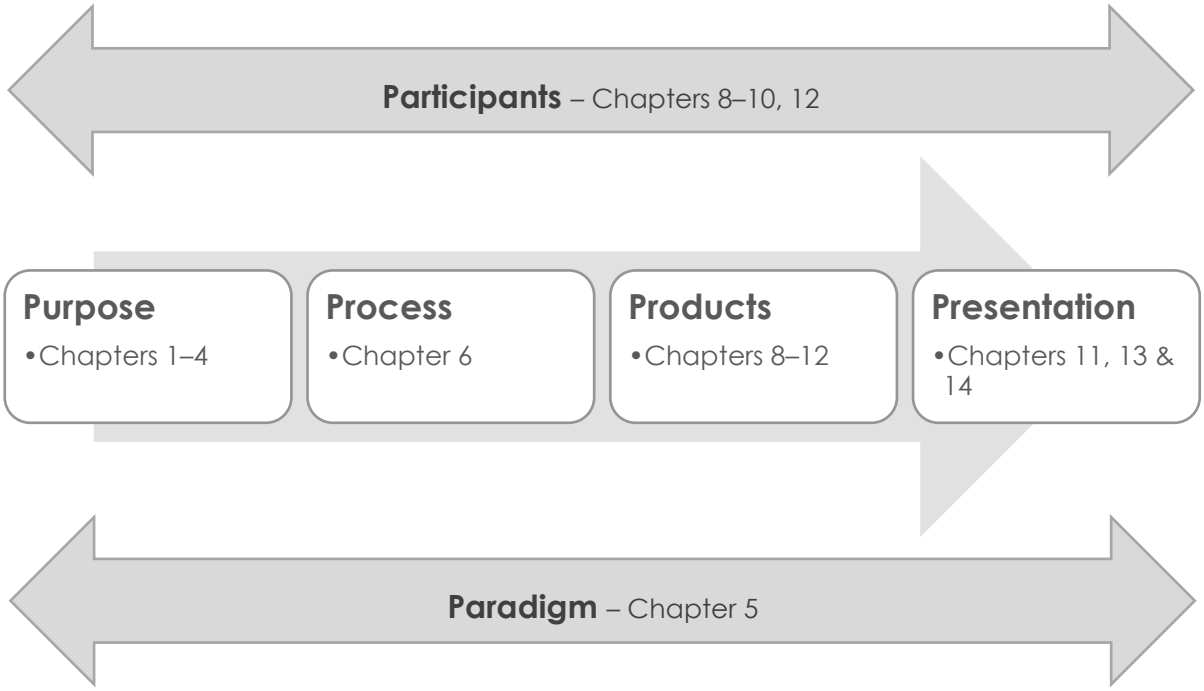


Figure 1.3: The 6 Ps of research (Oates 2005, p.12) mapped onto thesis structure.

The first four chapters of the thesis outline the *purpose* of the research, through introduction and the literature review, identifying current academic theories and industry practices. This provides background context to the research challenges.

Chapter 5 discusses research philosophy, including identification of the interpretivist *paradigm* as suitable for the research questions.

The methodology is described in Chapter 6, with a narrative of the selection of a qualitative approach, soft systems methodology and action case study

research. This forms the *process*, and is further developed into the research plan, incorporating task scenarios and the think aloud protocol as suitable user research methods. The overall study plan is described in Section 6.9 on p. 193.

Chapter 7 describes the attribute selection process, where existing information quality frameworks are analysed to identify the key attributes that may have an impact on online content quality.

Chapters 8–10 contain individual write-ups of the three action case studies to identify the informative attributes of content within three case study websites. Each of the chapters is an independent narrative informed by the findings from earlier case studies. These chapters form both the *product* and *participants* of the research.

Chapter 11 discusses the formation of the model identifying the taxonomy of guidelines and structure of guidance for professionals, resulting in the Informative Web Content Guidelines as shown in Appendix A on p. 425, *presenting* the *products* of the research study.

Following the development of the model, Chapter 12 provides a narrative for the two validation case studies on a further website, demonstrating the effectiveness of the practitioner guidelines. This further produces *products* of the research involving *participants*.

The final two Chapters 13 & 14 *present* the research study, identifying key themes for further exploration and research.

1.6.1. Chapters

This thesis is organised into 14 chapters, each of which is summarised below.

Chapter 1. Introduction ..... p. 1
This introductory chapter, with an overview of the research project, questions, and the thesis.

Chapter 2. Online content ..... p. 35
A literature review of online content, including definitions, relevant frameworks and an overview of industry standards, guidelines, and best practice.

Chapter 3. Information quality ..... p. 61
Literature review on information quality in the fields of Information Systems and Human-Computer Interaction, positioning the relevance of content quality within online systems.

Chapter 4. Information quality frameworks ..... p. 94
Literature review of general and domain-specific IQ frameworks identifying common patterns of definition and their contexts of application.

Chapter 5. Research philosophy ..... p. 127
Philosophical positioning of the research project, including researcher's perspectives on information systems theory.

Chapter 6. Methodology ..... p. 145
Outline methodology for conducting the research project, including an overview, selection, and justification of the relevant methods.

Chapter 7. IQ attribute selection study ..... p. 196
Description of the attribute selection study where the IQ attributes identified within the literature are reduced into the core set that have an impact on online content for future study.

Chapter 8. Case study 1 – University website ..... p. 222
Narrative on the first action case study, the University website including questions, think aloud study, and analysis of user feedback.

Chapter 9. Case study 2 – Non-profit website ..... p. 245
Descriptive narrative of the second action case study, the non-profit website including task scenario and analysis of participant responses.

Chapter 10. Case study 3 – eHealth website ..... **p. 263**
Narrative of the third action case study focusing on the eHealth website, including task scenario and participant responses.

Chapter 11. Model ..... **p. 281**
This chapter proposes the practitioner model for improving information quality of online content.

Chapter 12. Validation study ..... **p. 301**
Description of the validation study conducted to trial the proposed model with an additional fourth case study.

Chapter 13. Discussion ..... **p. 337**
Narrative on the study's progress towards research questions and contributions to theory, knowledge, and practice.

Chapter 14. Conclusions ..... **p. 366**
End-of-project conclusions incorporating contributions, outputs, and future areas of study.

1.6.2. Appendices

Each of the appendices included at the end of this thesis are summarised below.

Appendix A. Informative Web Content Guidelines (IWCG) ..... **p. 425**
Documentation of the Informative Web Content Guidelines (IWCG) developed during this research study in a practitioner-accessible format, aligned with the WCAG standards.

Appendix B. Ethical Approval documentation ..... **p. 452**
Copies of documentation relating to the ethical approval of this research study, including approval letters and organisational/individual consent information.

Appendix C. Data analysis worked example ..... **p. 458**
A worked example of the data analysis process as described in Section 6.7.10 and conducted during the first action case study cycle in Chapter 8.

Appendix D. Application of model ..... **p. 464**
A summary of modifications made to the case study website during the application of the model in Chapter 12.

Appendix E. Researcher achievements ..... **p. 473**
An overview of the researcher's key achievements during the period of doctoral studies.

1.7. Summary

This chapter has introduced the research problem, questions, and objectives in preparation for the study. The research questions identified during this chapter will be referred to throughout the study to ensure that all activities provide answers to the relevant problems. In Chapter 14, the project aim and questions are reviewed to demonstrate that appropriate outcomes have been achieved.

The following three chapters review relevant literature from both practitioner and Information Systems perspectives to identify online content and information quality definitions, concepts, and frameworks to aid the objectives of this research. This culminates in the analysis of existing information quality attributes, leading to the selection of those pertinent to the research questions.

Chapter 2. Online content

This chapter focuses on online content, synthesising the various academic and practitioner definitions, guidelines, and industry practices to develop a review of current research.

2.1. Introduction

In 2020, more than 96% of UK households have access to the web through a PC, mobile phone or tablet (Office of National Statistics, 2020). With almost ubiquitous Internet access, the primary purpose of the World Wide Web is 'providing immediate access to content' (Thielsch & Hirschfeld, 2019) alongside secondary purposes such as communications, entertainment, e-commerce and work. Most websites contain some form of online content designed to be *informative*, providing access to key content of interest to a wide range of users.

The origins of information theory can be traced to Shannon's (1948) seminal paper on the theory of communication. Within the mathematical model, information is sent from a transmitted to a receiver, with the possibility of noise or disturbance in the process. Content could be thought of as the information being transmitted, or the *message*: this possibility was considered by Shannon as 'a sequence of letters', what we would today call *textual content*.

As computer systems evolved, the idea of content and its usefulness has shifted: in their early work, McLuhan (1994, p. 7) suggested that the 'medium is the message', i.e. that more information can be drawn from the communications format and appearance than the content itself. This highlights a continual dichotomy within Human-Computer Interaction research: is the content or the way the content is transmitted more important

to the user? Does the system's appearance, navigation style and usability have an impact on the way users find information?

2.2. Definitions of content

Within the Information Systems literature, there is no commonly accepted definition of *content*. Many authors work towards their own definitions; however, these are frequently not explicitly stated within works relying on the cultural norms within a narrow field.

In order to understand what content is and how the term can be applied within this research, it is necessary to examine its use within several related fields across both academia and industry. A summary of these definitions is shown in Table 2.1.

Author	Definition of content	Field	Focus
OED (2021)	'The things contained or treated of in a writing or document; the various subdivisions of its subject matter'	General	General
Krippendorff (1980 updated 2018, p. xviii)	Meanings contained with a message	Content analysis	Academia
McCandless (1996, p. 8)	'any information that can be communicated electronically'	Engineering	Academia
Stallman (1996)	'Some sort of information'	Software development	Industry
ISO 9241-151 (2008, p. 3)	A 'set of content objects' which can be 'interactive or non-interactive object containing information represented by text, image, video, sound or other types of media'	Quality management	Industry
Rosenfeld et al. (2015, Ch. 2)	Content can 'include the documents, applications, services, schemas, and metadata that people need to use or find in your systems'	Information architecture	Industry
Barker (2016, Ch. 1)	Content is created and used to interpret raw data. It has a specific purpose and evolves over time.	Web content management	Industry
W3C (2018)	'Information and sensory experience to be communicated to the user by means of a user agent, including code or markup that defines the content's structure, presentation, and interactions'	Web accessibility	Industry

Table 2.1: Definitions of content within literature.

The table above shows that despite a general-purpose dictionary definition of *content*, the term is used with various meanings across different domains and the academic/practitioner boundary.

The term *content* is generally accepted to be of Latin etymological origin, with a link to the words *con-tenere*, meaning to *hold altogether* (OED, 2021). This derivation demonstrates a clear link to content being a collection of individual items considered together to form a coherent explanation of a subject.

Krippendorff (1980 updated 2018) focused on content within information systems, and defined this to be *meanings contained with a message*, building on Shannon's (1948) communications theory. This assigns an additional concept to content: that there may be more meaning derived from two pieces of content together than when they are perceived individually, with information being an additive operation (Gernert, 1996). This is analogous with psychology's Gestalt theory, where more meaning can be derived from the whole sum of the parts of a situation than from the individual parts together (Wertheimer, 1924 as translated and reprinted in Ellis, 2013).

Outside of Information Systems, the term *content* has had a varied usage pattern. According to Google's (2021) analysis of millions of literature items from the past four centuries as shown in Figure 2.1, the nomenclature suffered a decline in use until the early 1900s, when the word began to be used with more regularity to describe the burgeoning number of media items.

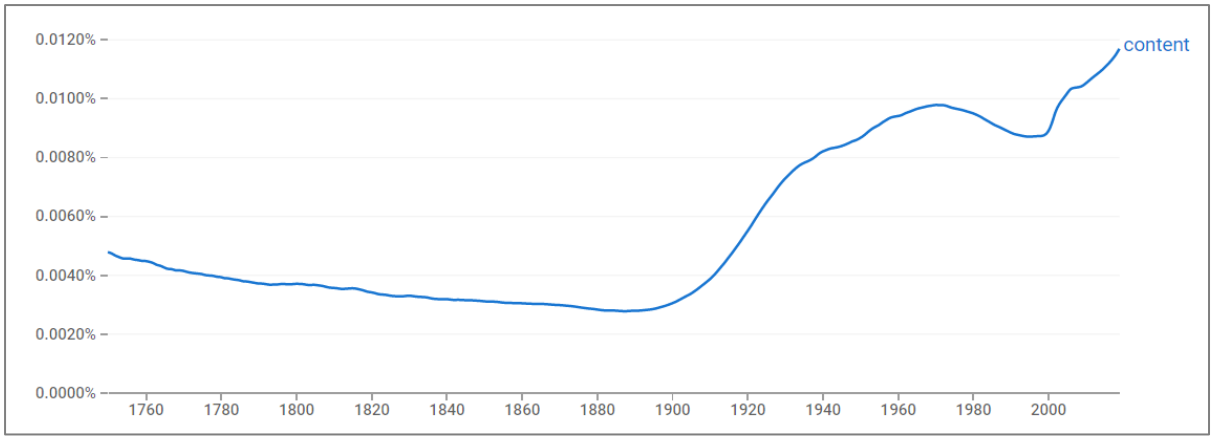


Figure 2.1: Google Books Ngram graph demonstrating the use of *content* within books and associated literature between 1750 and 2019 (Google, 2021).

The largest increase in the term's use dates from the late 1990's until present, reflecting the growth of the Internet and the information society. As we consume more online media, the term has been increasingly used to describe the items within a web page: textual content, images, videos, and other interactive devices. The web development world ubiquitously applies this definition with industry, with a loose understanding of the types of items that may be considered to be content on today's web.

Early web thought leaders such as McCandless (1996) and Stallman (1996) suggested a broad definition for content of any information stored or transmitted through the Internet. Whilst this maintains the generalised nomenclature, the lack of clarity from those involved in the growth of the online world is still reflected today in the variety of meanings used within industry.

The web of the early 2000's experienced a process of simplification and unification, leading to international standards such as ISO 9241-151 (2008), which returned to the etymological origins of the term to define content as a 'set of content objects' suggesting a contemporaneous taxonomy of such content object types to include text, images and media. This definition was not expected to be complete, reflecting the anticipated growth of future web technologies such as augmented and virtual reality.

From an Information Architecture perspective, authors such as Rosenfeld et al. (2015, Ch. 2) have expanded on the concept of a *content object*, including abstract ideas reflecting virtual technologies such as documents, schemas and applications. Whilst some of these such as documents can be clearly seen to contain content in an informative sense, the idea of code and designs as content reflects the way programs and data are often stored within the same memory space as first described by von Neumann (1945 reprinted in 1993).

Returning to the informative nature of content, Barker's (2016, Ch. 1) definition explores the link between *raw data* and *information* as described in the Information Systems field by Checkland & Poulter (2006, pp. 112–113; see p. 63). Considering the meaning of content as more than the raw data used in its construction returns to Krippendorff's (1980 updated 2018) pre-web definition of content connecting the user with meaning.

The concept of informativeness can inform the definition of content within an information system. More discussion on informativeness is included in Section 3.1.3.

The industry-standard Web Content Accessibility Guidelines (WCAG 2.1) defines *online content* to be, 'information and sensory experience to be communicated to the user by means of a user agent, including code or markup that defines the content's structure, presentation, and interactions' (W3C, 2018a). This broad definition encompasses all aspects of the user's experience of a web page, reflecting the comparative value of all elements (such as text, images, media, navigation, search) within the provision of informative content to communicate ideas in the online world.

As demonstrated in this review of content definitions, no single description can be given to provide a holistic view for all contexts and situations online. The specific use of the term *content* is often flexed to provide nuanced meanings within particular guidelines, standards, or industry practices.

For clarity within this thesis, online content is taken to mean any information that is expressly provided to the user to transmit meaning within the message, including text, images and media though excluding the technical means with which this is conveyed, for example systems architecture or underlying code.

2.3. Content is king

Bill Gates' (1996) seminal blog post *Content is King* proposed a future with an abundance of online content, drawing on experiences from prior new media such as the rise of TV broadcasting. The essay popularised the notion that the growth of the Internet would be driven by high quality content produced not just by large corporations, but also smaller organisations and individual users.

Within academia, there are many authors who agree with Gates' perspectives. Kang & Kim's (2006) study demonstrated that user satisfaction is often linked to the content provided within a website, with clear and accessible information providing positive experiences for users.

Since the original blog post, the development of the Internet has led to many opposing views on whether or not content remains king (Müller & Christandl, 2019, p. 46), though this leads to a key question: what determines the *effectiveness* of online content? This research examines content effectiveness from the viewpoint of *informativeness*: what makes informative content successful for users?

With the growing amount of information published online, many organisations have implemented content management systems (CMSs) to manage large amounts of text and image content. The graph in Figure 2.2 shows the rapid growth in CMSs over the past decade, rising from 24% of all websites in 2011 to 62% in 2021. This highlights the challenges in managing voluminous amounts of textual and multimedia data with the modern web.

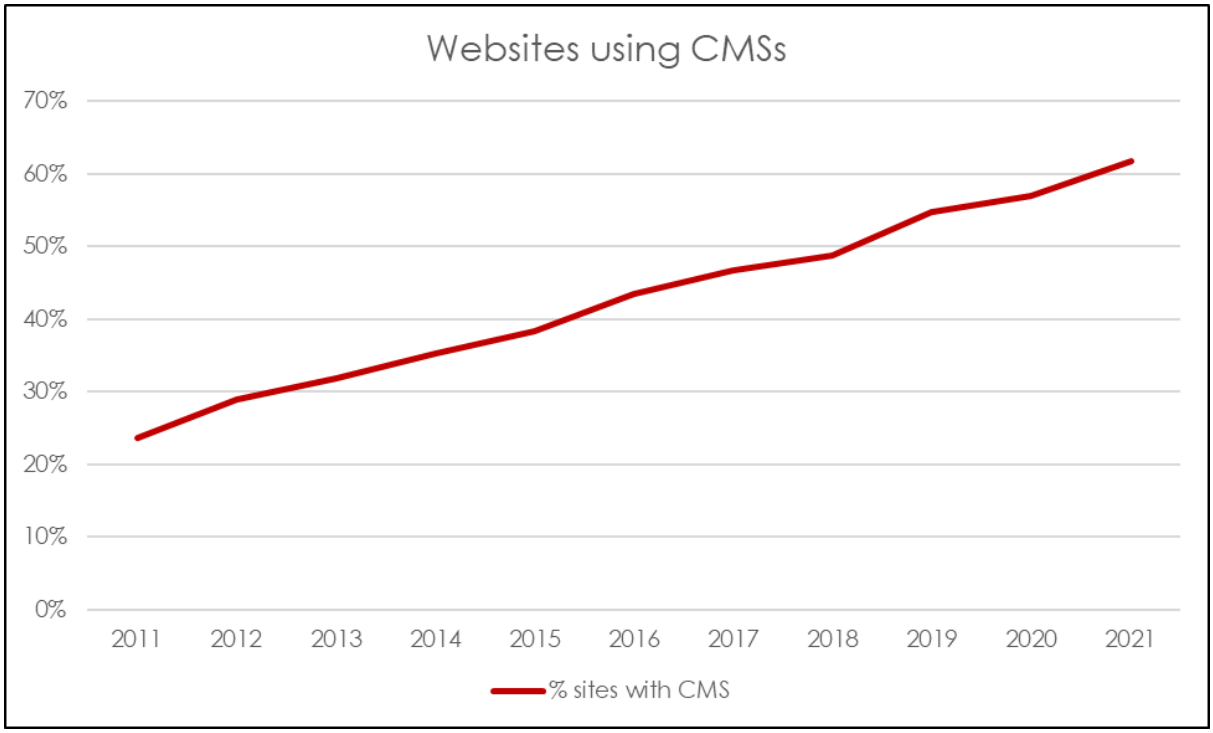


Figure 2.2: Proportion of websites using CMSs to manage content (based on data from W3Techs, 2021).

The primary purpose of content is to provide value for the user, though this is often restricted by the fast pace of online information change, rendering contents out-of-date, inaccurate, inconsistent, or hard to navigate. Investigating information quality within online content will provide a framework to ensure that content remains king, the most important element of experience on the web ensuring value remains for the user.

2.4. Content in the online world

Despite the origins of information systems theory within the offline world, online content differs in several key aspects have arisen since the widescale proliferation of the web. This section will explore a selection of the key issues facing modern web designers and the challenges they face in bringing informative content to users by exploring a series of key themes within the literature.

2.4.1. User generated content

As the web has evolved into a social platform, there has been an increasing reliance on *user-generated content* (UGC) to provide information for other users. There are a plethora of comparable terms used to describe UGC within different web contexts, though they share the common theme that there is a 'possibility for every user to publish content' (Wyrwoll, 2014, p. 11). George & Scerri (2007, p. 3) provide a concise definition for UGC and non-UGC, using the term engineered content to distinguish the former from more formally published information. They suggest that whilst the prevalence of user-generated content on the web is increasing, engineered content, 'usually has a high level of oversight and quality control and is generally considered to be more reliable and credible'.

Lazar et al. (2017, pp. 301–302) provide an alternative nomenclature for this divide between user-generated and published content, suggesting the terms *audience content* and *media content* instead. Websites with published information are considered as media content, with those informed by UGC as audience content, denoting the different methods in which such information should be treated within research contexts.

Understanding the context of the problem is important to identify both research methodology and outcomes of the study. The content described in this thesis pertains to published information, that has been authored by a professional for access by a wide range of audience members. User generated content cannot be engineered in the same manner as it is dependent on the understanding of users who cannot be influenced by the organisation operating a specific website.

To help address the question of user-generated content providing information for the user, the fourth action case study (validating the practitioner model) focuses on the examination of a Wikipedia article as a representation of such content. For more discussion on the selection of this

action case study and its impact on the analysis of inherent information quality, see Chapter 12.

2.4.2. Content management

The concept of published content on the web has led to the field of *content management*, the practice of authoring and managing information within organisation websites. As shown in Figure 2.2, organisations are increasingly turning to content management systems (CMSs) to categorise, organise and structure the process of writing information for websites. These platforms are a manifestation of the content management field, simplifying the process for professionals publishing online content.

One potential downside to the use of CMSs is the proliferation of bad content: when reused, the poor information spreads to multiple places within the system to the detriment of usability (Kostur, 2006, p. 193). Ensuring that the information contained within such a system is highly structured and appropriate to use is the role of the *information architect*, as described in the next section.

2.4.3. Information architecture

Organising information for ease of access has been required for as long as humans have held large catalogues of data. Many of these early catalogues were based on indexes of other texts, such as clay tablets or the books contained within the Alexandria Library (Rosenfeld et al., 2015).

The original classification of information belonged to professionals: the huge growth of libraries and the role of librarians as custodians of information in the late 19th and early 20th centuries is largely due to the introduction of systems such as the Dewey Decimal classification system. The Dewey system, still used in libraries around the world today, introduced a single, unified scale for classifying, storing, and accessing material by grouping themes of materials

within subject areas, applying hierarchical rules to generate specific labels or codes to aid users.

The use of *information architect* to describe the role of a person classifying and organisation information for others was first coined by Wurman in 1975, with its use constrained to Business Information Systems thinking of the 1980s and 1990s, where all problems involved organisation issues resulting in business and resource requirements (Resmini & Rosati, 2012, pp. 38, 41). This concept was brought to the early online world by Morville & Rosenfeld (2006), who used the parallels of physical architectural design as a parallel to the virtual world. The main difference between the concepts of Wurman and Morville & Rosenfeld is what is included within the content being organised: the former focused on elements of a page with the latter identifying the patterns required for navigating between and around several interconnected pages (Resmini & Rosati, 2012, p. 39).

Weinberger (2007) draws parallels between information within physical and digital worlds, leading to the classification of users as either *seekers* or *browsers*. The first category of user knows what information they are looking for and require efficient access to the relevant resources. By designing intuitive, easy to follow interfaces we can easily account for the behaviour of seekers and improve their experience of a system. The browsers require an alternative pattern: the ability to discover and experience new resources that they did not know they needed. The web must cater for both of these types of audience; however, the type of information can help to refine the problem of effective design. When working with informative resources, users are much more likely to be seekers, whereas for entertainment and social purposes, they are more likely to be browsers.

Moving towards the modern web, the role of information architects within web projects are recognised to be beneficial (Rosenfeld et al., 2015). When examining the process of designing effective user experiences, Garrett (2010) breaks this down into five key planes of design as shown in Figure 2.3.

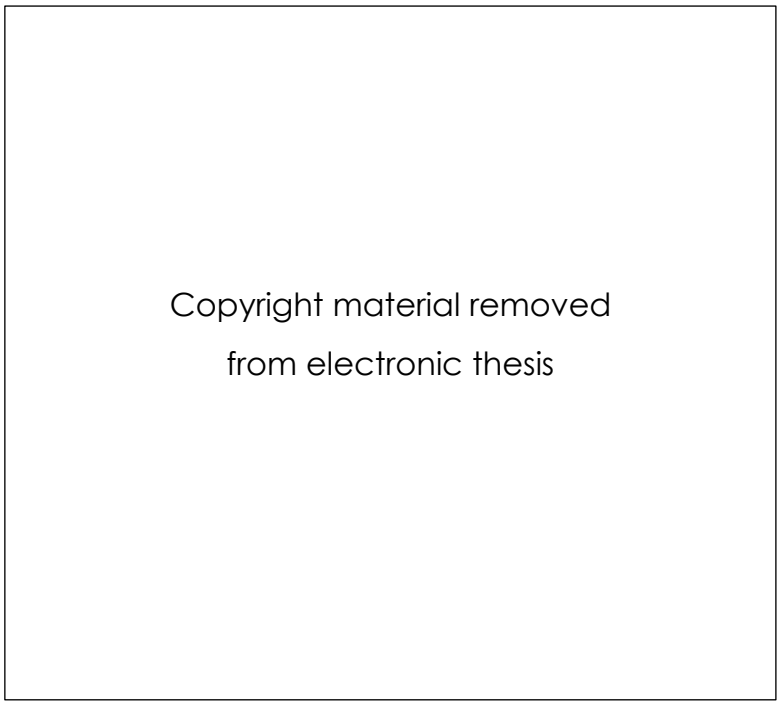


Figure 2.3: The five planes of online user experience (Garrett, 2010, p. 24).

Information architecture is crucial at all of the abstract stages of design, and information architects often work across all of these levels with a scope expanded from being purely information organisers.

To start at the most abstract level, information architects need to design a strategy for coping with large volumes of data, ensuring that this is appropriately scoped. Having large volumes of information is unproductive for users if there is no unified understanding of the purpose of the site and its design.

Structure is the most crucial element where IA needs to be involved: this is the point at which information must be classified to make it accessible. There are several different strategies for generating taxonomies of data, ensuring that such records can be easily found using relevant metadata. This must then be fed into wireframes, or the *skeleton* plane, to ensure users can successfully navigate between pages and content areas.

The oldest approach to online IA is to define the entire site within a hierarchical structure based on an inherent taxonomy such as chronological,

geographical or topical ordering (Dillon & Turnbull, 2005, p. 3) as shown in Figure 2.4. These can be used to aid users in accessing content, though the derivation of taxonomies requires careful consideration to match user expectations. Activities such as card sorting can be used to identify patterns in the way users explore topics to create appropriate categorisations for a website's audience (Rosenfeld et al., 2015). This may result in deeper hierarchies, with sub-categories within other categories to allow users to easily *drill down* into data (Whitenton, 2013).

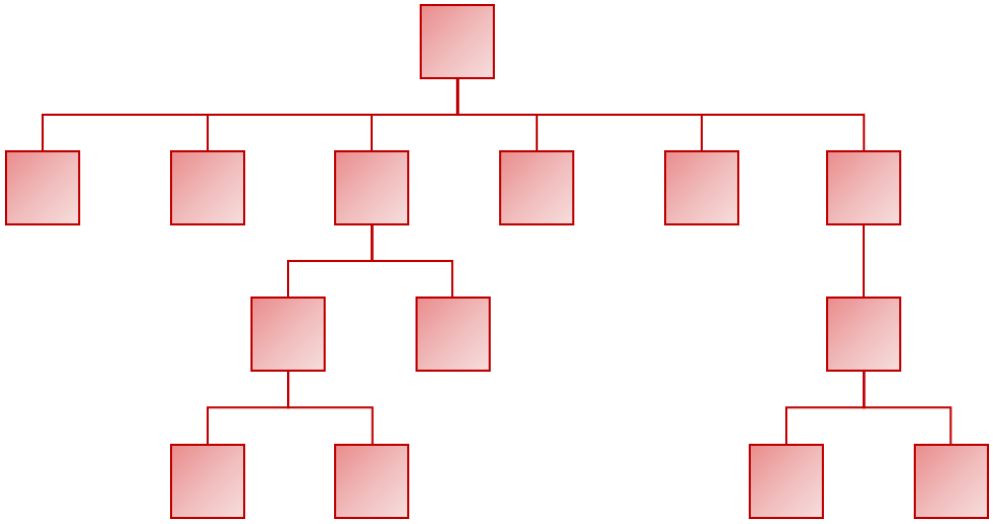


Figure 2.4: Example hierarchical website structure. Each page is linked from the home page, with a simple tree structure. Sub-categories provide further specific information as users explore the taxonomy.

As the volume of content hosted online has grown, single hierarchical structures are unable to fully match all users' expectations. Beginning with a hierarchical approach, introducing deep linking between thematically similar concepts can help through systematic linking to related content (Rosenfeld et al., 2015; Ruzza et al., 2017, p. 168).

Whilst information architecture may not directly influence the visual aspects of a website, the classification and organisational processes are crucial to allow users to access content using information seeking patterns that fit with their cognitive understanding of the subject. Wuman's original concept of

Information Architecture has historically been used widely in print media to identify the optimum layouts for visual information (Fenn & Hobbs, 2014, p. 15). Visual IA is often explored through wireframes, low fidelity prototypes that can be used for initial user testing and client approval prior to online development (Rosenfeld et al., 2015). These allow exploration of visual and navigational structures to evaluate effectiveness of typical user journeys.

The use of Information Architecture can aid designers and developers in the production of websites that provide information to users within expected patterns. The use of techniques such as card sorting and wireframes can elicit suggestions for contextual information design, highlighting issues prior to a website's launch.

2.4.4. Trust

Since the widespread growth of ecommerce in the early 2000s, many researchers have focused on the role of trust within website use. There have been numerous studies which examine consumer outcomes based on trustworthiness of websites (Büttner & Göritz, 2008; Ou & Sia, 2010) in addition to the importance of *credibility* as a factor in browsing habits.

Sillence et al. (2004, p. 666) examined the relationship between design and trust, highlighting that 94% of online credibility is governed by design aspects and aesthetics of a website. Whilst users tend towards familiar designs with reputable information, the credibility of a particular page is somewhat subjective and based on the experience and task of the user (Kąkol et al., 2013).

2.4.5. Speed

When navigating a website, users expect to find the information they require within a timely manner. Miller (1968) first explored the psychological needs of users for mainframe systems, identifying a discontinuity of conversational

response times at 10 seconds, with users waiting longer becoming disrupted by the slow access to information.

Figure 2.5 below shows the three thresholds Miller identified for user attention in mainframe use.

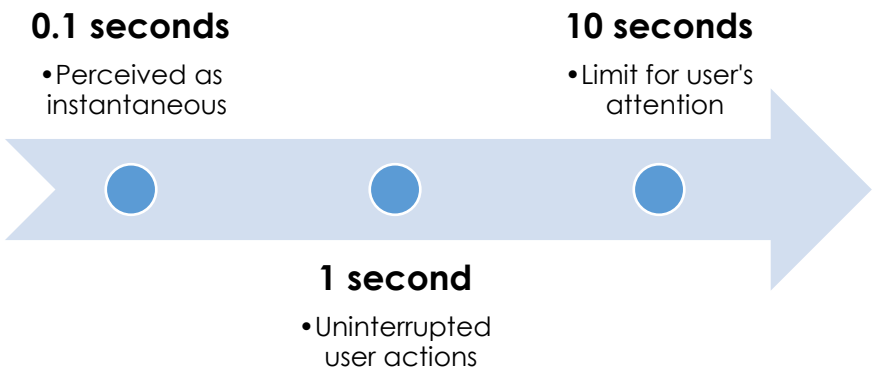


Figure 2.5: Computer use attention span thresholds based on Miller (1968).

Miller's thresholds described computer response times of 0.1 seconds as instantaneous, with the user not perceiving any difference with working with a human. 1 second response times are noticed by users; however, their actions are uninterrupted whilst waiting for a response. Waits of over 10 seconds not only interrupt the flow of user actions, often they lead to a break in attention span. Observing these effects in practice, Miller suggested an optimum response time of 2 seconds based on the mainframe technologies of the 1960s.

With the growing use of PCs in the 1990s, similar observations were repeated. Bickford (1997) observed web users and concluded that 'it takes only 8½ seconds for half on the subjects to [give up]', a figure which entered industry practice with a benchmark page load time of 8 seconds. This became known as *the 8-second rule*, a generally accepted principle that any page

accessed by the user should load within the first 8 seconds, otherwise the user will lose interest and abandon the site.

As the speed and ubiquity of Internet connections increased during the early 2000s, repeated experiments determined that the task being conducted can vary user expectations; for example, a simple page load is expected within 2 seconds whereas a complex request involving databases may be tolerated for 8–32 seconds. This can be considered as a quality issue as response time can affect a user's ability to access information and their overall website satisfaction (Kim & Stoel, 2004).

2.4.6. Search

As the web grew rapidly during the 1990s, many users accessed information by browsing, i.e., following links from one known page to another without the means to discover previously unknown content. This resulted in the need for a mechanism to enable users to easily find information across any number of different websites and information resources. By the end of the decade, *search engines* offered a low effort mechanism for seeking specific content within the growing number of websites. Search engines consult a set of indexed web pages for keywords, concepts or metadata to find relevant pages for the user, ranking them based on their query to improve the usefulness of results (Seymour et al., 2011). With the increasing volume of online information, search has become the 'user's lifeline for mastering complex websites' (Nielsen, 2001).

Searching a website is one mechanism for users to access specific content, and this can be useful to find individual items of content using metadata and keywords when a user's understanding does not match the underlying architectural model.

Typical site searches are for specific terms, with one large government website reporting half of searches using three words or fewer (Reynolds, 2013). Most studies of search engine effectiveness focus on average users

rather than their use by professionals (Lewandowski, 2005, p. 141) who may combine queries with complex operators to search for specific pieces of information that may be challenging to access through hierarchical architectures.

Despite users becoming accustomed to site search as a backup information seeking strategy, many navigational frameworks do not account for search (Karanam et al., 2015). When designing a website, content needs to be created in a searchable format with appropriate metadata and keywords, with prominent search areas visible for the user. If information is not able to found by users, the increasing reliance on external search engines means consumers may leave the website and search for information elsewhere (Dinet et al., 2012).

2.4.7. Accessibility

As the web has evolved, many formal and informal guidelines have been developed to demonstrate achievable methodologies for practitioners to deliver informative and accessible websites. The most widely followed web guidelines related to content is the Web Content Accessibility Guidelines (WCAG) 2.0 (W3C, 2008). The document defines three levels of compliance with online accessibility ensuring that all users, no matter their mental or physical impairments can access websites. Many governments and organisations define WCAG 2.0 compliance as a minimum standard for commissioned projects, with the EU standard (EN 301 549, 2014) encapsulating WCAG 2.0 for accessibility aspects.

2.4.7.1. Importance of accessibility

Within online communities, there has been an increasing focus on the social aspects of accessibility: the desire to provide comparable experiences for all. Whilst many practitioners are drawn to accessibility through social aspects, for example by closing the digital divide for those with disabilities, organisations are often driven by standardisation or regulatory requirements

(Yesilada et al., 2012). Providing equal access to information for all citizens and users is both a political and moral imperative, with such rights part of the United Nations Convention on the Rights of Persons with Disabilities (UNCRPD) agreed by most international governments (European Commission, 2021). Many high-performing companies opt to comply with accessibility guidelines over and above their basic legal responsibilities as this has been demonstrated to enhance brand perceptions, increase market share and provide innovation for all customers (W3C, 2018b).

2.4.7.2. Evolution of WCAG standards

Over the past two decades, the WCAG standards have evolved based on practitioner feedback and agreement by working committees. The diagram in Figure 2.6 shows this evolution between the three major revisions of WCAG.

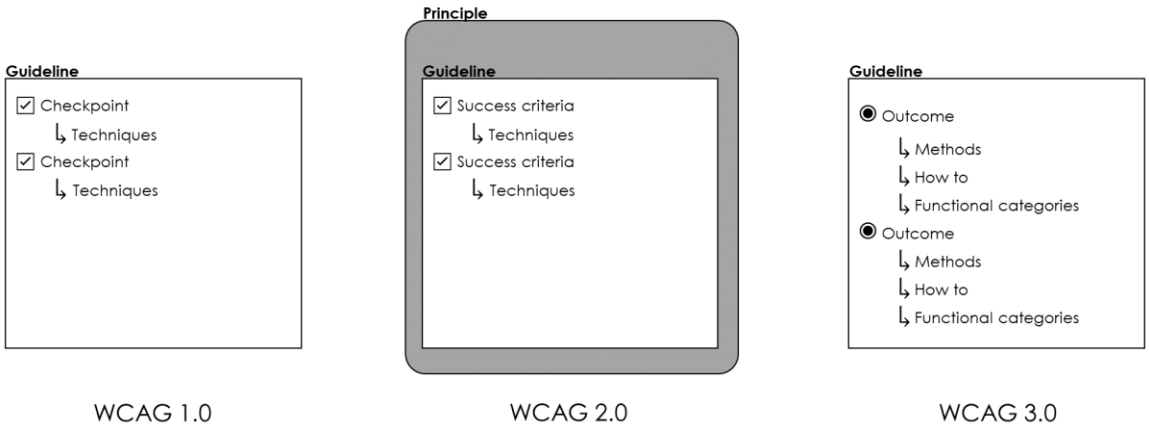


Figure 2.6: Comparison of WCAG structures.

The first version, WCAG 1.0, was based on a simple checklist approach to ensure that minimum criteria were met by websites. The checkpoints were generated based on small adjustments for users with common physical disabilities, whilst the authors included suggested *techniques* to demonstrate

to practitioners how the checkpoints could be achieved. This was important in the first edition as accessibility was often an unknown consideration for web developers with no prior or specialist knowledge.

The second edition of WCAG 2.0 changed the nomenclature of *checkpoints* to *success criteria*. This represents a shift away from a checklist basis, incorporating scaling criteria to represent the varying levels at which websites could comply with the guidelines. This was accompanied by a change to levels of compliance, representing the ways in which partial compliance with the criteria could help improve accessibility without reaching an unaffordable and challenging gold standard. Due to the proliferation in the number of guidelines, four groupings called *principles* were also included to make compliance easier to manage.

WCAG 3.0 is in draft mode and is planned for full publication during 2022. This represents a large step change to the approach for evaluating accessibility, with guidelines now represented by outcomes, each of which have practical advice such as methods, how to examples and functional categories to improve online implementation. Individual outcomes are scored on a metrics-based scale, allowing for automated compliance testing and the ability to evaluate and compare individual sites.

Each of the three evolutions of the WCAG are discussed below.

2.4.7.3. WCAG 1.0

The first version of the Web Content Accessibility Guidelines was released in 1999 as a community-driven attempt to create a short checklist for developers to improve content for those with accessibilities. This was driven by three rising trends: making telecommunications more accessible, increasing political attention on rights of those with disabilities and the growth of the Web as an informational platform (Ellcessor, 2010). The confluence of these trends resulted in a focus of users with known physical disabilities,

developing key criteria for accessible websites without reference to content, design aesthetics or non-physical or unknown disabilities.

As the Internet developed with increasing numbers of users, newer multimedia technologies and novel applications, there was a growing need to address the general accessibility concerns of a typical user, not just those with known and classified disabilities. This led to the development of WCAG 2.0, the second version of the guidelines in 2008.

2.4.7.4. WCAG 2.0

The WCAG 2.0 guidelines were released in 2008, built upon the earlier WCAG 1.0 guidelines. The main differences between the two standards is in the introduction of approaches to deliver accessibility through new multimedia elements that became commonplace on the web as design standards, devices and browsers evolved (Termens et al., 2009, p. 1171). Keeping up with technological developments has often been a challenge for standards and best practice guidance, a challenge still faced by practitioners on today's standards compliant and accessibility aware web.

WCAG 2.0 has brought about a more accessible Internet, however it focuses more on the system and presentation aspects (commonly considered as HCI in academia and UX in industry) than the content itself. In addition, many studies have demonstrated that the guidelines are insufficient for their purposes as they fail to address the needs of those with the impairments considered in the creation of the document (Brajnik, 2009; Debevc et al., 2011; Lewthwaite, 2014; Power et al., 2012). The guidelines can also be difficult for designers and developers to interpret and implement (Donnelly & Magennis, 2002, p. 57). Empirical studies have demonstrated the difficulty of applying the guidelines consistently, with an average agreement between auditors of 70-75% (Brajnik et al., 2012, p. 8:1), furthering the notion that WCAG 2.0 leaves many subjective choices to those involved in the development lifecycle without a thorough consistency of approach.

From September 2018, all public sector websites within the EU and UK are mandated to comply with WCAG 2.0 as a minimum standard of accessibility (European Parliament, 2016). As many existing UK-based websites do not fully comply with every aspect of WCAG 2.0 to AA standard, there will be a staged approach with the use of accessibility statements to denote content that is not universally accessible (Government Digital Service, 2018).

2.4.7.5. WCAG 2.1: Sufficient for the modern web?

After ten years' use in practice, WCAG 2.0 has received its first revision to become WCAG 2.1 (W3C, 2018a). This new version addresses some of the emerging technology issues, incorporating new success criteria for gestures, screen orientations, touch screens and text inputs reflecting an evolution of web design. Whilst providing the criteria aids in the development of sites that work seamlessly across all devices, many of the attributes discussed have been used by practitioners for many years, with industrial best practice providing sufficient guidance to exceed the minimum standards required by WCAG 2.1.

2.4.7.6. WCAG 3.0 and the future

WCAG 3.0, currently in draft stage, represents a step change in the approach to web accessibility evaluation. The change to outcomes rather than success criteria moves the assessment process from checkbox-based activities to a focus on the use of web pages by users, particularly those with accessibility concerns. This includes the reliance on techniques and how-to documentation to educate professionals on the most contemporary techniques required to provide accessible websites. There is a general aim to author all content within plain language to avoid technical nomenclature and lower the barrier to professionals working with the guidelines (W3C, 2021).

The shift towards metrics-based assessment aids in the development of automated tools and comparison of techniques, however quantitative data can often be hard to interpret without understanding the reference points

and nuances of user behaviour (Budiu, 2017). Such techniques are useful for large-scale evaluation of websites, for example by researchers (Freire et al., 2008) or search engines though cannot cover every possible interaction with the page. When working on specific pages, the meaning provided by qualitative user research can provide deeper insights into behaviours and identify patterns which require accessibility adjustments.

The scope and scalability of the changes within WCAG 3.0 are yet to be evaluated by the community. As the standard is developed and published, the insights into its effectiveness from accessibility and usability professionals will add to the discourse on the applicability of the new metrics-based approach.

More discussion on the WCAG 3.0 model and its differences is included in Chapter 11, where the practitioner model is defined.

2.4.7.7. Summary

Despite the latest update, WCAG 2.1 still fails to meet the criteria of providing accessible and usable content for all. The scenarios of use demonstrate how the individual points satisfy the use of assistive technologies, however these do not refer to patterns of use by real users and are simplified into desired patterns of behaviour (Kreps & Goff, 2015). Whilst implementing WCAG-compliant websites, developers still need to perform sufficient user testing and understand industry best practices in order to meet user needs. This reflects the trend towards the third wave of human-computer interaction where metrics-based analyses of task completion are insufficient in supporting an integrative user experience (Bødker, 2006; Cooper et al., 2012) within a complex environment (Sloan & Kelly, 2011).

2.4.8. Design

Aesthetic design of web pages can support both informative content and functionality (Thorlacius, 2007, p. 68), though there has been a trend towards

design minimalism to 'present content and features in a simple, direct way' (Moran, 2015a). The goal of minimalist website design is to present information as cleanly as possible, without distracting the user with irrelevant elements (Moran, 2015b). It has been demonstrated that simplistic websites can improve searchability and user recall (Baughan et al., 2020) by limiting the number of distracting or irrelevant elements.

Whilst pages can be developed using minimal features, another key aspect is the visual appeal of a website. The desirability of a web page's appearance can depend on the colour scheme, typography, and culture of the user, in addition to the content itself (Varela et al., 2013, p. 74). There is also a difference in aesthetic appeal between expert and non-expert users, with the former spending more time examining details with the site rather than the overall appearance (Pappas et al., 2018).

The density of information provided by a website can be affected by user culture: for example, Chinese websites may contain larger numbers of items within the same screen space as those in the English language (Chu & Yang, 2010). This demonstrates that minimalism cannot be applied without considering the user context and norms, though extraneous features should always be avoided wherever possible. This has been encapsulated by the Government Design Principles (Government Digital Service, 2012) which suggests that simplicity is key to creating accessible online services, even though 'it's usually more and harder work to make things simple, but it's the right thing to do'.

2.4.9. Conclusion

There are many aspects to consider when examining online content within websites. Industry practitioners have developed several standard approaches such as content management methodologies, information architecture and accessibility frameworks to aid in the production of informative websites without the need to re-evaluate interfaces with every

new project. This has helped to drive innovation on the web, with new products built on existing principles and frameworks matching users' functional expectations and information seeking patterns.

Many of these expectations are shaped by users experience of the entire web, as demonstrated by Jakob's Law: 'Users spend most of their time on other sites, and they prefer your site to work the same way as all the other sites they already know' (Yablonski, 2020). As users do not enter a website free of any notions of structure, navigation, or presentational aspects it is important to follow conventions wherever possible to match their expectations. These can be in many forms such as following hierarchical architectures, minimising page load times, and providing accessible features. Within large organisations this often leads to the creation of design systems with common features to enable the rapid development of consistent, standards-based, and brand-compliant websites whilst minimising costs.

Separating the informative and visual/structural aspects of a website can be challenging as many information seeking patterns rely on facets of accessibility and usability, though this is important to link practice with academic theory. The next section will discuss the varying levels of clarity between content and design.

2.5. Content and design separation

A continual contention within the web development industry is the separation between *content* and *design*.

Within academic theory, the separation between content and design was first identified by Huizingh (2000, p. 123), who defined the divide as, 'Content refer[ing] to the information, features, or services that are offered in the Web site, design to the way the content is made available for Web visitors'. This provides a clear distinction between the two terms which was easy to apply to websites at the time the article was written, however contemporaneous

use of web technologies has blurred the distinction.
Hernández et al. (2009, p. 363) further identified the distinctions between content and design within website analysis metrics.

This dichotomy is proliferated by Thielsch & Hirschfeld (2019, p. 283) who suggest that there is a large focus on content from a quantitative perspective, e.g. search engines and automatic classification tools neglecting the qualitative aspects of user perceptions. They further conclude that perceptions of content must be clearly delineated from the design and usability aspects to optimise information access.

The evolution of web technologies such as HTML5 and CSS3 has aided development teams to separate content and design by keeping styling and presentation aspects separate from structure and information – this is a core principle of the W3C standards (W3C, 2016a).

Clark (2007, pp. 57–58) argues that the separation between content and design is rarely as straightforward as suggested by practitioners: for example, some presentational aspects are required for content structure, and such divides are often dependent on the technologies in use. This is leading to an increasing reliance on design principles in technical communications, requiring presentational aspects to convey key information.

One demonstrable example of this shift is the rise in infographics to represent challenging statistics within relatable media. First used by journalists in the 1980s to illustrate news articles (Siricharoen, 2013), infographics are now a popular means to communicate information to non-technical audiences. Such visual representations have been demonstrated to improve access and retention of key information, especially where the presentation also narrates a story (Dunlap & Lowenthal, 2016, p. 42).

The content and presentational aspects are often challenging to separate, as described by the *information quality* and *service/system quality* aspects of

the DeLone-McLean (2003) model (see p. 78). Ensuring that user perceptions are based on either cognitive understanding or visual affordance can affect the outcomes of qualitative studies and requires careful consideration to avoid conflating the two areas of web design (see p. 85).

2.6. Summary

The Internet would not exist without content. Even though the first web page was text-based, and multimedia was introduced early in the existence of the World Wide Web, content remains the major driver in use of online spaces. Depending on the context, the word 'content' can have various definitions. For the purposes of this research, all content is assumed to be text-based.

The main driver towards a more accessible web is the application of the various WCAG guidelines, whether through voluntary developer use or Government-mandated standards. Despite the widespread focus and understanding of the WCAG model, few suggestions have been made to extend the approach to other content-related themes such as information quality.

The following chapter discusses the theme of information quality, including an analysis of existing frameworks and guidelines to identify potential IQ attributes that may influence the quality of online content.

Chapter 3. Information quality

This chapter reviews information quality attributes within the context of the field of Information Systems. This overview positions the study within existing frameworks, building on practitioner and academic models to situate the research.

Information quality is an often-quoted term within academic models, occurring in many guises and with a variety of context-dependant meanings. To understand what is meant by information quality, this chapter begins with an overview of information systems and definitions of information, information quality, and informativeness within the context of online content.

3.1. Information systems

The field of information systems (IS) has grown in parallel to the rapid development of computer hardware and software since emerging in the 1940's. Despite this fast pace of change, the rate of thinking about the social and cultural impacts of such technology has not developed to the point where we have a thorough understanding of the issues involved (Checkland & Holwell, 1997, pp. 8–9).

Information systems is a hybrid of many different domains, primarily computer science and management. This is supported by a large number of other disciplines, including 'psychology, sociology, statistics, political science, economics, philosophy and mathematics' (Boland & Hirschheim, 1985, p. vii).

The history of the field is varied and not commonly explored by contemporary researchers, leading to activities that routinely fail to build on prior studies, theory and paradigms (Bryant et al., 2013; Hirschheim & Klein, 2012). Despite the growth in information systems research, there is still a lack of recognition of the need for the development of IS-specific theories, particularly those

targeted towards ontological questions (Gregor, 2006, pp. 611–612; Watson, 2001; Weber, 2003). As a result of this, many information systems studies base their philosophical foundations within other fields: computer science, business & management, and the social sciences. The influence of computers in our everyday personal lives and organisations has blurred the boundaries between computing and every other field (Markus, 1999, p. 176), leading to topics traditionally categorised as information systems being studied within other fields.

Hirschheim & Klein (2012) describe the field as being in a fourth era of research, representing the rising use of information systems technology within our everyday lives. The prevalence of Internet enabled communications has changed our ways of working, learning, and socialising resulting in the emergence of new theories to help develop more effective strategies for successful systems. Much of the academic research now focuses on web technologies such as analytics, social media, and search engine optimisation (SEO), building specific devices to aid technology aided research. The global reach of the Internet has led to researchers contrasting approaches within different cultural contexts and within virtual environments.

The distinction between information systems and other systems development activities is observed by Davis (1999, p. 196) who notes that technical activities cannot be outsourced, whereas the core planning and implementation activities of IS require specialist internal knowledge. Much of the information systems literature is practitioner focused due to the commercial requirement for employees knowledgeable about IS theories and strategies (Hirschheim & Klein, 2012, p. 219), however it is often noted that academic theories lag behind practice in industry due to the long lead times of publication and the common role of academics as observers and reflectors on existing systems (Benbasat & Zmud, 1999).

A key characteristic of information systems research is the focus on questions such as (Boland & Hirschheim, 1985, p. vii):

- How can IS technologies be applied to produce the desired outcomes?
- How can systems be managed effectively?
- What are the wider implications of systems?

The study of such questions has produced a wide body of knowledge combining insights from numerous different domains to produce the theories and application to practice which combine with the contemporary field of information systems.

3.1.1. What is information?

Within the literature, there are conflicting definitions of data, information and knowledge based on the origins of the author's works. Checkland & Poulter (2006, pp. 112–113) provide one such relationship model between these terms, leading the researcher to the definitions in Table 3.1 that are used throughout this report.

data	Symbols that represent the properties of objects and events (Ackoff, 1999, pp. 170–172) in their raw format.
capta	A subset of data that 'we have an interest in knowing' (Checkland & Poulter, 2006, p. 112).
information	Data that has been processed to improve its usefulness for its intended purpose by creating meaningful facts.
knowledge	Larger, longer and living structures of meaningful facts (Checkland & Poulter, 2006, p. 113).

Table 3.1: Definitions of data and information within this report.

As shown in the table, the smallest and most raw symbols are referred to as data. These are unprocessed properties without any attached meaning. A subset of data can be selected as a collection, focusing on a particular interest. Once processed, these will form information or meaningful facts. Finally, humans can use information to create knowledge, the larger structures that allow us to connect and understand meaning within data. Although Artificial Intelligence is working some way towards creating and maintaining knowledge, the general computer systems of the modern web can only contain information ready for interpretation by the user.

Whilst these definitions suggest a clarity of the distinction between data and information, this does not become apparent when studying the quality techniques applied to each; for example, content that is traditionally thought of as raw data may have already been processed into information, prior to further processing for additional purposes. Additionally, Lee (2004, p. 10) notes that the term 'information has come to be used interchangeably with data and knowledge'.

The ambiguity of information has been documented and explored since before the growth of the web; for example Buckland (1991, pp. 351–352) used this uncertainty to define the four aspects of information as shown in Table 3.2.

	Intangible	**Tangible**
Entity	Knowledge	Data or document
Process	Being informed	Data processing

Table 3.2: Buckland's (1991, p. 352) four aspects of information.

Buckland's model places the four aspects on two dimensions: tangibility, and whether the terminology refers to an entity or a process. Using these aspects, all of the various definitions of information can be interpreted and included

with their meanings. For example, information in a knowledge sense is intangible and an entity, whilst information contained within an online web page could be considered to be a tangible entity.

Mingers (1995, p. 287) questions the idea of such objective or absolute information as the same data or documents may not have the same impact on a different user as they will form their own resultant knowledge. This results in the observation that information is an objective commodity: it is available whether or not it is being extracted to form knowledge (Mingers, 1995, p. 290).

The use of the term *information* can be fraught with confusion due to the multiplicity of contextual meanings and their relationship with the production and maintenance of knowledge. Whichever meaning is intended by the researcher, the notion of information is often based on intuition (Floridi, 2005).

In this research, information means the processed data within a web page for access by the intended user. This means that all information is tangible and can be incorporated into the study, though different users may have alternate perceptions of its meaning and the way it incorporates with their existing knowledge.

3.1.2. What is information quality?

Information quality has been a much-explored field within information systems literature, with many studies demonstrating that improving the quality of a system will result in higher rates of acceptance and ultimately higher rates of success. The challenge of such a term within a multi-disciplinary field such as information systems is that there is no single common definition. The next three subsections will explore the meanings of information, quality, and information quality, defining the terms within the context of this research.

Much of the Information Systems research defines quality to be equivalent to fitness for purpose (Bovee et al., 2003; Wang & Strong, 1996, p. 6), i.e. does

the provided information serve the purposes for which it is intended. This implies that quality is contextual and dependent on the data processed, the intended purpose of the information and the user's expectations of the information. Strong et al. (1997, p. 38) suggest that 'poor information quality can create chaos. Unless its root cause is diagnosed, efforts to address it are akin to patching potholes'.

The use of the terms data and information are mixed within the quality literature as throughout information systems, although Madnick et al. (2009) document a 'tendency to use data quality to refer to technical issues and information quality to refer to nontechnical issues'. This is further confirmed by a systematised study of data quality literature within the Information Systems and Computer Science fields which concludes that the 'IS community is focused on the use of information for decision making and the business values of high levels of quality of information' (Sadiq et al., 2011, p. 158), that is the production of information using data to inform c-level decisions.

Information quality is 'a complex and multi-dimensional phenomenon, which has yet not been fully understood' (Ge et al., 2011, p. 2). Organisational failure is common, with an estimated 60% of medium-size companies suffering from known information quality issues (Wand & Wang, 1996).

Drawing on the information systems theory described in this chapter, Petter et al. (2008) define information quality as, 'the desirable characteristics of an information system's outputs'. This definition focuses on the use of attributes as nomenclature for the aspects of a system, and is often measured as an element of user satisfaction and not accounted for within an individual aspect (Al-Debei et al., 2013, p. 103).

The link between the information quality of content within the information systems and computer science fields is weak with the IS field focusing on issues of satisfaction and impact on business decisions, without the consideration of technical content quality (Sadiq et al., 2011, p. 159). The

authors further suggest that there should be collaboration between the two groups of researchers to produce literature with a common focus on the success of content.

Assessing quality within an information system 'involves measuring the quality dimensions that are relevant to the information consumer and comparing the resulting scores with the information consumer's quality requirements' (Bizer, 2007, p. 23). This demonstrates the importance of domain-specific criteria for quality: what constitutes quality in one domain may not be consistent with the criteria within another. As a result, many of the generic information quality results may not be applicable within online content, a specific presentation to users. Domain-specific IQ frameworks aiming to address this inconsistency are discussed in Section 4.4.

3.1.3. What is informativeness?

The definition of informativeness widely varies within the literature whether the researcher is conducting studies within a positivistic or interpretive paradigm (see p. 133).

Within a positivistic epistemology, data is considered with a quantitative approach, and this leads to informativeness referring to 'the perceived amount of valuable and useful information given in a website' (Thielsch & Hirschfeld, 2019, p. 15), though this may not match the absolute amount of information available on a website (Chakraborty et al., 2005, p. 422).

The amount of information is considered a separate concept to the quality of information contained within a website, with the suitability of online content considered more subjectively by the user (Barnes & Vidgen, 2002). This provides a qualitative viewpoint with scope for the inclusion of user perceptions instead of metrics. Sadiq et al. (2011, p. 159) propose a one-way dependency between good data quality and good information quality, suggesting that the latter can only be produced with the former. This focuses on the use of metrics which may not always be appropriate to the context.

When considering informativeness online, most prior studies take a metrics-based approach. This is valuable for analysing known corpora in structured formats, however, may be unsuitable for the web with a varied audience who rely on individual perceptions of information to add to their own existing knowledge base.

3.1.4. Why study informative content?

Since the web become ubiquitous in everyday lives, there have been numerous studies demonstrating the positive impact of informative content on website effectiveness. Despite the widespread correlation between these studies, there is still no unified measure of user perceptions of web content (Thielsch & Hirschfeld, 2019, p. 4). Along with the attribute of usability, the informativeness of content online can have a large impact on the user's ability to find and access information pertaining to their task. This study will focus on providing the missing link between informative content and webpage success by producing evidence-based advice for practitioners to improve the outcomes of online websites.

3.1.5. Summary

The historically varied use of the term *information* has led to ambiguity within literature and online practice. Models such as Buckland's (1991) aspects of information aim to remove this vagueness by identifying the specific nomenclature based on the context of study. Ultimately informative online content should be improved to increase knowledge gained by users from use of a website.

The following section discusses theoretical models of information systems success, and how they can be applied by academic studies to evaluate individual case study websites.

3.2. Information systems success

Information quality is an important measure of IS success from a user perspective (DeLone & McLean, 1992), and can be used to determine website success (Liu & Arnett, 2000).

'An information system (IS) has many stakeholders, each with a different definition of system success' (Briggs et al., 2003, p. 6), challenging researchers to develop encompassing definitions of such success: is financial, time or other metrics-based criteria a true measure of the system? As one of the oldest research traditions within IS, the concepts of success and failure have been extensively researched even though the terms are difficult to precisely define (Dwivedi et al., 2015, p. 144).

In their survey of information systems projects, (Nelson, 2007, p. 74) identified that 37% of project failures are caused by quality issues. Despite decades of research into the causes and risks of information systems failure, the rate of unsuccessful projects has not changed (Nelson, 2007).

This section of the thesis will examine what impacts on information systems success by first discussing failure, then using this to explore academic models which can be applied to research contexts.

3.2.1. Systems failure

To understand what information systems success is, it is necessary to examine the opposite: what is information systems failure?

Flyvbjerg & Budzier (2011) note that, 'new research shows surprisingly high numbers of out-of-control tech projects – one that can sink entire companies and careers', with public sector projects more likely to fail than those in the private sector (Holgeid & Thompson, 2013, p. 1).

The failure of information systems has been much explored within the literature, with one such definition provided by Ewusi-Mensah (2003): 'Either

the implemented system not meeting the user expectations or inability of creating working or a functioning system'.

This generic definition means that information systems failure can be apparent in many different forms: whilst it is most frequently noted within systems which lack particular functionality, any element of a system not working as intended can be described as a failure.

Systems failure can be caused by many different factors during a system's development. To help analyse the reasons for systems failure, Dwivedi et al. (2015, p. 146) identified five perspectives within existing literature. Two of these, social and user resistance, are particularly relevant to online websites.

3.2.1.1. Social

The social perspectives identify two key attributes to systems failure: failing to meet expectations or termination failure.

The first of these is a broad attribute corresponding to a system not meeting user needs due to resistance to change. This resistance can manifest in many forms which are often not apparent during the analysis and design stages of a project (Hirschheim & Newman, 1988, p. 399). Users must feel confident that a system meets their needs and does not cause future uncertainty over their role within the task whether through lack of information, poor quality design or misunderstanding the system. These can have profound effects on organisations where the human element is not fully considered at the design stage, leading to systems which are technically capable yet do not anticipate user needs.

The second attribute is *termination failure*, first described by Sauer (1993). This relates to the perception of a project as seen by the stakeholders – once the amount of potential small failures or flaws reaches a certain threshold, a project may be cancelled or abandoned. This a different outcome that failure: an abandoned project has a perceived inability to meet system

requirements rather than an abject failure to perform the task (Ewusi-Mensah & Przasnyski, 1994, p. 185).

3.2.1.2. User resistance

Klaus & Blanton (2010) identified four types of user resistance: individual, system, organisation, and process. Together these types can be used to classify resistance to the use of new technologies based on technology use behaviours.

Individual resistance is based on intrinsic psychological determinants that must be satisfied for a user to accept a system. If these four determinants are not satisfied, then there will be a tendency towards uncertainty: whether this is through automation, threat to current working practices or change of routine. Within the web space, this can be likened to users stumbling across an unfamiliar website using a non-standard navigational structure. If users do not know how to interact with a system, then their intrinsic beliefs about information seeking patterns will be threatened.

System issues can be caused by technical problems or complexity. Firstly, technical problems are often caused by bugs or unintended features. Websites with such errors can be hard to use, frustrate users and reduce the likelihood of successful information patterns. A similar effect can be observed with complex systems: if systems or their inherent features are too hard to understand, navigate or interact with then users will not be able to find and process the information required to satisfy their needs.

Organisational issues are more apparent in expert systems targeted at employees of particular organisations, though one key aspect of organisation issues is that of communication. If information is written clearly with adequate signposting, it is more likely that user tasks will be successful.

Finally, resistance can be as a response to the process, particularly when familiar systems are adjusted and changed. Whilst this is most relevant to

users with a vested interest in the organisation such as employees, this may become an apparent issue when users are required to repeat tasks or conduct more work to obtain the same outcome. A manifestation of this in web systems could be interfaces which require repeated user actions or repeated data where clearer steps and interfaces could reduce the amount of activity required to obtain the same results.

3.2.1.3. Summary

On the web, there may be many reasons for systems failure. Beyond the traditional organisational thinking of the IS field, there are many expedient explanations for the small-scale failure of systems to meet user needs. This can be down to the system not meeting expectations, unfamiliar and complex navigational patterns or changes to processes that are inconsistent with user beliefs. These notions are further explored in Section 3.3 which discusses academic theory related to information seeking behaviours within an online perspective.

3.2.1.4. Conclusion

The study of the failure of information systems draws on theories and literature from many fields, including technical and sociological literature. The broad spectrum of factors represents the breadth of research within the field and an inability to study systems within isolated contexts. This leads many studies to *interpretivistic* research, where particular cases are studied in-situ and findings are based on user perceptions and observations within individualised contexts.

3.2.2. Systems success

The success of specific information systems has been the subject of empirical-based literature for many years, with researchers applying a variety of different taxonomies of related variables to conceptualise the complex relationships involved. The relationship between IS failure and success is not a simple dichotomy, and as (Glass, 2005, p. 110) notes, 'How do you categorize

a project that is functionally brilliant but misses its cost or schedule targets by 10 percent? Literalists would call it a failure, realists a success'. Even with a concrete model of normative success factors, systems neglecting the political context can still result in failure (Lee & Myers, 2004), with the method of implementation (Sherer et al., 2003), the background of employees (Heeks, 2006) and the existence of many stakeholder groups (Grover et al., 1996, p. 183) other salient issues to be considered.

Whilst the main focus of information quality is the perceived usefulness, (Castañeda et al., 2007) note that users require a 'minimum level of ease of use' in order to evaluate a site's usefulness. This is based on the Heuristic-Systematic Model, where the user focuses on the aspects requiring a minimum effort until they are sufficiently motivated and experienced to not abandon the site. This is confirmed by other studies (Morris & Dillon (1997) and Teo et al. (1999)) that conclude that usefulness depends on ease of use. Studies such as Klein (2002) have demonstrated that users often confuse information and technical problems with online websites, leading to assumptions that a system is failing when it could be due to poor quality information. Even where an organisation is aware of poor-quality information, the impact is often difficult to calculate beyond the cost of maintaining the data through improved processes (Haug et al., 2011).

Petter et al. (2012, pp. 354–355) observe that the present changing scope of information systems challenges researchers to develop practices to incorporate the needs of a wider group of stakeholders within success criteria. The use of more qualitative methods is required to capture intangible benefits that may not be accounted for easily with existing objective research frameworks. This is particularly important for the study of websites where observational qualitative studies can discover insights into common user experience concerns (Budiu, 2017).

3.2.3. TAM

The Technology Acceptance Model (TAM) was developed by Davis (1989) as an instrument to predict new technology adoption within a group of people. This is based on Fishbein & Ajzen's (1975) theory of reasoned action. The model provides 'a basis for tracing the impact of external factors on internal beliefs, attitudes and intentions' (Davis et al., 1989, p. 985). Davis devised the TAM model following a software program with a well-designed user interface having low user take-up due to the system not being perceived as useful (Benbasat & Barki, 2007, p. 212).

TAM is based on the concept that perceived ease of use and perceived usefulness are both important user behavioural determinants (Chuttur, 2009, p. 4; Davis, 1989), based on the work of Swanson (1982). Definitions and examples of these two concepts are shown in Table 3.3.

Term	Definition	User experience example
Perceived ease of use (PEOU)	'The degree to which a person believes that using a particular system would be free from effort' (Davis, 1989).	PU is a measure of how easy a user finds the system to use, for example an online system may by intuitive to navigate yet lack informative content.
Perceived usefulness (PU)	'The degree to which a person believes that using a particular system would enhance his or her job performance' (Davis, 1989).	This term relates to attitude towards the system helping produce better task outcomes, for example relaying larger volumes of more accurate information to users.

Table 3.3: Definitions of TAM concepts.

The two concepts of perceived ease of use and perceived usefulness can thus be understood as relating to two different aspects of website use: the former the availability of online information, and the latter be navigable with a lack of expected content. Focusing on these two aspects in isolation does not address many of the other considerations when producing informative aspects.

3.2.3.1. TAM2

As a result of the plethora of studies applying TAM-based approaches, Venkatesh & Davis (2000) propose an updated model to incorporate *social influence processes* (subjective norm, voluntariness and image) and *instrumental processes* (job relevance, output quality, result demonstrability, perceived usefulness) as additional variables important to system acceptance, based on works completed during the intervening period. This development is often termed *TAM2*, and it is suggested that this enables the model to explain 60% of technology acceptance within systems (Venkatesh & Davis, 2000).

3.2.3.2. TAM3

As a further development, Venkatesh & Bala (2008) adds extra influences based on user experience, denoting their impact on computer anxiety, perceived usefulness and perceived ease of use. This recognises the development of additional variables within contemporary TAM-based studies. The TAM1, TAM2, and TAM3 models are compared in Figure 3.1.

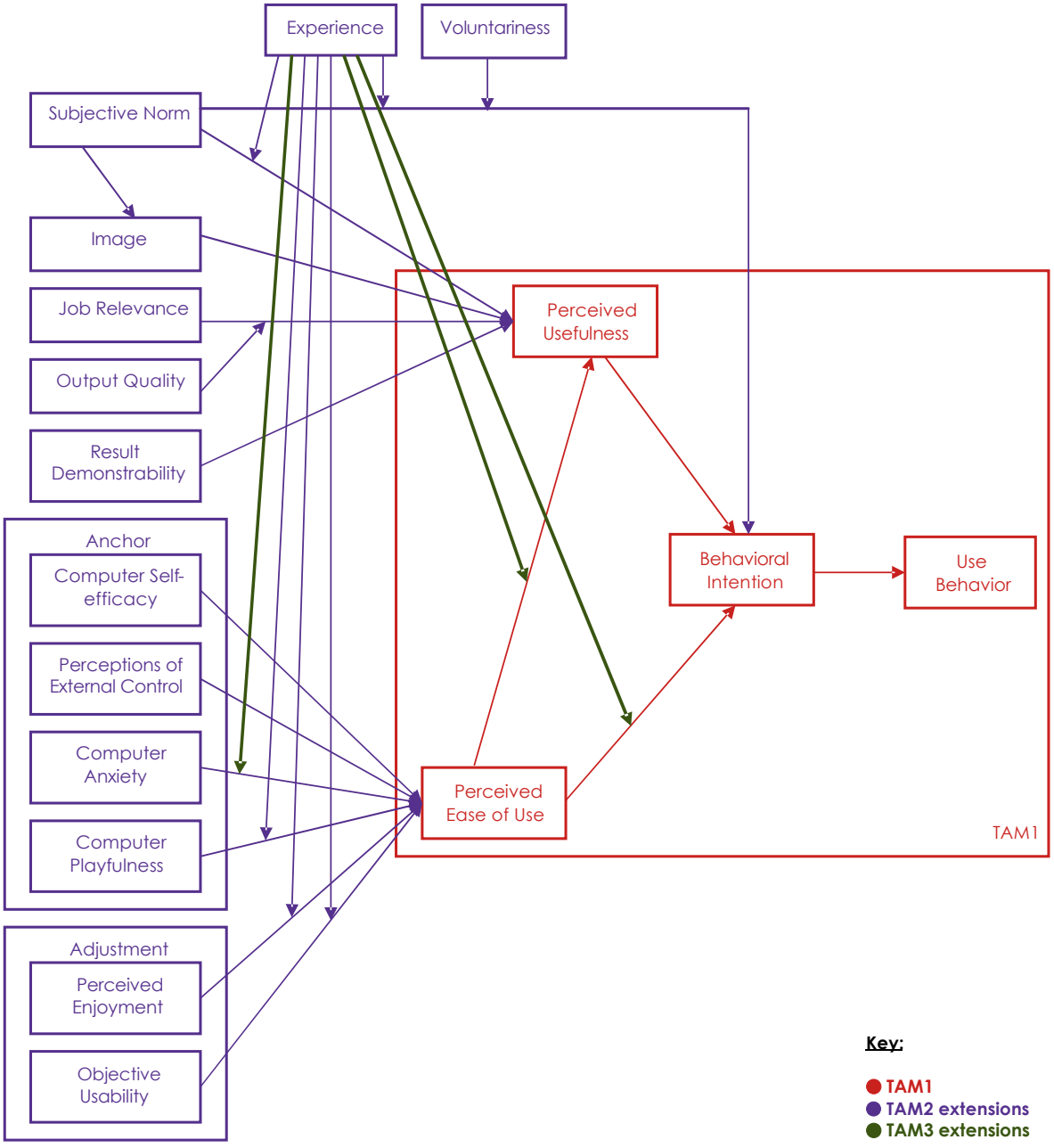


Figure 3.1: Comparison between TAM1, TAM2 (Venkatesh and Davis, 2000), and TAM3 (Venkatesh and Bala, 2008) models.

The most major difference between TAM1 and TAM3 is the removal of *attitude* from the models. This 'disconnects TAM from the theories of reasoned action and planned behavior (sic)' (Hornbæk & Hertzum, 2017, p. 33:22) on which it was constructed and removes the link to attitude as a fundamental element of the construct to explain use of information systems.

3.2.3.3. Applicability

Even though TAM is used in many studies, the literature often focuses on applying the model to specific technologies without considerations of the possibility of generalisation across different categories of information systems (Hassenzahl & Monk, 2010). Coupled with the changing focus from single-user computing to multi-user international and inter-connected always-on devices, TAM has become a dominant paradigm incorporating many formerly disparate areas of IS research into consistent knowledge about a small area of the domain (Benbasat & Barki, 2007, p. 214).

3.2.3.4. TAM and the web

In order to understand the acceptance of a specific website, Castañeda et al. (2007) developed a derivation of the TAM model for online use. This demonstrated the importance of perceived usefulness as the main determinant in the intention to continue visiting a website, with a higher value placed on this aspect by more experienced users (Castañeda et al., 2007, p. 392). Newer users seek novelty, resulting in a need for organisations to develop designs with a permanent focus on the specific use case of the site (Castañeda et al., 2007, p. 393).

Despite not being a major element of TAM, literature using the model to study the world wide web have demonstrated that information quality is the highest predictor of usefulness (Lederer et al., 2000).

3.2.3.5. Conclusion

The technology acceptance model has been widely used to explain users' behaviour when electing to use individual information systems and measure the aspects that make widespread acceptance more likely.

Whilst TAM can be used to explain the uptake of websites, it cannot be applied to individual informational pages where users may be forced to access information. For example, information published by a government

body on their website would only be available authoritatively from that site: users do not have a choice where to seek answers to their questions. Thus, acceptance of use is not an appropriate concern as there is a forced usage pattern.

In addition, the TAM model provides a generalised approach without considering the contextualisation of application. The general gap in the research identified by this study represents such a refinement to limited elements of the model: the study of online content quality requires a context-specific perspective that may suggest enhancements that focus on a small number of inherent properties. This does not encompass the holistic overview of what defines website content success.

The DeLone-McLean model discussed in the next section is concerned with the overall success of an information system and how this relates to users achieving their goals.

3.2.4. DeLone-McLean model

The DeLone-McLean model of information systems success defines the impact of a technological artefact on individuals and organisations, based on three aspects of quality: system, information, and service (DeLone & McLean, 1992, updated 2003). Broadly, these can be considered as corresponding to the three areas of literature drawn upon in this study:

1. **System quality**
   As this aspect focuses on the provision of the system itself, it can be considered to encompass the techniques provided by HCI.

2. **Information quality**
   This aspect focuses on the content provided by the system.

3. **Service quality**

   The overall quality of the service is parallel to the industry idea of UX (User eXperience), that describes the entire journey of a user through a system.

The model is a generalistic approach to encapsulating the key factors in the success of information systems incorporating many different aspects. Due to its use of dimensions and potential for contextualisation, it is suitable for re-use in usability and user-centred design studies (Hellstén & Markova, 2006, p. 4). The DeLone-McLean model is routinely applied at individualistic levels, with only occasional applications within an organisational space (Urbach et al., 2009, p. 321).

Information systems research often focuses on rigour, leading to the lack of practitioner focus within the literature. Through their works, Rosemann & Vessey (2005, 2008) demonstrate that the DeLone-McLean model could be immediately applied to industry contexts with sufficient explanation of the constructs involved.

3.2.4.1. Evolution of the model

The original DeLone-McLean model (1992) was developed based on the parallels between information systems and communications systems (Mardiana et al., 2015, p. 172), building on the theories of Shannon (1948), Shannon & Weaver (1949) and Mason (1978). The models have become two of the most cited literature within IS (Lowry et al., 2007, p. 150), demonstrating their widespread application by researchers within the field.

The major two modifications to the original DeLone-McLean model (1992) are the introduction of service quality and net benefits as additional constructs (DeLone & McLean, 2003, pp. 18–19). The former follows Pitt et al.'s (1995, p. 173) observation that most IS measurements focus on products and not the overall service to users, therefore failing to fully capture system effectiveness. The concept of net benefits captures the growing number of information

systems impact measures, reflecting the contextualisation of the model within research studies.

3.2.4.2. Dominant influences on systems success

In their review study of literature based on the application of the DeLone-McLean model to individuals, Petter et al. (2008) identified the strongest influences on information systems success within the model (see Figure 3.2). This demonstrates that the dominant factor in producing net benefits for individuals is user satisfaction - that is, positive user satisfaction will lead to greater success of the system. In addition, the study demonstrated that the two main factors in producing user satisfaction are the quality of the system and the quality of information within the system.

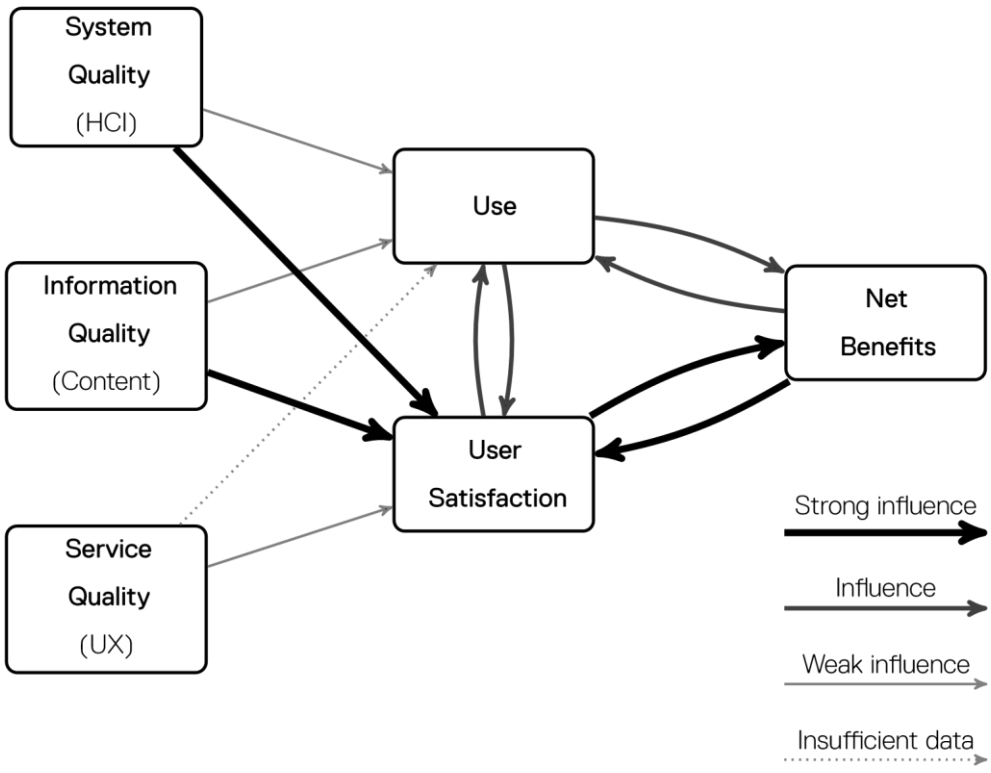


Figure 3.2: DeLone-McLean model (1992, 2003) of information systems success, highlighting the strongly correlated influences as identified by Petter et al. (2008).

3.2.4.3. Information quality

The first evolution of the model categorised information quality as a purely semantic property of information systems success. This is in contrast to earlier works that focused on the quality of output produced rather than the information inherent within the system (DeLone & McLean, 1992, pp. 62–64). A lack of information quality can lead to high organisational costs and disrupted operations due to not meeting user needs (Swanson, 1997), leading to the potential for a commercial entity to lose customers (Gorla et al., 2010, p. 215). Improving system quality can help improve users' perceptions of information quality, although this may not be possible depending on the scale of graphical and online processing features available to the system developers (Gorla et al., 2010, p. 222; Nelson et al., 2005).

Using the research questions defined on p. 4, it would be tempting for the research to focus solely on the information quality aspect of the model for this project. Gable (1996) cautions against this style of approach as single constructs cannot be utilised as a measure of success. Over the wealth of literature using this model, many studies report mixed results due to their lack of holistic approach and failure to account for the complete collection of success dimensions (Gable et al., 2008, p. 380).

Many researchers have successfully applied the DeLone-McLean model to a variety of contexts, and generally found a positive and significant relationship between information quality and intention to use (Halawi et al., 2008; Petter et al., 2008; Rai et al., 2002). This has led to researchers such as Iivari (2005) and Wu & Wang (2006) conducting studies to ascertain users' intention to use by measuring perceived information quality using Likert scale-based questionnaires. Many studies have successfully used subjective survey techniques to measure user satisfaction as a proxy for information systems success (Petter et al., 2008, p. 256), although researchers cannot rely on self-reported measures as fully representative of system use Heo & Han (2003).

Petter et al. (2013, pp. 38–39) note that whilst there have been many studies examining information quality as an independent variable, none have identified strongly correlated antecedent variables, demonstrating its importance in systems success as a separate concept.

3.2.4.4. User satisfaction

User satisfaction is strongly related to the perceived usefulness of an information system, to an extent that when a user perceives a system as providing more value, they are more likely to be satisfied with its use (Mawhinney & Lederer, 1990). This has an impact on the information quality, as it demonstrates the importance of user perceptions over empirical measures. Petter et al. (2008, p. 241) note that whilst information quality is inherent in many user satisfaction instruments, it is rarely distinguished as a separate construct, making measurement difficult within IS studies.

The relationship between information quality and user satisfaction has been studied within many applications of the DeLone-McLean model. There is a strongly positive relationship between the two aspects, strongly supporting the notion that higher information quality can lead to higher user satisfaction with the system (Al-Debei et al., 2013, p. 103; Halawi et al., 2008; Petter et al., 2008). This provides a demonstrable link between the two aspects, as well as a basis for measuring information quality through user satisfaction. User satisfaction with a system has a high significance with intention to use, demonstrating that a system is more likely to be successful if it meets user requirements (Petter & McLean, 2009, p. 164).

3.2.4.5. Comparison

The previous sections have discussed the DeLone-McLean model of systems success and related theories. Selecting the correct theory for IS studies can be a challenging process, especially as this can divert researchers away from key themes and research objectives due to the pre-defined constructs and potential for lack of contextual awareness (Benbasat & Barki, 2007, p. 212).

This can lead to an abundance of literature forming an impression of 'accumulation of knowledge' without recognising the developments or providing actionable advice whilst enforcing a cumulative research tradition (Sabherwal et al., 2006; Benbasat & Barki, 2007, p. 213; Gable et al., 2008, p. 379). In addition, it cannot be assumed that 'more use of an information system will lead to higher performance' (Goodhue, 2007, p. 220).

3.2.4.6. Conclusion

Despite their similarities and popular use in information systems studies, TAM and the DeLone-McLean model differ in one key aspect: the focus of measurement. TAM aims to measure acceptance by users, i.e., how likely is it that a system will be adopted. In contrast, the DeLone-McLean model measures empirical success of a system, making it more applicable to online systems that users cannot choose whether or not to use and websites where there are variable levels of success. This is important to the philosophical approach of this research, as the questions aim to discover insights into user perceptions on the ease of use, not just the ability to complete a given task.

3.2.5. Summary

This section has discussed the key information systems success theories that are relevant to this project. There are many existing models of acceptance and success of web systems based on the two major models: TAM and DeLone-McLean. Within this research, the DeLone-McLean model is a better fit due to its focus on success of systems rather than acceptance in use, as users are often forced to use the types of websites to be examined in this study. In addition, information quality is a key element of the DeLone-McLean model, providing a wealth of literature on its impact within information systems.

The next section of this chapter examines the state of online quality within both academic and industry domains, identifying the key standards and guidelines that aid practitioners in the delivery of online content.

3.3. Information seeking behaviours

With the growth of information systems for organisational and business purposes in the 1960s–1980s, many researchers studied the provision of information services, focusing on the availability of texts. This resulted in the development of several models for information seeking within institutional contexts (Ingwersen & Järvelin, 2006, p. 55).

The turning point for information seeking behaviour research occurred during the 1980s when there was a shift towards empirical studies focused on users identifying their behaviours rather than solely focusing on their need for information (Ingwersen & Järvelin, 2006, p. 56). This change in research track occurred in tandem with the shift in information systems use; PCs becoming commonplace in offices and affordable computer equipment enabling digital transformation.

By focusing on the behaviour of users, many of the information seeking models of this era incorporate ideas and theories from cognitive psychology combined with sociological studies. This shift to a focus on achieving the individual's desired outcomes has continued with contemporaneous empirical studies examining web use across a wide range of devices and contexts.

3.3.1. Problem solving

Despite the variability in information seeking models, a common shared trait is the focus on the user as a problem solver. This informs model development to focus on how a user can transition from a problem or question to an appropriate solution. This is clearly demonstrated by Dervin & Nilan's (1986) *Sense-Making Approach* as illustrated in Figure 3.3.

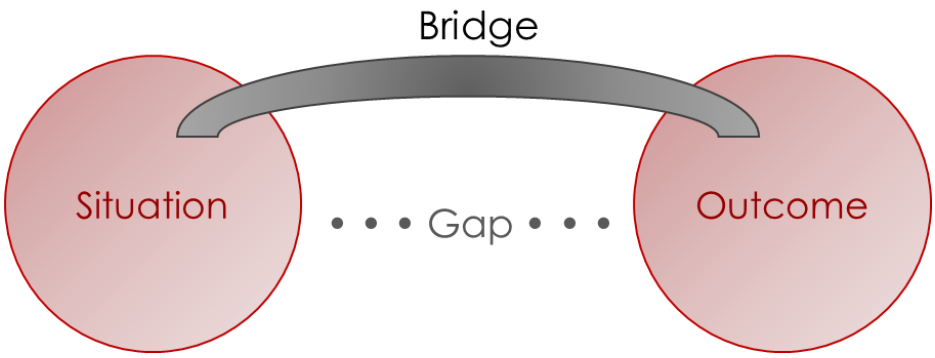


Figure 3.3: Dervin & Nilan's (1986) Sense-Making Approach (as illustrated by the researcher).

The Sense-Making Approach enables researchers to analyse and understand how the user makes the transition from situation to outcome. This is often bridged across a gap in knowledge, analogous with a problem-solving approach.

Problem solving approaches are ubiquitous throughout information seeking behaviour models as this explains both the reason for using for the system and the result of the user's activity. Information is meaningless without a context and purpose, both of which are provided by bridging the gap in knowledge.

3.3.2. Information retrieval

The subfield of information retrieval (IR) has a similar history to information seeking behaviours. The systems-oriented approaches of the 1960s-1980s provided many different models examining quantitative retrieval processes with small-scale empirical studies identifying organisational requirements for successful access to known information snippets (Ingwersen & Järvelin, 2006, p. 111). With an increase in computing power, the ubiquitous spread of the web and a demand for systems to provide answers in a more natural human-like manner, the 1990s and early 2000s provided an opportunity for researchers to diversify the subfield whilst still maintaining a cognitive user and task focus.

Information retrieval research follows a mechanised approach, pertaining to 'the representation, storage, organization of, and access to information items such as … Web pages' (Baeza-Yates & Ribeiro-Neto, 1999). This is a broad focus without consideration of the inherent quality of the information contained within a document.

Despite the lack of focus, content quality is a concern for traditional applications of information retrieval such as search engines. When seeking information, users are often seeking accurate and reliable sources, a challenging outcome given the mechanised processes of IR (Pokorny, 2004, p. 46). This is considered to be a key challenge for the future of information retrieval where many systems do not have automated means of checking the correctness of supplied documents (Ranpara & Kumbharana, 2021, p. 270).

3.3.3. Summary

The quality of documents contained within information systems is a consistent challenge in both the subfields of information seeking behaviours and information retrieval. Whilst there have been attempts to automate and mechanise the evaluation of quality, these have focused on quantitative processes rather than the subjective elements required to understand user perceptions of content.

The next section discusses online content quality, exploring the shift brought on by the increase in information available through the web.

3.4. Online content quality

Despite the breadth of information quality literature, online content quality has remained a 'vastly undefined concept' (Aladwani & Palvia, 2002, p. 468), leading to widespread fragmentation focusing on contextual subsets of critical factors (Aladwani & Palvia, 2002, p. 469). Furthermore, Rieh (2002, p. 145) suggests that quality has become a more important factor to users

with the growth of the Internet. With the increased amount of information available to users through a largely uncontrolled environment, it has become more difficult for individuals to perceive the informativeness of the content they are accessing.

This section discusses several key areas of research which contribute to the collection of knowledge surrounding content quality within both online and offline contexts.

3.4.1. Website success

The quality of content can have a positive impact on website success (Liu & Arnett, 2000), and there are many established measures in industry to help evaluate this. These range from usability guidelines (such as ISO 13407 (1999) and ISO TR 16982 (2002)) to best practice documents (such as WCAG2.0 (W3C, 2008) and Walker et al. (2013)). These provide varying amounts of guidance to content authors, however there is no definitive answer as to what contributes to 'good' content, and how organisations can improve quality. Rather there is a focus on the way the content is delivered and the structure of the system in which it is contained, leaving specific content creation to non-practitioners who may not have the necessary skills to write for the web. Within the academic literature, this covers three key areas: information quality, usability, and user satisfaction.

Kang & Kim (2006, p. 1188) first considered website content as a separate factor within website success identifying the lack of study on non-interactional elements. They divide content into two different forms: for information and for entertainment purposes. This highlights the differences between types of websites, with informational resources treated separately to those which are intended to be fun to explore. The study provides a path for future research to determine the unique factors in informational resources, whilst also exploring the relationship between *quantity* and *quality* of website content.

3.4.2. What makes 'good' content?

The discussion in Chapter 2 demonstrated that writing 'good' online content is not an easy process: the effectiveness of a website depends on many contextual factors such as the user, their organisation, technical constraints, and the task to be achieved. As such the question of *what makes good content* is abridged and requires a more complex answer.

Moving into the online world has compounded the problem of producing content that is effective for its use, with fewer constraints on the audience, time, and place of use, leaving user behavioural patterns less predictable than for prior media forms.

In order to analyse content for its effectiveness, a number of dimensions or *attributes* need to be considered. The next section will discuss existing frameworks for information quality analysis that may be applicable to online content.

3.4.3. User perceptions

Thielsch & Moshagen (2015, p. 4) provide a model for the incorporation of user perceptions into websites as shown in Figure 3.4.

Figure 3.4: Website perceptions and antecedent variables (Thielsch & Moshagen, 2015, p. 4)

The research demonstrates that user perception of web content is based on several interrelated factors. It can depend on external factors such as the

user's age or technological proficiency, the web browser or computing environment. Internal factors include the content itself, its usability/navigability, and the visual appearance. These three internal factors are parallel to the primary qualities considered in the DeLone-McLean model (see p. 78).

When studying the effectiveness of online content, studies should account for all of the antecedent factors in addition to the content itself. This is the positioning of most studies with HCI, which examine either the technical aspects (e.g., device context of use), usability (whether through accessibility or analysis of page elements) or aesthetics (e.g., design aspects). The missing link within website studies is the inherent quality of the content itself. This study focuses on this aspect to produce guidelines for improving information contained within websites, notwithstanding due consideration for interlinked aspects such as accessibility, navigational patterns, and design choices.

3.4.4. Frameworks

Many frameworks exist for quantitative information quality assessment within the fields of management and information technology. In their review study, Eppler & Wittig (2000) evaluated 20 such frameworks, and discovered that many are specific to their particular domain without providing generalisable results. This is to be expected, as user perceptions of information is contextual, dependent on situational norms and difficult to repurpose within different systems. There is also little focus in the frameworks on how to measure and analyse the quality aspects, focusing on the existence or absence of individual criteria.

A similar approach is taken by Thielsch & Hirschfeld (2019) who evaluated a subset of quality frameworks within literature to determine the attributes or *facets* with the largest impact on user perceptions of a website. The researchers followed a quantitative approach, though future research

avenues include the exploration of facets identified as subjective and dependent on the context of the user.

In their empirically derived model, Wang & Strong (2006) categorise the characteristics of information quality into four areas, as shown in Table 3.4.

Intrinsic	Believability, accuracy, and reputation of the data, as perceived by the user.
Contextual	Relevancy, timeliness, and completeness of the data within the task context.
Representation	Format and meaning of the data.
Accessibility	Ease of user access to the data.

Table 3.4: Four categorisations of information quality characteristics (Wang & Strong, 2006).

This conceptual framework attempts to reduce the number of disparate areas for which researchers need to analyse, however their study notes that many of the characteristics are non-exclusive and may be contextually appropriate to be included in several different areas of the framework (Wang & Strong, 1996, p. 19). This highlights the challenge of providing a 'one size fits all' tool for application for multiple domains without specialisations for the particular challenges. Considering the context of data is an important distinction between data and information; information theory suggests that meaning can only be derived when value is added to the user's knowledge, and this can only be understood for known situations.

Following Wang & Strong's (1996) seminal work, researchers have defined and created numerous frameworks and collections of attributes for data analysis within information systems. These are reviewed further in Chapter 4.

3.4.5. Extending industry models

In their study of online applications, Olsina et al. (2009) propose an extension of the ISO9126 (2001) software model to incorporate content quality (also termed *infoquality*) as an additional feature (see Figure 3.5).

Figure 3.5: Extended ISO quality model incorporating content quality (Olsina et al., 2009, p. 8).

The addition of the *content quality* attributes of accuracy, suitability, and accessibility address some of the differences between data and information, ensuring that the data has purpose for the intended user's activity. Whilst infoquality has interrelated dependencies with many of the existing attributes, adding a single pillar produces a stepwise enhancement for practitioners without the need to understand and comprehend new models.

This recognises the lack of content-related attributes within ISO standards in general, resulting from the prior focus on technical aspects of quality. Many of the other frameworks produced for information quality purposes lack consideration of web development lifecycle (Olsina et al., 2009, pp. 13–14) and are not compatible with existing ISO models (Lew et al., 2010, p. 219), whilst this model is an extension of existing practice.

A further approach is taken by Rafique et al. (2012, p. 571) who extend the existing ISO 25012 data model with two new characteristics, *value added* and *representational adequacy*. This relates data to information via the context of use with the observation that quality can only be maintained where data is processed into something meaningful.

Extending industry standards in this manner enables the research to build upon prior studies, incorporating stepwise improvements into existing

validated models. Enhancements such as those discussed in this section have been used to transform data quality models into those for information by considering the extra value of content within systems.

3.4.6. Summary

Despite the extensive literature analysing data and information quality using fully quantitative approaches, the quality of online content and analysis using qualitative techniques have not been fully explored. Where there have been attempts to produce models for content quality, these have generally involved the specification of new frameworks that do not build on existing practitioner approaches. This research project aims to create a tool to aid practitioners when improving the quality of online content, whilst relating the results to relevant academic theories.

3.5. Conclusion

This chapter has explored the key definitions of information, information quality and informativeness, in addition to the reasons why this presents an opportunity for further study. These themes can be explored using one of three existing academic approaches: information systems acceptance (i.e., TAM), information systems success (i.e., DeLone-McLean's model), or information seeking behaviours. For the purposes of analysing information quality within informative web systems, the DeLone-McLean model presents the best opportunity for alignment with existing academic theory as users can choose where to receive their information, with the key qualities of a website (system, information, and service) separable according to existing practitioner conventions and nomenclature. The development of 'good' content is essential for the delivery of information through web-based information systems.

Through the link to the DeLone-McLean model, this research will provide additional insights into one of the key aspects of the academic theory, contextualising information quality advice for the consideration of online

content. This situates this study within a large body of prior information systems literature, extending existing theories within a new domain with its own research challenges.

The following chapter explores information quality frameworks: collections of attributes used to assess and evaluate online information sources. This leads to the selection of appropriate attributes to form the basis for the case studies.

Chapter 4. Information quality frameworks

Within the last twenty-five years, there have been many attempts to provide frameworks to aid researchers in quantitatively analysing, evaluating, and assessing information quality across different datasets. This is an important development for positivistic studies as assigning relative values of success to areas of quality demonstrates strengths and weaknesses within the information provided.

This chapter provides an overview of key information quality assessment frameworks and their relevance to this project.

4.1. Introduction

Since the growth of the Internet in the late 1990s, there have been various attempts to define frameworks for aiding the development of quality online information. Many of these focus on the specification and numeric evaluation of dimension-based metrics to provide qualitative metrics to suggest areas that researchers can focus on to deliver benefits to information.

4.2. Categorisation of frameworks

Pipino et al. (2002) categorise information quality assessment frameworks into two categories: *objective* and *subjective*. Objective assessments are based on the use of software rules measuring the quality of information, broadly applying a positivistic philosophy (see p. 133). Subjective assessments incorporate user opinions to evaluate the perception of the information by consumers, typically incorporating surveys and/or interviews following an interpretivistic paradigm (Caballero et al., 2007; Price et al., 2008; Ge et al., 2011).

To identify the perspectives of existing information quality frameworks, the remainder of this section presents an overview of the key frameworks proposed by authors within the field of Information Systems for generalised systems. Following this section, domain-specific frameworks will be reviewed in Section 4.4.

4.3. General IQ frameworks

With an increasing focus on generalised and user-accessible information systems within the 1990s, a renewed focus on data quality from consumer perspectives led to Wang & Strong's (1996) conceptual framework that underpins many later works. This paper contained an initial focus on four key quality groups:

1. **Accessibility**
   Data must be accessible to the consumer.

2. **Interpretability**
   Data must be able to be interpreted by the consumer.

3. **Relevance**
   Data must be relevant and timely for the consumer's decision-making process.

4. **Accuracy**
   Data must be correct, objective and from reputable sources.

Using a survey-based approach, the authors evaluated user perspectives of data quality, resulting in a hierarchical collection of 20 attributes for objective evaluation. These original attributes underpin much of the later information quality research, with the additional or removal of attributes demonstrated within specific domains.

A key challenge of works based on Wang & Strong (1996) is the original paper's focus on *data quality*. As discussed in Section 3.1.1, data and information are often used synonymously yet are separate concepts.

Ackoff's (1999, pp. 170–172) definition that data are raw symbols representing object properties, combined with Checkland's (1981) definition of information as processed data with an intended purpose demonstrates that these terms should be considered as discrete constructs. The criteria that may make data useful may not be identical to that which provides quality information, often resulting in a mismatch between the consideration of attributes within the literature. Furthermore, this demonstrates that information quality can only be considered within the context of a domain or purpose, as information cannot exist without understanding its intended use.

Many researchers have attempted to improve on the work of Wang & Strong (1996) with frameworks considering alternate research perspectives. The remainder of this section provides an overview and narrative of the development of general academic frameworks within information systems.

4.3.1. Web Assessment Index

The Web Assessment Index (WAI) was an early attempt at creating a general-purpose device for analysing and comparing information quality within websites. By providing a systematic score out of 100, the WAI provides a measure of website success at meeting user requirements, identifying weaknesses within the information provided (Miranda González & Bañegil Palacios, 2004, p. 314) and is based on an earlier analysis of University websites (Buenadicha Mateos et al., 2001).

The WAI provides assessments within four key categories as shown in Table 4.1. These categories provide a holistic measure of website success.

Accessibility (20%)	**Speed (10%)**	**Contents quality (50%)**	**Navigability (20%)**
Presence in search engines (5%) Popularity (15%)	Site size (bytes) (10%)	Informational content (30%) Communicational content (20%)	Site map (15%) Keyword search function (5%)

Table 4.1: Web assessment index metrics (based on Miranda González & Bañegil Palacios, 2004, pp. 318, 324).

Although several of the measures are focused on metrics which are no longer important in the modern web (e.g., sitemaps or presence in sitemaps), the broad categories remain as important to users seeking information within modern websites. The authors highlighted the need for positive content quality within websites by assigning the category 50% of the overall score, the largest component of which is informative content.

The original validation of WAI identified two relationships as the most important within websites: that of navigability/content, and accessibility/content (Miranda González & Bañegil Palacios, 2004, p. 326). This highlights that the user's ability to navigate around the content and access the information they need can be key factors contributing to website success. Navigation is discussed as an information quality attribute in Section 7.3.5.

4.3.2. TDQM

Total Data Quality Management (TDQM) is a theoretically-grounded methodology for delivering information products to consumers (Wang, 1998, p. 58). Developed at MIT, this framework draws parallels with Total Quality Management (TQM) practices from product manufacturing in turning raw inputs into processed outputs (see Figure 4.1), adding more detail to the transition described in Ackoff's (1999, pp. 170–172) works. This draws a clear parallel with information as a commodity (see p. 65), produced from raw data being processed within an information system.

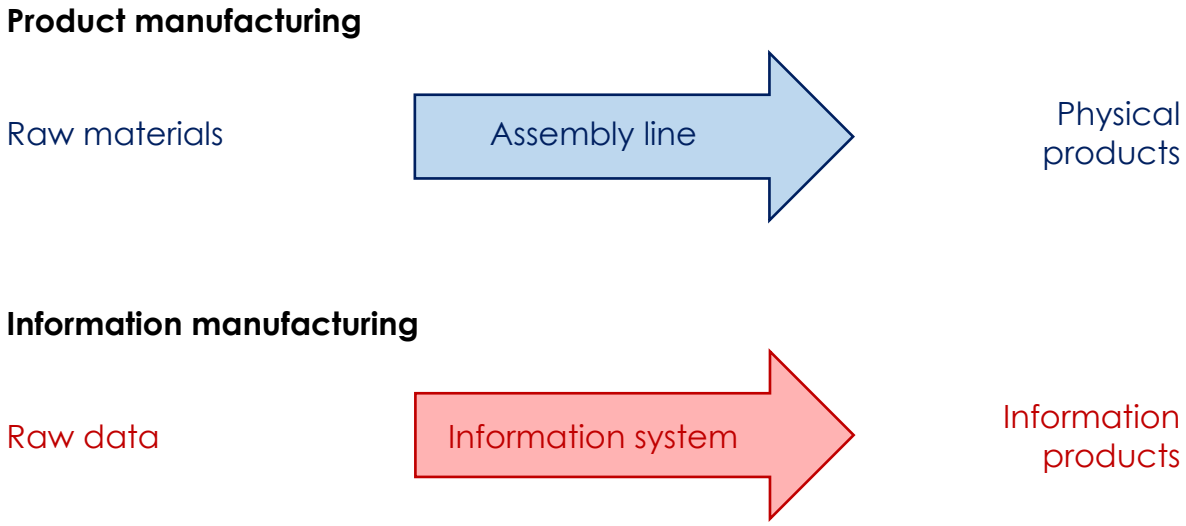


Figure 4.1: Total data quality management pipeline as a parallel to product manufacturing (based on Wang, 1998, p. 59).

TDQM is an iterative methodology to identify requirements, measure characteristics, improve quality and demonstrate benefits to the information. This approach helps to address issues outside of the system that could have an impact on information quality, including production errors, technical storage or access issues and evolving consumer needs (Strong et al., 1997, p. 39). A key aspect of this methodology is the need for long-term commitment to follow the principles as part of institutional practice (Funk et al., 1998, p. 4).

4.3.3. WebQual

The WebQual framework was first developed by Barnes & Vidgen (2000) to identify which qualities are expected of websites by users, identifying means for improvement through quantitative questionnaire analysis. The initial study displayed a number of websites to participants, asking for their ranking on 24 criteria simplified into eight subcategories. Of these, the most relevant to this study is the information content, incorporating questions on IQ attributes such as relevancy, ease of use, and appropriate level of detail. With evaluation performed using quantitative methods based on participant rankings, finer grained understanding of the quality of content is not possible within the model.

Over the subsequent two decades, several refinements to the model have been made to reflect the evolution of both academic theory and web technologies, resulting in WebQual 4.0. The Information Quality dimension has been expanded to include the timeliness of information, appropriate formatting, and reputation (i.e., authoritativeness) of the source (Loiacono et al., 2007). The framework has subsequently been applied in a range of domains with high informational content such as higher education, governmental, and health services.

4.3.4. PSP/IQ

The PSP/IQ model builds on TDQM to define the difference between product quality and service quality (see Table 4.2). In terms of the updated DeLone-McLean model (see p. 80), the system and information aspects have been conflated to form a single category: product. This approach allows for a more generalised model; however, it obfuscates the delineation between the external and internal boundaries of the product.

	Conforms to specifications	**Meets or exceeds consumer expectation**
Product quality	Sound information	Useful information
	The characteristics of the information supplied meet IQ standards.	The information supplied meets information consumer task needs.
	Relevant IQ dimensions: Free of error Concise representation Completeness Consistent representation	Relevant IQ dimensions: Appropriate amount Relevancy Understandability Interpretability Objectivity
Service quality	Dependable information	Usable information
	The process of converting data into information meets standards.	The process of converting data into information exceeds information consumer needs.
	Relevant IQ dimensions: Timeliness Security	Relevant IQ dimensions: Believability Accessibility Ease of operation Reputation

Table 4.2: PSP/IQ model of product and service quality
(Kahn et al, 2002, p.185; Lee et al., 2002).

Quality has a dual definition of fitness for use and the extent to which a product successfully serves the purposes of consumers (Kahn et al., 2002, p. 185). This focuses the model on two traditionally separate concepts: whether the data is what is specified, and whether this specification is suitable for the users. The duality of these concepts means that the model can analyse the difference between producing information with users in mind or not considered in the process, separating two important aspects of models such as the DeLone-McLean model of information systems success (see p. 78).

The two levels of the model represent different aspects of information quality; early literature in the field focused on a product-orientated approach (Kahn et al., 2002, p. 186) for expert, trained users with dedicated terminals.. This matches with information systems transition away from full-service products

(such as those provided within organisations) to services provided for access by users both outside and inside the organisation. Much of this transition followed the development and spread of the web to a wider audience, providing the opportunity for those outside of organisational boundaries to access and explore information. A service-based approach matches the understanding that customers require information from an organisation, and a website is one means of meeting this need.

PSP/IQ provides a quantitative and data-driven framework for information quality benchmarking, demonstrating how researchers can compare levels of quality across substantially different datasets using common attributes. This allows for a generalised approach, with the researcher selecting the appropriate measures for each attribute.

4.3.5. IQ measurement

Eppler & Muenzenmayer's (2002, p. 192) information quality measurement framework takes a more technical approach to analysing websites. Following the Plan-Do-Check-Act cycle in TDQM (see p. 96), this defines an iterative process for using qualitative data gained from five different tools as shown in Table 4.3.

#	Type of tool	Data collection type
1	Performance monitoring	Passive
2	Site analyser	Active
3	Traffic analyser	Passive
4	Web mining	Active
5	User feedback	Active

Table 4.3: Five types of web-based tools for qualitative measurement of website performance (based on Eppler & Muenzenmayer, 2002, p. 189)

Performance monitoring and traffic analysers collect passive data as users navigate a website without the need to conduct any extra activity. These are sources of data that already exist due to the architecture of web infrastructure and can be interrogated to provide metrics to help with understanding a page's performance.

Site analysers and web mining take a more active approach, evaluating a website by parsing content and performing live analysis based on predetermined metrics to simulate use of a website. These can give a good indication of the technical performance of the infrastructure yet will rarely relate to individual use of the site.

Finally, user feedback can be used to collect information directly from users to inform future development activities. These are often elicited using numerical data and Likert scales to facilitate metrics-based evaluation and graphical representation of results (Eppler & Muenzenmayer, 2002, p. 191). This may support the business improvement processes; however, it does not deliver the feedback required to evaluate the user's perception of and use of the data.

Eppler & Muenzenmayer's (2002) framework takes a technical approach to evaluating website performance, with little focus on the content on the website, thus its suitability for general purpose applications. As discussed in Section 3.2.4, the information within a web page is often overlooked, with focus being placed on the system and service quality, i.e., the HCI and UX aspects. This approach can determine some missing attributes of the data, however more evidence would be needed to suggest improvements to the content of a site.

4.3.6. AIMQ

The AIMQ methodology is designed to evaluate and improve the quality of a system using questionnaires and features a three-step process (Lee et al., 2002; Vaziri & Mohsenzadeh, 2012, pp. 57–58), identifying the effectiveness of multiple data quality attributes within controlled study contexts.

The first activity of AIMQ is a questionnaire to measure the dimensions of information quality using the PSP/IQ model. Each quadrant receives a score based on the component dimensions, allowing for data quality comparisons between organisations.

Following the questionnaire, a matrix is completed to identify the gap between intended properties of the data and user perceptions of those attributes.

The third component of AIMQ is the conducting of a gap analysis, identifying where improvements can be made.

AIMQ is a comprehensive methodology, however it uses a pre-determined list of attributes for the analysis of data quality. This does not permit the questionnaire to be easily contextualised to the environments that have developed since the original paper was published (Vaziri & Mohsenzadeh, 2012, p. 59). The individual questions in the questionnaire are based on the opinions of the researchers involved in the development of the tool without a thorough grounding in academic and industry-based research (Vaziri & Mohsenzadeh, 2012, p. 60). As a framework designed for improvement of existing systems, AIMQ provides little guidance on how to make systems more effective, leaving the identification of weaknesses and appropriate actions to the researcher.

4.3.7. IQIP

The analysis of information quality for search engine results is another area explored by some researchers. The IQIP model provides a framework for assessing information quality, by applying the results to help rank content within web search results (Knight & Burn, 2005).

The unique feature of this framework is the focus on three dimensions: user, environment, and task. This makes the process highly context-specific, with assessment attributes chosen dependent on these dimensions (Knight & Burn, 2005, p. 168). Whilst this allows IQIP to be refined for different online contexts, the authors make no attempt at defining which attributes are appropriate for which environments, reducing the generalisability and repeatability of projects applying this approach.

The results of the assessment process are metrics-based, enabling the quantitative comparison of information sources. This will produce data to inform computerised search results, however qualitative aspects of the user perceptions are not considered.

4.3.8. Ge & Helfert

In their review of information quality research, Ge & Helfert (2007, p. 3) identify a common set of components for assessment across many works in the literature (see Figure 4.2). This framework encompasses the three main elements of the existing literature: methodologies, dimensions, and metrics.

Figure 4.2: Framework for IQ assessment (Ge & Helfert, 2007, p. 3).

This conceptual framework provides a generalised approach to understanding IQ assessment in any domain or context, facilitating the researcher's selection of the appropriate dimensions and methodologies for the information being studied.

Through their literature review, Ge & Helfert (2007, p. 6) suggest 22 common attributes analysed in information quality papers, categorised within six dimensions (see Figure 4.3). Whilst this provides a good starting point for generalised IQ assessment, many of the attributes are not relevant to the content within an online website, focusing on the system properties instead (Bovee et al., 2003; Helfert, 2001; Kahn et al., 2002; Naumann & Rolker, 2000; Wand & Wang, 1996; Wang & Strong, 1996).

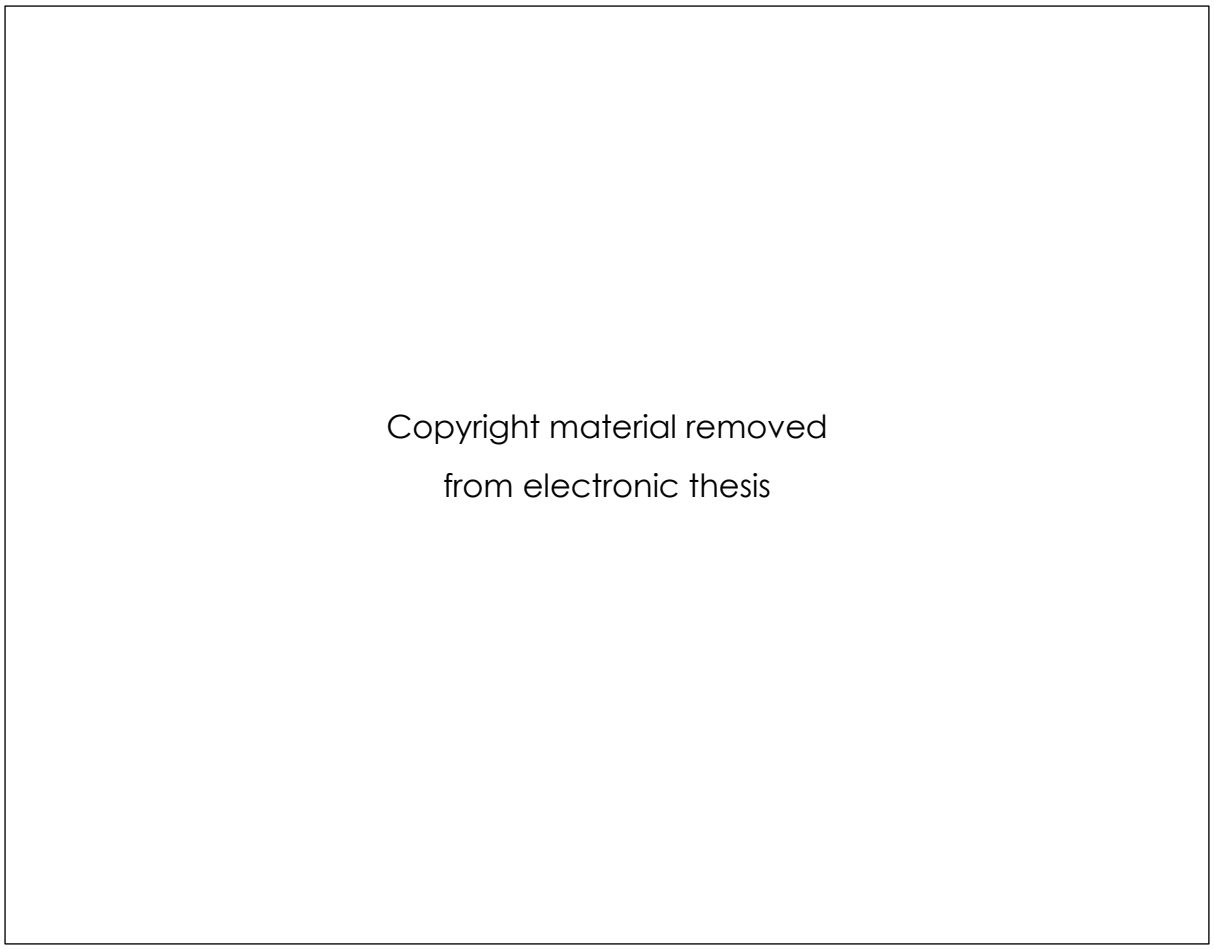


Figure 4.3: Ge & Helfert's (2007, p. 6) categorisation of IQ attributes, with dimensions relevant to this study highlighted.

A later iteration of Ge et al.'s (2011) model incorporates objective and subjective assessment approaches by determining the actor in the IQ process (see Figure 4.4). This aids the researchers in understanding the fitness to purpose, as purposes may be different for different actors.

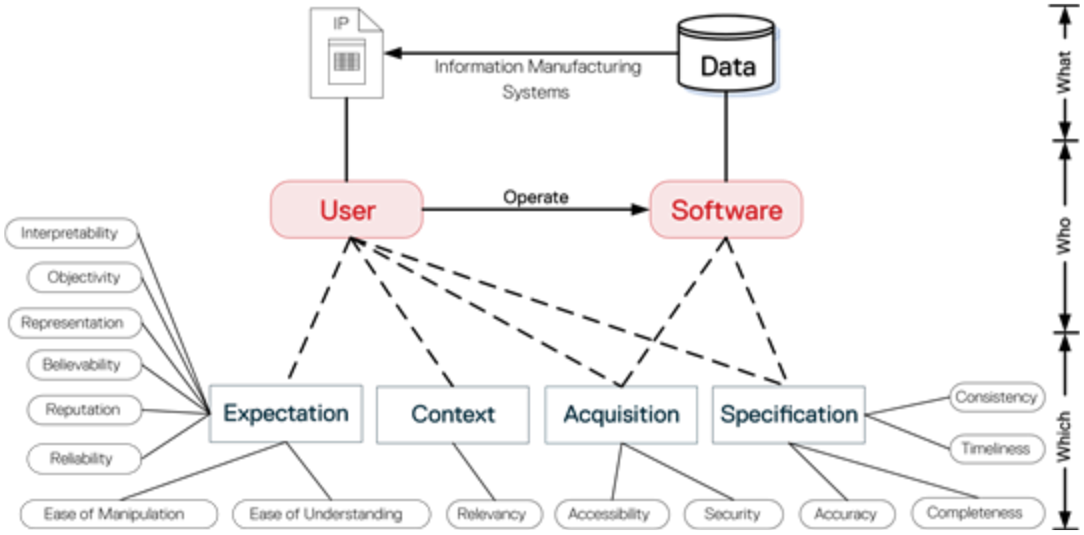


Figure 4.4: Ge et al.'s (2011) information quality assessment framework.

This framework also distinguishes between raw data and information quality, refining the dimensions that may apply to the assessment process. The dimensions are drawn from a frequency-based analysis of prior works following a systematised literature review.

The model provides a comprehensive framework for establishing IQ assessment methodologies, incorporating the objective/subjective and data/information dichotomies that have become apparent within the literature. The authors suggest that the model has been designed to support conversion to a practitioner tool that could be applied within business contexts.

4.3.9. Web-CLIC

As an alternative to the quantitative approach favoured by the above frameworks, Web-CLIC intends to assess the *subjective* user perceptions of web content in accordance with four key facets: clarity, likeability, informativeness, and credibility (Thielsch & Hirschfeld, 2019). Amongst these four facets, credibility was demonstrated to have the strongest impact on perceived informativeness of a web page (ibid, p. 16).

The primary research instrument is a 12-question Likert questionnaire based in eliciting user perceptions after exploring a case website. The authors note that the subjective character of the results must be considered: high informativeness scores can be due to perceived value of the information and not only the provided text.

The primary Web-CLIC study evaluated the approach against 60 websites from different domains, showing generalised validity even though this may not mean universal applicability within any web context.

As an emerging instrument for the evaluation of web content, Web-CLIC provides a useful starting point for considering development of the framework within this research study, especially the demonstration of how user input can feed into the understanding of informativeness of content provided by a website.

4.3.10. Comparison

This section has summarised many of the suggested information quality assessment and improvement frameworks within the prior literature.

Each of the frameworks reviewed in this section can be considered through six comparative criteria:

1. **Purpose**

   The general intention of the research framework, including domain. This

may have an impact on its applicability to web spaces as some frameworks are specialised for a chosen context.

2. **Attributes**
   The number of quality attributes considered. This provides a measure of the complexity of the framework as those with larger number of attributes will be more challenging to apply.

3. **Dimensions**
   The number of overall dimensions, i.e., groups of attributes used to simplify the problem space. Dimensions can be considered in isolation or in combination with the rest of the framework.

4. **Iterative**
   Indicating whether analysis of information quality requires the framework to be applied iteratively or once.

5. **Context aware**
   Whether the framework is applied differently based on the research context.

6. **Online specific**
   Whether the framework is specifically intended for use online or is more generalised quality research.

Table 4.4 below shows a summary of the nine information quality attribute frameworks reviewed in this section, identifying the six criteria.

§	Name	Year	Purpose	Attributes	Dimensions	Iterative	Context aware	Online specific
4.3.1	WAI	2001	General purpose website evaluation	20	4	No	No	Yes
4.3.2	TDQM	1998	Information as products	15	4	Yes, PDCA	No	No
4.3.3	WebQual	2000	Website evaluation	24	8	No	No	Yes
4.3.4	PSP/IQ	2002	Information as services	16	4	No	No	No
4.3.5	IQ measurement	2002	Website analysis	16	–	No	No	Yes
4.3.6	AIMQ	2002	Gap analysis	15	4	No	No	No
4.3.7	IQIP	2005	Web crawlers	–	3	Yes, PDCA	Yes	Yes
4.3.8	Ge & Helfert	2007	Organisational	41	9	No	Yes	No
4.3.9	Web-CLIC	2018	Subjective evaluation	12	4	No	No	Yes

Table 4.4: Comparison of generalised IQ frameworks.

As can be seen from the table, most frameworks were derived during the early 2000s when the new opportunities presented by online information

became a focus for IQ researchers. Of these, only two consider an iterative approach (following the Plan-Do-Check-Act) as a key element of their application. Four frameworks can be considered as specialised for the online context; the other five are suitable for application though the attributes and dimensions are not specifically tailored for web-based content.

4.3.10.1. Recency

The development of generic information quality frameworks peaked during the mid-2000s, a point at which the web had become sufficiently developed to provide searchable information, with search engines such as Google becoming predominantly used as a starting point. This matches with the use of the term information quality in literature, as shown in Figure 4.5.

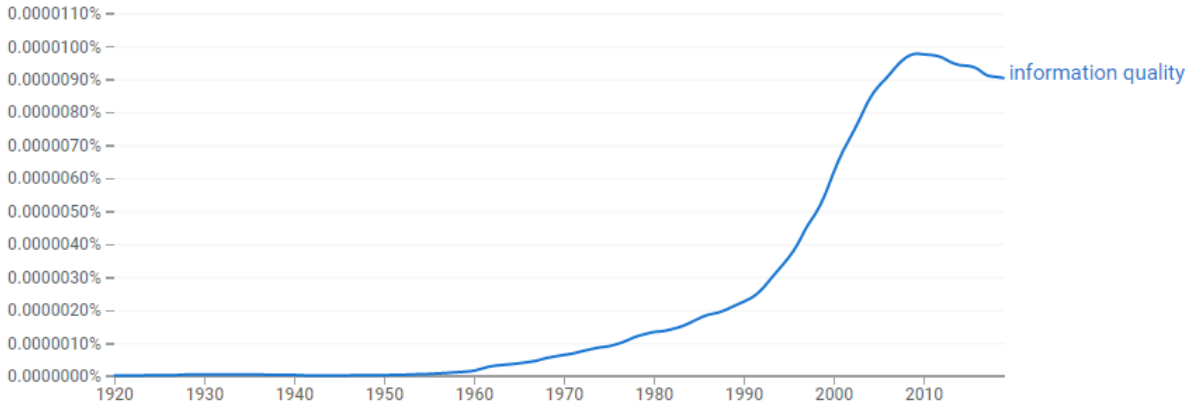


Figure 4.5: Google Ngram for the term *information quality*, highlighting incidence within published works (Google, 2021).

As can be seen in the graph above, information quality was first explored as a concept with the development of 1950s information systems. With an increasing reliance on the processing of data to provide insights, the proportion of published works using this term increased to a peak in 2009. This corresponds with the widespread use of online search engines, content management systems, and users more adept at finding information online.

The development of generic IQ frameworks began with Wang & Strong's (1996) focus on data quality, leading to several alternative re-definitions of

their original attribute set. Except for Web-CLIC, all the generic frameworks were developed before the peak in IQ literature, following the same positivistic approach to analysis.

Despite the lack of development of generic frameworks, researchers in specific fields focused on the development of specialised attribute sets for their domains. Three such examples of fields with their own activity include e-commerce, e-government, and e-health, both of which are further explored in Section 4.4.

4.3.10.2. Multiple attributes

The frameworks discussed above generally consist of multiple attributes for information quality analysis. Hernández et al. (2009, p. 363) identified this as a trend for detailed website assessment with general purpose scales consisting of multiple criteria and others containing fewer factors specialising on more criteria. Most frameworks are general purpose with little research dedicated to comparative studies between sectors. Research such as that conducted by Kandari et al. (2011a, 2011b) moves towards identifying the differences between attributes in different sectors, with little impact observed on the informativeness of content despite their perceived differences.

4.3.10.3. Alternative analyses

Two attempts have been made in the literature to categorise and synthesise generic information quality attributes. The first of these was performed by Parker et al. (2006) with the aim of returning online information appropriate to a user's context. As a literature-based study, the authors perform a frequency-based analysis of 13 contemporaneous data and information quality frameworks to find the most common attributes. The authors note that *accessibility* and *timeliness* are the most frequently defined attributes in the IQ frameworks studied, though no attribute appears ubiquitously in every framework.

The frequency-based approach cannot provide a complete set of attributes as the nomenclature varies between frameworks with no conversion of context. Some attributes such as speed and timeliness, or accuracy and completeness, are linked concepts, with one attribute potentially informing the other, so cannot be studied in isolation (see Section 7.2.2).

Kandari et al. (2011a, 2011b) follow the same approach, extending the frequency analysis to cover 20 frameworks from the literature. Whilst this provides an overview of the 23 attributes discovered, the same inherent issue with related attributes prevents the sole use of a frequency-based approach.

This study introduced a second-stage approach to the identification of inherent information quality attributes using a small focus group of five graduate students. Whilst this provided a non-representative sample of users or practitioners, the impact of individual related or previously unexplored themes was incorporated by the researchers providing a final set of 22 attributes relevant for web-based content. This results in a quantitative approach to analysing websites in a variety of domains.

As this research study intends to produce a practitioner model for qualitative analysis, a renewed approach to important attributes needs to be undertaken to ensure that the guidance provided to inform the development of informative content is effective and appropriate for industry use.

4.3.10.4. Domain-specific frameworks

The generic frameworks in this section provide a useful starting point for identifying data and information quality attributes as they can be applied to general-purpose content and contexts. To compare attribute sets with those used within domain-specific frameworks, three key areas (e-commerce, e-government, and e-health) are explored below in Section 4.4.

4.3.11. Summary

The major challenge in taking a positivistic approach to information quality evaluation is the definition of suitable metrics. By defining assessment metrics within the framework, researchers are contextualising the heuristics and ensuring that their methodologies only work within specific contexts, even if this is not explicitly specified. The use of quality dimensions deconstructs the challenge of incorporating many aspects into a single model, providing a method for analysing, evaluating, and improving individual characteristics of the information within a website. Although none of the models discussed within this section are fully compatible with the focus on online content desired in this study, the approach of selecting attributes to describe aspects of the information provides the basis of a methodology for breaking down the different user requirements for website informativeness. This approach will be followed in this study to incorporate research in many different aspects of content to provide an overall tool for practitioner use.

The next section describes domain-specific IQ frameworks derived from those discussed during this section for specific applications. These facilitate a focused approach, highlighting the attributes demonstrated to be important within these fields.

4.4. Domain-specific IQ frameworks

Since the first applications of information quality within an online context, there have been many attempts to generate domain-specific frameworks. These works recognise the limitations of all-encompassing checklists of features that may not be applicable for all contexts.

Many information systems challenges are common across different domains; however, practitioners may perceive their challenges to be unique. Whilst general theories may be applicable within an academic context, there have been many instances where a domain-specific IQ framework is tailored to a

specific context. Practitioners may believe the narrower scope of attributes to better fit the content being provided (as with e-health frameworks below).

This section discusses key frameworks within three domains: e-commerce, e-government, and health.

4.4.1. E-commerce

Early uses of the web were limited to business-to-business transactions, with sales completed through limited, expert-targeted systems for purposes such as holiday bookings in the early 1980s (Feldman, 2011, p. 59). This was followed in 1984 by the first consumer-focused e-commerce platform enabling online home shopping with a range of stores including Tesco groceries, Lloyds Pharmacies, and Greggs bakeries. Since these early experiments, e-commerce in the UK has grown to the point where more than £1 out of every £10 of Gross Domestic Product (GDP) is spent online (Lone et al., 2021, p. 43).

Concurrently with the development of generic IQ frameworks, several researchers recognised the need to improve the quality of e-commerce websites, especially due to the strong correlation between the provision of quality content and consequent sales to consumers. The main four IQ frameworks defined and applied within e-commerce are discussed below.

4.4.1.1. SITEQUAL

The SITEQUAL framework was developed in 2001 to evaluate and compare Internet shopping website performance (Yoo & Donthu, 2001). Grouped into four dimensions (ease of use, design, speed, and security), the nine dimensions are evaluated through a survey-based approach based on subjective questions. The instrument was designed to perform an evaluation of user browsing behaviour and its link to site performance. As an early measure for site performance, the authors acknowledge the need for further development of the framework.

4.4.1.2. PIRQUAL

As an alternative approach, the PIRQUAL questionnaire examines consumer expectations and perceptions by providing a survey for a representative sample of website users (Francis & White, 2002). Organising the questions into six dimensions (functionality, product descriptions, ownership, delivery, customer service, and security) this framework provides a holistic overview of the online shopping experience beyond the website. This could be useful from a service evaluation perspective although the quality of the information is only covered by the first three dimensions.

4.4.1.3. eTailQ

The eTailQ framework moves away from performance analysis towards assessment of quality of e-commerce websites (Wolfinbarger & Gilly, 2003). Through an analysis of existing frameworks, key attributes of e-commerce website quality are selected for evaluation in a three-stage study: focus group, sorting exercise, and online surveys. A combination of the results produces a framework based on four dimensions (see Figure 4.6).

This framework demonstrates the same holistic approach as PIRQUAL, with the information quality focused on the first dimension. Separating the users' perceptions of the information quality from other aspects of the site is challenging given the potential interdependencies not explored in the framework.

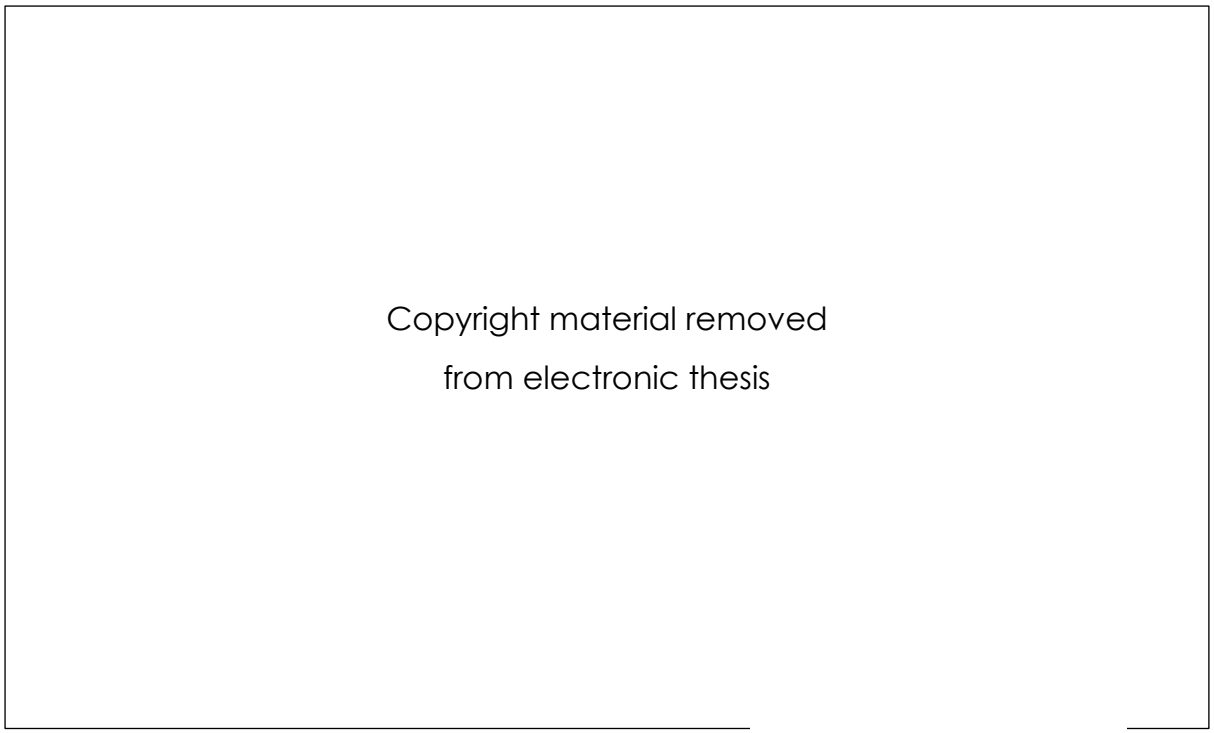


Figure 4.6: eTailQ dimensions of e-commerce website quality (Wolfinbarger & Gilly, 2003, p. 193).

4.4.1.4. eTransQual

The eTransQual framework (Bauer et al., 2006) moves beyond goal-orientated behaviours to focus on a transactional model of e-commerce quality with a four stage of process. Of these, the first (information) is most relevant to online content, where key considerations include functionality, accessibility, efficiency of navigation, content, design, and enjoyment. These begin to explore aspects of information quality such as clarity, accessibility, and relevance of information though only as part of the entire purchasing process.

4.4.1.5. Discussion

The general trend within e-commerce research after the early 2000s has been away from fixed frameworks of attributes towards an analysis of user perceptions using information systems success models. This has largely been

due to the need to have an overall view of the customer journey including all touch points, with a recognition that no one aspect of the site can be a panacea for increasing online sales.

Despite this trend towards information systems models, the analysis of attributes within e-commerce website case studies can still be relevant for analysing online content. The key challenge for an information quality researcher is the separation of attributes between what are the informative aspects of the content and which elements of the site are not.

4.4.1.6. Summary

With the huge growth of e-commerce during the early 2000s, researchers responded with a range of frameworks to both allow academic analysis and provide practitioners with advice on which aspects of sites could be improved to drive sales. The quality of information within websites has been repeatedly demonstrated to have a positive impact on overall success, though the quantity of informative content (as opposed to marketing materials) has not been factored into any of the prior research studies. The definitions of content, information, and information quality explored in Chapter 3 demonstrates the need for clarity of nomenclature and clear delineation between informative and non-informative content.

4.4.2. E-government

Within the academic literature, there are many different definitions of e-government centred around the common theme of using information technologies such as the Internet 'to improve the delivery of government services to citizens, businesses, and other government agencies' (Palvia & Sharma, 2007, p. 2). This wide-ranging group of definitions relates the provision of web-based information by government organisations to a varied audience that includes several different types of actors. Understanding the information required by a user can help inform the challenge of providing the e-government content expected by the modern citizen.

The study of governmental services is a classic application of information systems theory as there are often complex organisation structures that require an understanding of the problems and user needs. To aid in the study of online information quality, four key frameworks have been identified from the literature.

4.4.2.1. E-GovQual

Recognising the need for a specialised website evaluation framework, the authors of E-GovQual developed a framework with thirteen dimensions of quality from a citizens' perspective, including content as a key area (Ataloglou & Economides, 2009). Of the 100 criteria, 9 relate to content quality including attributes such as accuracy, completeness, and uniqueness. The researchers identified the need for some websites to contain larger volumes of organised content depending on the intended audience whilst not necessarily rating highly on presentational aspects. The need for informative websites to convey large volumes of information clearly to a wide range of users is one of the key drivers of this research project.

4.4.2.2. Summary

Whilst much of the e-government research is based on earlier frameworks such as SERVQUAL, the application of models such as DeLone-McLean's Information Systems success are growing in prevalence (Nkanata, 2019), reflecting the recognition that user perceptions can have a large impact on the use of a system, and a key aspect of that success can be attributed to information quality. The lack of a contemporaneous model for information quality in e-government websites reflects the division between academic and industry developments in this domain, with services such as the UK's Government Digital Service (GDS) aiming advice and research at practitioners rather than contributing to IS theory. The model proposed for development in this thesis could reduce the divide by producing a method for information quality evaluation equally applicable to both governmental and non-governmental informative websites.

4.4.3. E-health

The development of information quality frameworks targeted at healthcare websites is a recent phenomenon, with a renewed focus on content provided by medical practitioners. This section reviews three contemporary approaches to analysing the quality of e-health websites.

4.4.3.1. What are e-health websites?

Prior to analysing the frameworks, it is important to define the scope of e-health websites.

As a term, e-health has been used since 1999 as a generic name for the interaction between healthcare domains and computer-based systems (Eysenbach, 2001) with the intention to digitally enhance the doctor/patient relationship (Ball & Lillis, 2001).

A key consideration of online health information is the incorporation of ethics in the form of accurate and trustworthy content (Rippen & Risk, 2000). This continues to be a challenge on the modern web, as demonstrated by the frameworks discussed later in this section.

Within a UK context, the largest health information provider is the NHS. The NHS app has seen a rapid uptake in both downloads and usage patterns, with 2 in 5 English citizens installing on their mobile device for content access (NHS Digital, 2021). The continuing digital transformation of NHS services is driven by NHS X, providing frameworks and advice for the provision of quality online platforms.

The provision of online health information is currently disparate, varies between locales and does not provide a consistency of experience. Health information systems researchers have proposed several IQ frameworks to aid consumers and practitioners in the identification of quality content, with the major developments outlined below.

4.4.3.2. Early frameworks

Following the development of Wang & Strong's (1996) data quality framework, several health researchers developed domain-specific tools during the late 1990s.

The JAMA (Silberg et al., 1997) and HON (Health On the Net; (Boyer et al., 1998) frameworks were devised to provide accreditation or approval to sites complying with between 4–8 key criteria, including attribution, transparency and disclosure of ownership. As basic sets of criteria, these aimed to measure 'good quality' with the former requiring satisfaction of ¾ of criteria and the latter 100% compliance.

An alternative approach is provided by Charnock et al.'s (1999) DISCERN instrument, which devises a set of 16 Likert scale questions for analysis by health information consumers, resulting in an overall numerical score for all attributes.

Despite their domain-specialised approach, such metrics-based frameworks do not identify the nuances of content provided to users. For example, specific information may be embedded within a web page though not easily recognisable, accessible, or readable by the typical information consumer.

The frameworks discussed below evolve the approach of these early models, incorporating additional dimensions and analyses whilst still incorporating awareness of the medical context.

4.4.3.3. Health IQ

The Health IQ framework (Al-Jefri et al., 2018) builds on existing quantitative frameworks such as JANA, HON, and DISCERN to consider information quality within five specialised dimensions. In addition to the inclusion of health-specific information such as symptoms, treatments, and side effects, the framework also incorporates accountability and ethics as key considerations. This is demonstrated as an effective practice for health websites with users

ranking these attributes consistently highly, linking their impact to the user's perspective of the content provided.

As a framework for automated analysis of website content, Health IQ is intended to be used for quantitative studies with datasets, providing a metrics-based result for multiple cases. The attributes are mapped into overlapping sets as shown in Figure 4.7 below.

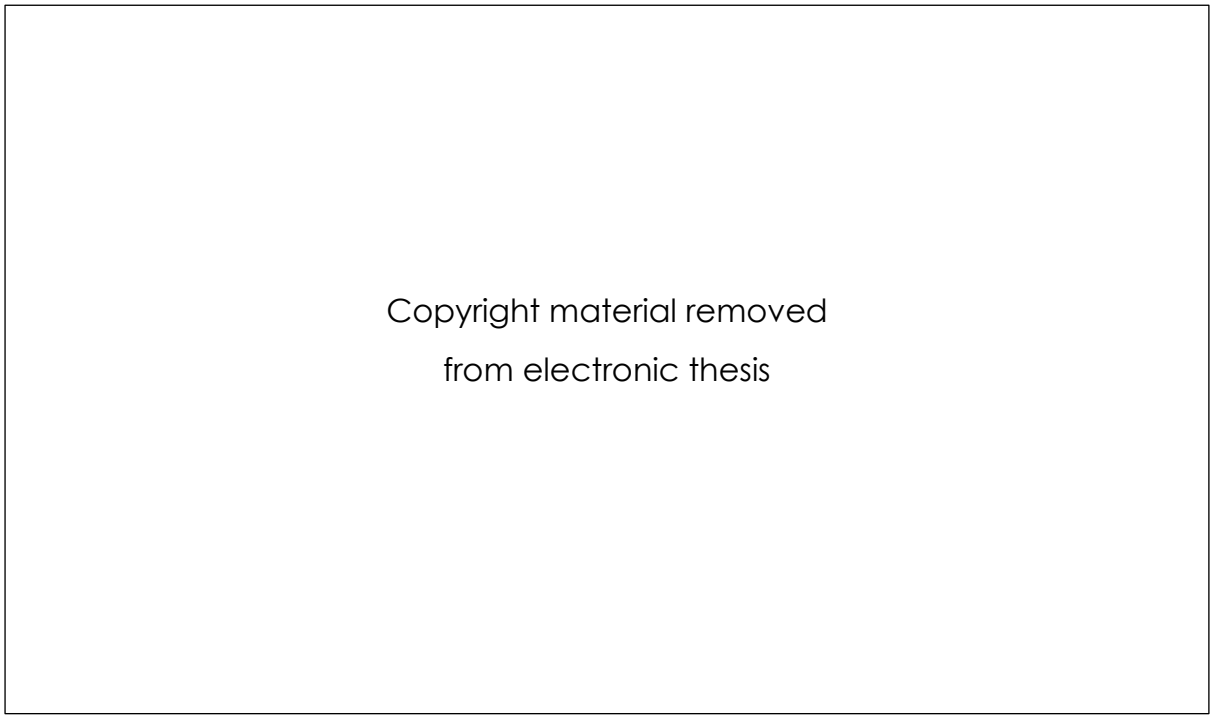


Figure 4.7: Classified mapping of attributes and dimensions within the Health IQ framework (Al-Jefri, 2019, p. 135).

This framework departs from the notion that each attribute is only applicable to one dimension, incorporating attributes into multiple sets based on potentially subjective analysis. Whilst this provides an academic model with impacts shared between different elements, the proposed automatic assessment of websites leads to a re-categorisation of the attributes within intrinsic and extrinsic sub-groups (Al-Jefri, 2019, p. 144).

4.4.3.4. FACILE

The FACILE model aims to introduce health empowerment by enabling a paradigm shift in the relationship between health practitioners and their clients by provision of online health materials (Alfano et al., 2021). This is achieved through automated recognition of key aspects of website quality (such as language complexity, customisation, reliability, and timeliness) and their consolidation into a zero to five-star rating.

For both non-expert and expert users of websites, the authors hypothesise that the most important attributes of information quality are reliability and timeliness, reflecting the early data quality frameworks discussed in Section 4.3. The automated approach facilitates integration of quality factors into search results, though does not incorporate further study of key online content quality metrics.

4.4.3.5. CLIQ

The Clinical Information Quality (CLIQ) framework is a proposed model in the health domain to evaluate online content (Fadahunsi et al., 2021). The model is derived from a systematised literature study, resulting in three dimensions with 13 attributes as shown in Figure 4.8.

Figure 4.8: The CLIQ model for clinical evaluation of online content quality (Fadahunsi et al., 2021, p. 7).

The CLIQ model is designed for clinical application, with key considerations drawn from both generalistic and domain-specific information quality studies. One of the three dimensions, *informativeness*, directly relates to the content quality with considerations such as accuracy, trustworthiness, relevance, and completeness. These are properties that can only be assessed by a practitioner as they do not account for user perceptions of the content.

Based on validated IQ frameworks, CLIQ presents the most complete domain-specific model within the literature. Although it has not been validated with user research studies, the strong basis of prior literature demonstrates the opportunity to generate new, specialised IQ frameworks based on considerations within the literature.

4.4.3.6. Summary

The evaluation of information quality within health websites is an emerging area, with several models proposed by clinical practitioners to combine and extend general IQ research. Of the frameworks reviewed in this section, CLIQ

presents the greatest opportunity to build upon prior research with the dimensions representing academic concepts of quality from prior literature.

The health IQ frameworks explored above consider several key contextualisations of content to this domain, identifying the presence of elements of information as predictors of information quality leading to user success.

4.4.4. Challenges of domain-specific frameworks

The works explored earlier in this section have evidenced three concurrent themes in domain-specific frameworks: specialised contextualisations, definitions of informative content, and a trend towards quantitative approaches. Each of these are discussed below.

4.4.4.1. Contextualisation

The purpose of domain-specialised frameworks is to provide contextualisation to the evaluation and guidance provided. This has been the goal of many researchers in different fields, recognising the chasms between each specialisation. For example, e-commerce websites are driving towards sales and e-health towards provision of accurate information. These disparate goals can be challenging to draw together, leading to researchers reducing the focus on those seen less important to their domain of study.

This research has been framed to focus on informative content; even though such a specialisation has not been previously explored, the emphasis on a particular type of online content follows the pattern of contextualisations followed by earlier literature.

4.4.4.2. Informative content

Whilst many of the works explored in this section have identified the important correlation between content and website success, there is little focus on information quality. Academic models such as DeLone-McLean (see

Section 3.2.4) highlight the important role of information quality in systems success, yet this crucial aspect is not a focus of domain-specialised frameworks.

4.4.4.3. Automated and quantitative approaches

The majority of the domain-specific frameworks reviewed above perform a quantitative analysis of results, often leading to suggestions of automation. Taking such a positivistic approach can often miss the nuances in user experiences, where subjective opinions of participants are required to be categorised for large-scale comparison. Whilst such studies have their place in the analysis of content quality, distilling the informative aspects can prove more challenging where conclusions are drawn based on statistical analysis.

4.4.5. Summary

Researchers within Information Systems and cognate fields are defining and refining several frameworks related to information quality, particularly related to the use of websites. Whilst many of these contain attributes that explore user perceptions of the site, little focus is made on the inherit quality of online content as opposed to properties of the system itself. The domain specific frameworks can be applied to improve the overall system quality.

The field of health IQ research, particularly applying quantitative techniques, is well developed within medical information contexts. Considering the mature stage of such models though the lack of qualitative analysis, this study will include a case study focused on the health content space (see Chapter 10) to demonstrate the generalisability and applicability of the practitioner model.

4.5. Conclusion

The development of information quality frameworks over the past 25 years has led to the proliferation of IQ attributes for both generic and specialised contexts. This chapter began by exploring general IQ frameworks, including

those based on Wang & Strong's (1996) attribute set for data quality, comparing approaches to analysing informative content. IQ Frameworks in specialised domains such as e-commerce, e-government and e-health were also discussed, with a comparison between the key features within each area. Information quality frameworks are further explored in the attribute selection study in Chapter 7, where the researcher identifies those that are most appropriate authoring informative online content.

The next chapter details the research philosophy leading to the methodological design of the studies in this research project.

Chapter 5. Research philosophy

This chapter provides an overview of the epistemological positioning of the research by examining the relevant information systems philosophical paradigms and justifying the selection of an approach for the study described in this report. The research philosophy is an important element of doctoral study as epistemologically-grounded research, developing new knowledge based on practice (Winter et al., 2000).

5.1. What is research?

Saunders et al. (2003, p. 3) suggests that research is 'something that people undertake in order to find out things in a systematic way, thereby increasing their knowledge'. This generic definition of research highlights two important factors in such activities:

1. The research must be based on demonstrable logic, not just the researcher's beliefs
2. The research must have a clear purpose and research questions.

The research described in this thesis satisfies both factors. The process and purpose are concurrent with prior literature and established methods (see Chapters 2–4), whilst the research questions have been clearly defined (see Section 1.2).

The next section contains a brief description of the information systems field, providing background context for the academic perspective applied to this report.

5.2. Information systems

Since its inception as a separate community in the 1970s, the field of information systems has developed in many different ways, and as such resulted in a diverse range of definitions of what is and is not part of the

domain (Goles & Hirschheim, 2000, p. 250; Robey, 1996, p. 400). Lee (2004, p. 10) states that 'the terms information, systems and information systems have fallen into such careless use that they seemingly no longer denote anything different from one another.' Benbasat & Zmud (2003, p. 186) describe this as an *identity crisis* and suggest that one of the dominant factors in any Information Systems research be the consideration of human behaviours in the use of technological artefacts. Davis & Olson (1985, p. 22) further argue that 'Information systems as an academic discipline is more an extension of organizational behaviour and management than computer science', focusing on social rather than technical perspectives. This leads to information systems as an applied rather than pure discipline (Baskerville & Wood-Harper, 1996, p. 235; Moody, 2000, p. 351; Weber, 1997), 'focusing on the application of information technology in practice rather than the technology itself' (Moody, 2000, p. 351).

This involves two primary objectives:

1. **Theoretical**
   Increasing knowledge by understanding why things happen.

2. **Practical**
   Improving practices to deliver social outcomes, for example finding better ways to develop information systems.

These two objectives highlight the need for practice to inform research, and research to inform practice, resulting in positive improvements and increased understanding (Avison et al., 1999). Bringing these two objectives together requires the researcher to accept that practice and theory are related and different forms of knowledge (Van de Ven, 2007). For example, practice-based knowledge is highly contextual and often contained in methods and standards, with research knowledge explicitly specified through publications (Mathiassen & Sandberg, 2013, p. 476). Checkland & Holwell (1997, p. 11) describe this relationship between theory and practice as a *learning cycle*

(see Figure 5.1), recognising that neither can exist in isolation and new ideas are often generated based on the practical experiences of the researchers.

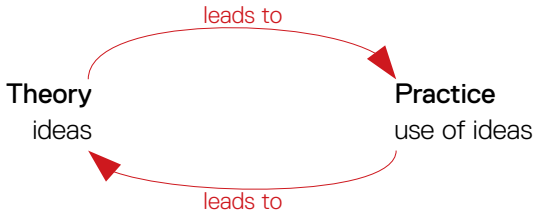


Figure 5.1: Theory and practice learning cycle (Checkland & Holwell, 1997, p. 11).

In the learning cycle, ideas or theories lead to their use in practice. By observing this, new theories and ideas can be identified. The cycle can continue in perpetuity, evolving the basis of knowledge.

Furthermore the need for clear boundaries within the field is highlighted by errors of inclusion, where information systems research attempts to answer questions better explained by constructs and methodologies outside the field (Benbasat & Zmud, 2003, p. 190).

Due to the diversity of the field, information systems research can be based on several philosophical perspectives. These will be explored in the next sections.

5.3. Epistemology

According to Babbie (2010), 'epistemology is the science of knowing'. This view is confirmed by Crotty (1998) who explains it to be 'how we know what we know'. Furthermore, Crotty (1998, p. 4) defines four elements crucial to successful research. The researcher has extended this model (see Figure 5.2) to include ontology and axiology as discrete elements, as well as data collection and generation methods at a separate level, highlighting the

importance of selecting appropriate methods for primary data collection during doctoral studies.

The continuation of this thesis mirrors these elements to provide a discourse on the theoretical perspectives and positioning of this research, along with a justification for the methods chosen to investigate the research questions.

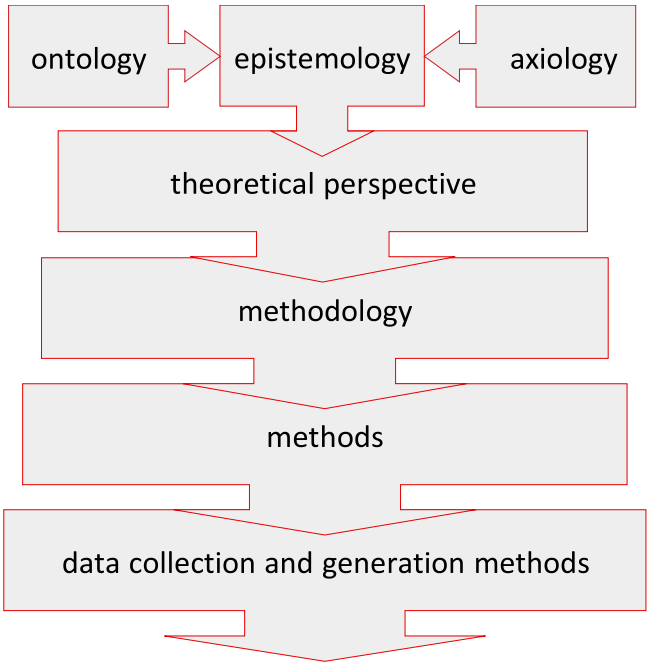


Figure 5.2: The seven elements of research, as extended from Crotty (1998, p.4) by the researcher.

Crotty (1998) categorises epistemologies into three core perspectives: objectivism, constructivism, and subjectivism. Each of these is described below.

5.3.1. Objectivism

Objectivism is the view that entities exist independently of experience, and have an absolute truth residing in them as objects (Crotty, 1998, p. 5). By conducting studies, the researcher can obtain the meaning of the objects as an absolute truth. Positivism (see Section 5.6.1) is largely grounded in this

area, and such research often focuses on quantitative methods and statistical analysis.

5.3.2. Constructivism

The constructivist perspective places the realities of the world as a construct of individuals' personal engagement. Meanings are personal constructs of the mind, produced within our consciousness to provide individualised understandings of the objects around us (Crotty, 1998). As constructivism explains truth as existing in individual realities, phenomena only exist as a social construct within the context of the problem space. Different groups of people can have different constructs, limiting the generalisability of results to the specific context being studied.

5.3.3. Subjectivism

Similar to constructivism, subjectivism explains knowledge as meaning applied by people to objects, however the subject may impose limitations on understanding objects. This means that meaning is independent of the object and only exists as assigned within the individual's mind (Crotty, 1998).

5.4. Ontology

Formally, ontology is considered to be 'a particular system of categories accounting for a certain vision of the world' (Guarino, as cited in Zúñiga, (2001, p. 194)), and from the researcher's point of view, describes the way the world works (Crotty, 1998). In the interdisciplinary field of information systems, there are two primary aspects: objectivism and subjectivism, that have been described in the previous section. The main difference between the two is the effect that the specific individuals in the context have on the situation. If the situation is entirely dependent on these people, then this is often considered as objectivism; whereas a subjectivist approach would consider the context as separate from the individuals in the study (Saunders et al., 2003).

5.5. Axiology

Axiology is concerned with the values and ethics that a researcher brings to the study, and considers the human nature that individuals possess during a research project (Creswell, 2008; Mertens, 2010). By clearly stating the researchers' perspective, their influence can be documented, explaining the potential impact on the results of the study. This is particularly important in qualitative research, where the researcher's values are visible in the approach to sourcing and coding data (Creswell, 2008).

5.6. Theoretical paradigms

Traditionally, the research methods used within information systems can be broadly categorised into two different theoretical paradigms: positivist and interpretivist (Braa & Vidgen, 1999, p. 2; Galliers & Land, 1987). Orlikowski & Baroudi (1991) and Goldkuhl (2012) expand on these two perspectives to include critical research and pragmatism as additional spaces applicable to Information Systems research. Each of these will be explored in this section.

5.6.1. Positivism

Positivism is the oldest and most widely applied paradigm in Information Systems (Oates, 2005, p. 283; Orlikowski & Baroudi, 1991). This paradigm assumes that the subject being studied can be observed objectively, with a focus on the repeatability of research, reductionism and refutability (Checkland, 1981). This view is confirmed by Oates (2005, p. 284), who states that the two main characteristics of positivist research are the assumptions that:

1. the world is ordered
2. the world can be studied objectively.

This leads to a position where research is reduced to a hypothesis that can be proven to be true for all situations at the moment, however proving falsehood just once can disprove a theory. Often, researchers consider positivist

research to follow a scientific method, with unbiased and repeatable experiments providing confirmation of the theory.

The patterns leading to the definition of a theory may be observable in practice, however they are often a construct of people. Without studying the people within a context, it may be difficult to fully comprehend the problem being investigated. Whilst making research problems easier to compute, positivism often dehumanises the context by removing the social element of the system being studied (Kreps, 2018, p. 8).

In addition, the principle of reductionism demands that the researcher can break complex things into smaller parts, and this is not always possible. For example, the bigger picture may be missed without a holistic view (Creswell, 2008, p. 7).

5.6.2. Interpretivism

Interpretivist research is used to understand the underlying meanings used by people to make sense of their own activity (Braa & Vidgen, 1999, p. 2; Orlikowski & Baroudi, 1991). As a paradigm, this contrasts with positivism to provide a space to explore context-specific questions 'through social constructions such as language, consciousness, shared meanings, documents, tools, and other artefacts' (Klein & Myers, 1999, p. 69). Interpretivist approaches involve explaining the relationships between different themes, requiring the researcher to interact with the human subjects of the study, perhaps changing the perceptions of both the researcher and the participants (Walsham, 1995, p. 376). To study the themes in a realistic environment, the researcher must explore the problem space outside of laboratories and controlled experiments, observing the actors within their natural setting. This may result in varying levels of agreement between study instances, leading researchers to focus their attention on the most convincing cases and the consideration of multiple interpretations within the context.

Within the IS field, there has been an increasing trend towards interpretive approaches, as demonstrated by the increase in journal articles following this

paradigm. This can be attributed to the increasing diversity of research, more mainstream adoption of non-positivistic studies and a shift in editorial policies (Mingers, 2001, p. 240).

5.6.2.1. Research questions

To define the boundaries of a study, interpretivist research tends to focus on research questions to be investigated within the problem space, as opposed to concrete assumptions to be tested. The results of the study depend on the interpretation of the data, a subjective matter leading to the importance of the researcher as an instrument (De Villiers, 2005, p. 2). Interpretivism only permits the researcher to view events through the people in the context, observing their direct experiences to understand the behaviour of the system being studied (Kreps, 2018, p. 15).

5.6.2.2. Principles for evaluation

Such field studies within IS require grounding within interpretive research methodology for both the active elements and post-evaluation. Klein & Myers (1999, p. 72) define seven key principles for the conduct and evaluation of IS research based on prior literature, each of which is explored below.

Principle 1. The Fundamental Principle of the Hermeneutic Circle

This principle is based on the concept that all human understanding can be understood by considering i) independent parts of knowledge and ii) the whole picture once the parts are combined. The relationship between parts and the whole is described in psychology as the Gestalt Theory, often characterised as the sum of parts providing deeper meaning than the whole (Koffka, 1935, p. 176).

The application of this within IS research can lead to iterative studies, where researchers explore individual fragments of data, followed by synthesis into a larger context. This builds on partial understanding to produce richer understanding of meaning to a system user.

The remaining six principles are based on the concept of parts and the whole, leading research to iterations of the *hermeneutic circle*.

Principle 2. The Principle of Contextualization

Research data cannot be viewed in isolation: rather the audience must be provided with historical and social context. Within IS research, this manifests as seeking meaning within the explored environment and justifying choices made by the researcher. Interpretive studies differ from positivistic approaches in this key area, where the former can explain phenomenon based on the evolution of theory with the latter based on the fundamental assumption that earlier data patterns will continue to be repeated in future studies (Klein & Myers, 1999, p. 73).

Principle 3. The Principle of Interaction Between the Researchers and the Subjects

The social construct of research studies can have an impact on the research data produced and must be carefully considered to ensure the effect of researcher decisions are accounted for within the context of the study. The effect of this within interpretive IS studies is often apparent in the reflexivity of participants, who often interpret their own context when providing commentary on their actions. Researchers must include this within the context of a research study, where preconceived notions may disguise unexpected and previously unexplored research outcomes.

Principle 4. The Principle of Abstraction and Generalization

Within IS research, observed phenomena should be explored in relation to existing philosophical and research theory whilst ensuring this is not based on assumptions such as universal natural laws of human nature or cultural norms. Studies should be related to broader abstractions whilst related to the field observations of the researcher. A common manifestation of this principle is

the grounding of research methodologies in existing theories of social construction.

Principle 5. The Principle of Dialogical Reasoning

Interpretivistic research studies can often uncover unexpected research outcomes based on prior assumptions or preconceptions of the researcher or participants. This principle suggests that research studies need to be conducted with minimised influence from prejudices, with the philosophical grounding clearly stated. Once previously unknown assumptions have been discovered, additional iterations of research should be conducted to remove and counter this new knowledge to provide unbiased outcomes. Reasoning should be explained within the overall historical and social context of the study.

Principle 6. The Principle of Multiple Interpretations

The reliance on research subjects for narratives can lead to differing interpretations of the same system or processes. These can often conflict, leading to a need to explore the social, power, economic and value influences, and context of these differences. This can lead to exploration of conflicting interpretations, providing heuristic value by deeper probing of system understanding.

Principle 7. The Principle of Suspicion

The final principle is that of suspicion, also known as *false preconceptions*. An unintended trait of interpretive research can be the discovery of false notions or socially created distortions. Such implicit misunderstandings can be discovered through analysis of participant narratives, identifying social and contextual constructs with a critical perspective.

Incorporating principles into study

Following the observation of the above seven principles, Klein & Myers (1999, pp. 78–79) argue that researchers should seek their own balance within research studies. The principles are described as guidelines for the collection, interpretation and evaluation of field study materials avoiding contextual or social biases. This leads to more generalisable and repeatable research outcomes, identifying multiple viewpoints which can be applied to future research contexts.

Within any interpretive research study, it will be important to incorporate all of the above seven principles, ensuring that data collection and evaluation activities provide a considered and contextually aware perspective of the research problem.

5.6.3. Critical research

Critical studies analyse and critique the current situation through the exposure of structural contradictions within social systems, giving the ability to transform these to improve implementation success (Orlikowski & Baroudi, 1991, p. 6). Whilst critical research is similar to interpretivist approaches, it does not attempt to distinguish between the researchers and other actors and lacks a solid grounding in evaluative approaches. Instead, evaluation occurs through self-consciousness, with a priority on the exposure of discrepancies and conflicts between the actors within the study.

With the focus on self-consciousness, critical research attempts to deliver a change to current practice through participatory approaches, helping the actors to free themselves from the structures they may find themselves constrained within (Creswell, 2008). Often, the outcome of a critical study is an action agenda, that can help participants gain their desired change through a collaborative approach. This is an alternative viewpoint to interpretivist studies, which aim to understand social structures rather than oppose them and bring about change.

Within the Information Systems field, critical perspectives are rarely used. The main area where critical research is conducted is in the Scandinavian participatory design culture, where studies aim to bring about change through 'workplace democracy' (Braa & Vidgen, 1999, p. 3).

5.6.4. Pragmatism

Pragmatism is a paradigm where researchers have freedom of method selection to use whatever approaches they think are necessary to resolve the research questions (Creswell, 2008, pp. 10–11), and has derived from the works of many authors in American, European and Asian thinking (Goldkuhl, 2012, p. 7). This paradigm draws the researcher to a mixed methods methodology, with a realisation that a truth can only be defined at the present instance, as results may not be generalisable across time or other contexts. The perspective suggests that there is an external world as well as that within the mind, although researchers should stop asking questions about reality and merely record the observable phenomena.

Taking a pragmatist approach may be aligned with industry practices, however it is not suitable for the completion of a PhD. The results produced by this project must be applicable outside the research context and for others in similar situations in order to create an impact, and this approach would not allow for thorough academic referencing. Few works within information systems take a pragmatist approach, perhaps due to the difficulty of building on the existing body of literature.

Goldkuhl (2012, p. 9) identified three types of pragmatism that are applicable to information systems research:

1. **Functional pragmatism**
   Knowledge is the basis of action and has direct influence on local practices. In addition, any knowledge gained from the activity must be generalisable to provide for both local and general contexts.

2. **Referential pragmatism**
   Actions become the subject of research, with knowledge gained

about those actions. As things and events exist independently of any observers, this position can be considered as taking a middle position between positivism and interpretivism (Goles & Hirschheim, 2000).

3. **Methodological pragmatism**
   The focus of the research is on the selection of methods appropriate to the current context, rather than subscribing to a pre-determined methodology. This permits a researcher to openly explore the problem space with a pluralistic mixed methods approach.

Despite its description as a wholly separate theoretical paradigm, there are suggestions that pragmatist thinking has been implicitly embedded in many types of interpretivist information systems research, particularly action research approaches (Baskerville & Myers, 2004; Goldkuhl, 2012, p. 10). Where action research aims to create new local knowledge based on participatory principles, the methodology can be considered to contain pragmatist practices. However pragmatism is rarely explicitly selected as a paradigm by information systems researchers (Goldkuhl, 2012; J. Mingers, 2001).

5.6.5. Summary

As a researcher, it is important to approach the selection of paradigm objectively based on the research context, research questions and traditions of the discipline (Oates, 2005, p. 304). The four main paradigms applied within information systems have been outlined above and represent the beliefs among researchers within the field. A summary of these paradigms is shown in Table 5.1 below.

Referring to the research questions (see p. 13), they tend to be more aligned with an interpretivist approach. This is particularly notable in RQ1, RQ2, and RQ3 being *how* questions that could lead to the existence of different answers within different contexts and requiring the exploration of complex social factors between the actors within the research environment. As each actor may assign their own meaning to the world to create their own realities,

this confirms that the study aligns closest with the interpretivist paradigm. Much of the prior literature within information quality takes a positivistic approach, and the application of alternative views form a major contribution of this study (see Section 1.5). Whilst the output of the research is a model, this needs to be contextualised to help practitioners understand the issues with generalising outcomes across different cases.

Theoretical paradigm	Positivism	Interpretivism	Critical research	Pragmatism
Epistemology	Data based on observations and testing. Elements reduced to simplest explanations.	Meanings applied to social phenomena.	Meanings applied to social phenomena through the lens of power imbalance.	Meanings and observations can provide sufficient knowledge based on the research questions.
Ontology	Independent of actors and experiences.	Constructivism. Multiple mutable social constructs.	Inequality is inherent in social situations.	Symbolic realism. Views chosen to answer research question.
Axiology	Independent and objective	Researcher cannot be separated from study; subjective	Objective or subjective to expose inequalities	Objective or subjective as researcher determines appropriate to context
Empirical focus	Formalisations	Beliefs (socially constructed cognition)	Time and context-sensitive beliefs	Actions and changes
Type of knowledge	Observations	Understanding	Critiques	Constructive knowledge
Role of knowledge	Concrete truths	Interesting	Identifying power	Useful for action
Type of investigation	Quantitative analysis	Ethnographic and case studies	Historical, ethnographic, and case studies	Inquiry
Data generation	Observations	Interpretation	Interpretive reflection	Assessment and intervention
Role of researcher	Observer	Engaged in understanding	Identifying conflict	Engaged in change

Table 5.1: Comparison of information systems paradigms (based on Chua (1986); Goldkuhl (2012); Oates (2005); Saunders et al. (2003)).

5.7. Philosophical selection

The previous sections in this report have outlined the various choices of epistemology, theoretical paradigm, methodology, methods, and data generation methods following the researcher's seven elements of research model in Figure 5.2. From this set of established practices within the information systems field, a specific research strategy is selected to answer the research questions outlined in Section 1.2.1. Positioning research within Information Systems, any 'study should inform the reader of axiology, ontology, epistemology, methodology, research approach, and criteria for evaluation of the form of AR' (DeLuca et al., 2008, p. 58).

This section contains a description of the selection process and the justification for the researcher's philosophical position in relation to this project. Explaining the philosophical assumptions of the researcher is important as these will heavily influence the methodological design decisions made in the next chapter (Coombs, 2017, pp. 2–3; Keutel et al., 2014).

An interpretivistic paradigm is applied to this research. The researcher selected this approach as the research questions are aimed at understanding how and why particular practices occur, and the meanings attributed by users to the content they are accessing. A positivistic approach to information quality has been much explored in the literature (see Section 5.6.1), however this 'hard' approach assumes that there is a single objective truth rather than assigning the differences in user perceptions to the variety of perspectives held by individuals. Exploring this version of reality involves identifying the underlying understanding users have of the content that they are processing. This can be a complex problem, exacerbated by the variety of individuals using online systems and the differences in their experiences and perceptions of computer-based information.

Whilst a pragmatist approach could be justified for this research, the application of a pre-determined methodology aids the researcher in addressing the research questions in a timely manner and help bound the

activities to create a manageable plan. The choice of the interpretivist paradigm for this project does not discount the use of pragmatist ideals within individual activities; Goldkuhl (2012) suggests that such elements are frequently implicitly present within information systems action-orientated research.

5.8. Summary

The chosen epistemology and theoretical paradigm are suitable for this research, as they will enable the researcher to address the research questions and identify user behaviour leading to the analysis of online content.

The following chapter discusses the research design based on the philosophical positioning discussed in the previous sections.

Chapter 6. Methodology

During the previous chapter, the researcher positioned this research project in terms of the relevant information systems philosophy. Applying an interpretive approach, the researcher will explore the information quality attributes which have the largest impact on the success of online content. By examining a series of case studies with an *action case* methodological approach, participants will help demonstrate the subjective improvement of specific web content through an application of the *think aloud protocol*. This will lead to the development and validation of a model for general practitioner use.

This chapter discusses the methodological selection process and justifies the approach to the research study.

6.1. Introduction

The chosen epistemology and theoretical paradigm are suitable for this research, as they enable the researcher to address the research questions whilst accounting for the problem context and social constructs. By answering the *how* and *why* questions, the study will contribute new knowledge in the form of theory and practice to enable the development of more informative online content within existing websites.

The following chapter discusses the research design based on the philosophical positioning identified and selected during this chapter.

6.2. Quantitative and qualitative approaches

Within the existing information systems research, there are broadly two categories of research methods: quantitative and qualitative (Venkatesh et al., 2013, p. 21). Quantitative methods produce data and evidence based on numbers (Oates, 2005, p. 245), and are most commonly associated with

positivistic approaches. Some examples of studies that can produce quantitative data are experiments and surveys, amongst other methods (Oates, 2005, p. 245).

On the other hand, interpretivistic studies tend to be closer to qualitative methodologies, where the researcher is concerned with non-numeric data (Oates, 2005, p. 266) and need to provide contextual analysis to process into conclusions, although this is not exclusively true. Oates (2005, p. 267) suggests that it is possible to apply quantitative measures to qualitative data by abstracting from the raw data to identify key themes and convey these to the reader. This is not a straightforward task as the researcher must select valid approaches for codifying and simplifying raw data and be able to justify this based on prior literature and validated research methods. Another challenge for researchers is the volume of data produced, which can be much harder to handle in qualitative studies based on verbal interviews and written information.

De Villiers (2005, p. 2) explains the difference between quantitative and qualitative methodologies by placing individual methods on a spectrum and indicating the potential overlap (see Figure 6.1).

Figure 6.1: Spectrum of research methods (De Villiers, 2005, p. 2).

This study will follow an interpretive paradigm, therefore will focus on the methods towards the right of the spectrum. The categorisation of methods into quantitative and qualitative is not straight-forward as there are many methods that could be used to obtain both forms of data. Two such methods are observations and surveys, where data can be obtained in quantitative, qualitative or both formats.

To inform the choice of research method, a summary of the differences between quantitative and qualitative methods is shown in Table 6.1 below.

	Quantitative research	**Qualitative research**
Purpose	Existing hypothesis tested and verified using data.	Discovery of meaning through interpretation of data.
Dominant paradigm	Positivist	Interpretivist
Type of questions	Closed	Open
Concepts	Distinct variables	Themes, taxonomies and generalisations.
Replicability	Easily replicated by other researchers as environment is controlled and standard procedures used.	Challenging due to contextual nature and difficulty in recording all variables.
Data	Numeric	Words and images from observations and documents.

Table 6.1: Summary of differences between quantitative and qualitative research methods.

This table demonstrates that whilst there are many similarities between quantitative and qualitative data, each begins from a different purpose. Quantitative studies will start from existing hypotheses, whilst qualitative studies will be constructing and exploring meaning through interpretation rather than existing propositions. Due to this different in purpose, qualitative studies ask open questions, obtaining data in non-specific formats such as words, images, documents, and observations. The researcher will then transform their interpretation of the data, informed by prior literature, into generalisations that can be applied during future studies. Replicating a specific study can be challenging due to the context of the data collection: much depends on the interpretations of the researcher who is informed by actions and descriptions provided by research participants. The general ideas produced during qualitative methods can be validated by applying the results to alternative data that fits within the constraints and assumptions provided by the researcher.

6.2.1. Mixed methods

It is possible for the researcher to more formally combine both quantitative and qualitative analysis through the application of a mixed methods approach. Within the information systems field, there is a 'dearth of research … that employs a mixed methods approach' (Venkatesh et al., 2013, p. 22), although there have been theoretical suggestions to apply such an approach to aid triangulation of a problem context to provide a deeper understanding of the issue being studied (J. Mingers, 2001; Venkatesh et al., 2013).

Undertaking a mixed methods approach can often place the researcher in a place of conflicting theoretical paradigms – for example, can research be considered truly interpretive when it is simplified to a numerical value for comparison with little context of the derivation? Venkatesh et al. (2013, p. 22) suggest that a pan-paradigmatic approach can aid a researcher to 'find theoretically plausible answers to his or her research questions', and

furthermore the choice of using mixed methods should be left to the suitability of the research question, purpose and context (Creswell & Clark, 2007).

Within this thesis, *mixed methods* is taken to mean research involving more than one worldview (quantitative or qualitative), with methods being applied concurrently or sequentially to build on prior findings. An alternative interpretation of this term would be Mingers' (2001) concept of multimethod research, where the methods may co-exist in separate worldviews. Applying the strictest definition of the term, all mixed methods research could be considered multimethod, however not all multimethod research could be labelled as mixed methods (Venkatesh et al., 2013, p. 24).

Traditionally, information systems research has been focused on qualitative research for theory testing, with quantitative methods applied for confirmatory studies (Venkatesh et al., 2013, p. 24). This reflects the notion that qualitative research will help broaden a researcher's understanding of a problem space (Walsham, 2006). Mixed methods address this divide by addressing both exploratory and confirmatory questions within a single study,

Responding to the difficulty in applying mixed methods approaches to information systems research, Venkatesh et al. (2013, p. 41) propose a set of guidelines for conducting such studies (see Table 6.2).

Guideline	**Summary**
Decide on the appropriateness of a mixed methods approach	Selection must be based on research questions, objectives, and context.
Develop a strategy for mixed methods research design	Strategy must be designed to answer the research questions and contribute to the overall conclusions.
Develop a strategy for analysing mixed methods data	Analysis must be rigorous for all methods used during the research.
Draw meta-inferences from mixed methods results.	All inferences must be appropriate to all methods in the research and show a consistent perspective.

Table 6.2: Summary of Venkatesh et al's (2014, p. 41) guidelines for mixed methods research.

These guidelines provide a framework for the researcher to justify selection of mixed methods approaches, ensuring that the methods are consistent with the research basis and paradigm-based perspectives. Analysis of the data is as important as data collection; without considering the interpretation of data then rigorous outcomes cannot be produced.

By developing pluralistic research using these guidelines, the researcher can effectively design and implement mixed methods research to produce cohesive results. Several reasons for using mixed methods are discussed in the following sections.

6.2.2. Triangulation

Within a study, researchers will often apply a mixed methods approach to evaluate the findings of multiple methods. This is termed method triangulation, and can minimise the limitations, weaknesses and biases introduced through the use of individual qualitative methods. In addition, a smaller number of observations can often be conducted, enabling the researcher to focus on improving the accuracy of the findings (A. Adams & Cox, 2008, pp. 25–26).

There are also several other types of triangulation that can be conducted (see Table 6.3), however these are outside the scope of this research as they need to combine multiple strategies or perspectives within one project.

There are other purposes for conducting mixed methods studies (as described by Greene et al. (2008, p. 127) in Table 6.3), however these are incompatible with the aims of this research as the project requires answers to the research questions within the situational context, as opposed to deeper findings with a broader generalisability.

Purpose	Rationale
Triangulation	Seeks corroboration and convergence of results to increase the validity of constructs and reduce method bias.
Complementary	For elaboration and enhancement of results, increasing interpretability and meaningfulness.
Development	Uses results from one method to inform questions and/or results in another, maximising the strengths of particular methods.
Initiation	Seeks the discovery of contradictions and recasts questions between methods, increasing breadth and depth of results.
Expansion	Used for multiple inquiry components to select the most appropriate methods.

Table 6.3: Purposes for mixed method designs (based on Greene et al. (2008, p. 127)).

The first purpose of triangulation aims to increase validity of outputs by reducing inherent bias, with data collection repeated between different methods to corroborate results. Complementary mixed methods aim to add extra meaning to the data through enhancing results, for example by performing additional data collection to provide further insights. The latter three purposes aim to help with developing and discovering research questions and their basis, through informing and evolving questions along with selecting research methods through application.

Selecting mixed methods can aid triangulation, though this must be justified in terms of the research questions. Many of the aims of triangulation can be achieved through other means using methodologies such as action research and action case, explored later in this chapter from p. 159.

6.2.3. Validation

During a research study, it is preferable to ensure that there is an appropriate level of validity within data collection and analysis to increase confidence in results and maintain the reproducibility of the research. There are two types of validation, each of which is described below.

6.2.3.1. Internal validity

Internal validity is often associated with quantitative studies, though can also apply to qualitative research. Researchers must ensure that variation between data points is limited to the research questions, reducing the impact of external factors such as organisational changes, maturation or subtle persuasion by researchers (Oates, 2005, pp. 131–132).

Good research design, including the establishing of suitable research questions, selection of appropriate methods and considering choice of participants can alleviate some concerns about internal validity, however the researcher must understand the consequences of external factors on research outcomes.

6.2.3.2. External validity

The second type of validation is *external validity*. This considers whether the outcomes of a study are generalisable and repeatable, and not determined by a particular set of circumstances that have limited the scope of the research. Oates (2005, p. 133) suggests that 'the main threat…comes from non-representativeness', i.e., using too specialised a case study or set of participants. This can produce research with hidden assumptions, not applicable within the contexts described by the research objectives.

In their review of qualitative methods in prior literature, Hayashi Jr et al. (2019, p. 101) identify several works describing external validity in terms of *trustworthiness* or *rigour*. This change in nomenclature shifts the focus of external validity from a rigid process to an understanding of the outcomes of qualitative research: outcomes should be relied upon to not contain additional contextual assumptions or method selection oversights, maintaining academic integrity with the researchers identifying, documenting, and justifying their own limitations. The authors reform validity into a construct of continual development, with researchers analysing and reflecting upon their own outcomes between data collection and analysis

activities. This approach increases reliability and facilitates the identification of non-representativeness at an early stage in the study.

In positivistic studies, researchers often apply representativeness to the point of *data saturation*, that is when no more data can be obtained that would change the outcomes. Applying qualitative methods within the interpretivistic paradigm rarely leads to data saturation as there is often no one-size-fits-all method for transforming more data into stronger research outcomes (Fusch & Ness, 2015, p. 1413). Thus, data saturation is not a necessary consideration for qualitative methods, as this does not always produce outcomes with stronger external validity.

6.2.4. Summary

By developing a mixed methods approach, the researcher can develop a study that will provide improved answers to the research questions. In order to aid identification of the particular methods to be used in the study, the next section will examine relevant methodologies.

6.3. Methodologies

Methodological choices represent the third layer of Crotty's (1998) research model (see Figure 5.2), and as such is informed by the epistemology and theoretical perspectives. The chosen methodology represents an overall plan or strategy for the research and guides the selection of methods for the study.

6.3.1. Definitions

Mingers (2001, p. 242) identifies three different uses of the term *methodology* within information systems literature as shown below in Figure 6.2.

1. The study of methods.

2. A particular research study's approach.

3. Determined combinations of methods that may be frequently used together, to form a set of guidelines.

Figure 6.2: Definitions of methodology within IS literature (Mingers, 2001, p. 242).

The first definition concerns the study of methods themselves, which is outside the scope of this research. The third definition provides loose sets of guidelines with methods that are often combined; however, this does not maintain the rigour of the second definition. Throughout this chapter and the research study, the term *methodology* refers to the approach of this study in collecting, analysing, and evaluating data to produce research outcomes for the research questions first described in Chapter 1. A number of appropriate methods will be identified and justified later in this chapter to develop the study.

As the methodology chosen in this research will be formed of a number of methods, each of these will 'make implicit or explicit assumptions about the nature of the world and knowledge' (Mingers, 2001, p. 242). Methodologies often 'develop implicitly or explicitly within a particular paradigm' (J. Mingers & Brocklesby, 1997, p. 490), highlighting the need for the selected methodology to be consistent with the interpretivist approach identified previously in this chapter.

6.3.2. Soft Systems Methodology

Soft systems methodology is built on the principle that all research is composed of three elements: a framework of ideas, a methodology and an

area of concern. Checkland's (1985, 1991) model[1], as shown in Figure 6.3, demonstrates that the process of doing research can aid the researcher to learn about all three elements. This resulted in the development of a methodology, SSM, to aid researchers in the resolution of softer interpretive information systems problems that are difficult to explain using harder positivistic methodologies more traditionally associated with computer and systems engineering (Checkland & Holwell, 1997, p. 24). Applying SSM to a problem will help to bring about change by exploring the problem space and refining the research questions. A different way of considering this situation is Argyris & Schon's (1978) double loop theory, where the dual cycles represent the researcher's concurrent learnings about the domain and the methodology leading to the investigation. This can lead to findings that challenge the current research norms, however this can help the researcher identify escalating systematic errors and issues with the governing variables in the framework (Hughes & Wood-Harper, 1999, p. 87).

[1] In Checkland's works, all diagrams are hand-drawn to reflect the organic and soft nature of investigating information systems with an interpretive approach, highlighting the elusiveness of absolute certainty in social contexts (Checkland & Poulter, 2006, p. 198). In this thesis, the researcher has reproduced the illustrations in a consistent digital style, however this is not intended to detract from the looser nature of research activities suggested by the original literature.

Figure 6.3: Elements relevant to any piece of research (Checkland, 1985, 1991; Checkland & Holwell, 1997, p. 23).

Developing this further, the researcher can construct interpretive studies to match the concept of repeated self-improvement and learning about the framework of ideas, methodology and area of concern, forming action research. In SSM, it is evident that both the participants and the researcher are actively involved in the activity, since the researcher is reflecting on both others and their own roles within the context being studied. To document this process, Checkland (1991) defined the cycle of action research in human situations, as shown in Figure 6.4. Throughout this cycle, the researcher will become involved in the action as a participant, and consciously reflect upon their experiences to extract learnings.

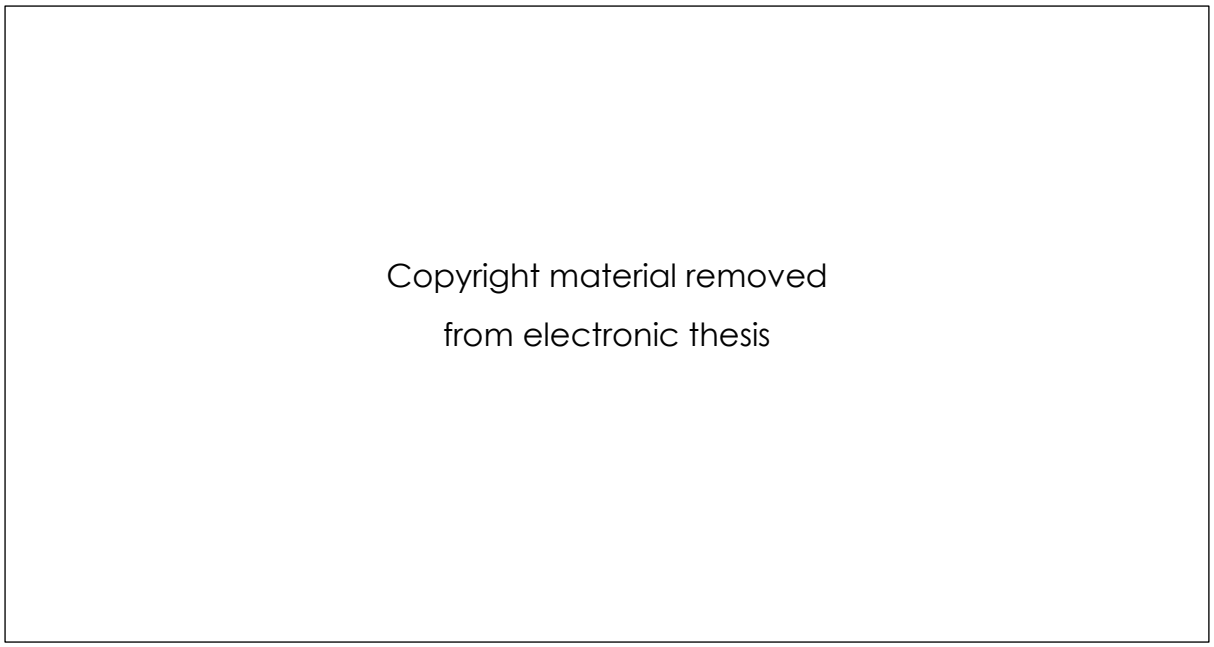


Figure 6.4: Cycle of action research in human situations (Checkland, 1991; Checkland & Holwell, 1997, p. 26)

Viewing this from another perspective, the researcher can be seen as bringing about change in a problem context by following the process of action research as shown in Figure 6.5. The amount of active participation can vary between projects, however the researcher should always be engaged at some level as a participant, and reflecting on their own activities. Checkland & Holwell (1997, pp. 26–27) note that methods other than action research exist for exploring this same problem space, although SSM is particularly suited to action contexts heavily relying on the influence of human actors. Unlike other methodologies, SSM does not impose a structure on the situation; instead, it aims to discover the patterns inherent in the contextual problem space (Checkland & Poulter, 2006, p. 158). This approach means that the researcher needs to take an open approach to the situation and recognise that the project may need to adapt as the context evolves.

Figure 6.5: Process of action research (Checkland, 1991; Checkland & Holwell, 1997, p. 27).

Soft Systems Methodology is a natural choice for this study, as the 'soft' problem space involves investigating human actors within social contexts, and the researcher will refine the research questions based on exploring the reflective learnings produced at each stage of the activity. This does not preclude the use of mixed methods within the action cycle as Mingers & Brocklesby (1997, p. 491) identify five possibilities for combining methodologies. Through these, the researcher can select the methodologies most appropriate to their situation, as often methodological isolationism will not provide enough evidence to sufficiently answer a research question. This also reflects the reality that interventions are often not discrete events, and the researcher needs to use a variety of methods to access data within the appropriate contexts.

6.3.3. Summary

For this research, a multimethodology approach is the most suitable, as it presents a strong basis in soft systems methodology, whilst permitting the methodological pluralism of mixed methods. This type of multimethodology is often considered problematic due to the possibility of incompatible paradigms, forcing the researcher to choose between conflicting fundamental philosophical assumptions (J. Mingers & Brocklesby, 1997, p. 496). Works such as Astley & Van de Ven (1983) and Burrell & Morgan (1979)

demonstrate that this observation is commonly observed due to the objectivist-subjectivist dichotomy as opposed to the other philosophical dichotomies where common characteristics of the paradigms are often identified. In addition, conflicting 'methods can be combined without crossing paradigm boundaries' (J. Mingers & Brocklesby, 1997, p. 498), reflecting a common approach by practitioners within industry. Choice of methodology often arises from the researcher's basic assumptions and world view, and provided that the methodologies selected are consistent with this, a multimethodology approach can still be valid.

Practically, the methodology selection relies on the experience and preference of the researcher as an agent, and such researchers cannot be forced to use methodologies that act against their beliefs and views (J. Mingers & Gill, 1997).

6.4. Research methods

The choice of research methods is critical to the success of a study, and must be aligned with the context, field, and theoretical perspectives. Selecting the correct method can often be a case of using 'the methods that work for us, nothing more and nothing less' (King and Applegate (1997) as cited in Goles & Hirschheim (2000). By applying the methodological considerations described earlier in this chapter, the research methods for this study will be discussed and justified in the subsequent sections.

6.4.1. Action research

Action research is a methodology for conducting studies with practitioners to try a theory in real situations and use the experience to gain feedback and modify the theory prior to trying it again (Avison et al., 1999, p. 95). This insight is partially gained through the emphasis on what practitioners *actually* do as a pose to what they *say* they do, often resulting in a more accurate picture of industry practices than can be gained from other methodologies.

6.4.1.1. The action researcher

Since its inception in the 1940s, action research has been used to investigate and improve researchers' own practices, blurring the boundaries between practitioners and researchers (Oates, 2005, pp. 154–155). As the roles within the activity may change depending on the situational context, researchers are often considered more as facilitators, collaborating with other practitioners to analyse current behaviour and seek methods for improvement.

Walsham (1995) explains an action researcher as an *involved researcher*, reflecting their role as a participant as well as an observer in studies. In Walsham (2006, p. 321), this view has evolved to become a spectrum, where the researcher can be placed anywhere from outside researcher to involved researcher, reflecting the muddled nature of the researcher's role within many information systems techniques. Having active involvement in the studies being conducted can be observed as a response to the frequent calls within information systems literature for researchers to increase the relevance of their work to industry (Baskerville & Myers, 2004).

To understand the boundaries of the study, it is important to define explicit criteria prior to commencing research, as unbounded action research can often lack a focus and fail to deliver results to substantiate the outcomes (Avison et al., 1999, p. 96), leading to 'research with little action or action with little research' (Dickens & Watkins, 1999, p. 131). This misinterpretation of the methodological outcomes has been countered by Avison et al. (2018) through their analysis of 120 empirical action research studies, demonstrating the overall success of such research.

Cecez-Kecmanovic et al. (2020, p. 256) highlight the numerable forms of action research within the information systems field and identify the 'neomethodological creativity' with which researchers can develop new forms of action research without committing to their long-term evolution. This

can often lead to the abandonment of subtle improvements or suggestions for canonical action research due to their erroneous classification as a new methodology rather than an evolution of the core body of knowledge.

6.4.1.2. Cyclical approach

One of the major characteristics of any action research is the use of a cyclical approach to improvement. Susman & Evered (1978) express this approach as a five stage plan (as shown in Figure 6.6), which is often termed *canonical action research* (Davison et al., 2004, p. 66).

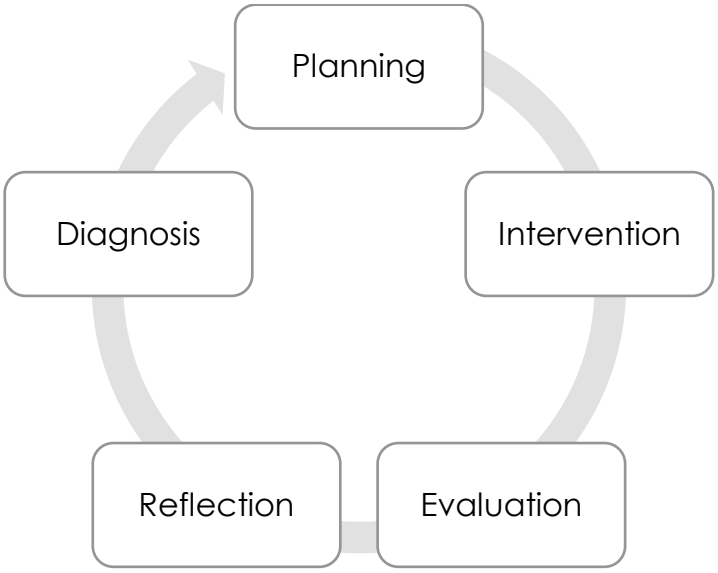


Figure 6.6: Five stage cycle of action research (Susman & Evered, 1978).

In this five-stage cycle, action research begins with the *planning* stage where the researcher will consider the research aim, objectives, and questions to determine the object of study. This is followed by an *intervention*, where the researcher immerses themselves in the research context and makes changes to the situation based on the research questions. The consequences of this activity are then *evaluated*, with the researcher *reflecting* upon their own actions and the impact on the context. Finally, the problem is *diagnosed* to identify further potential changes that could alter the situation. A further

change is then selected by the researcher, which is progressed during the next cycle of the study.

The researcher will stop the action research once enough cycles have been performed to identify and document patterns of behaviour based on the interventions, thus providing outcomes to the research questions.

6.4.1.3. Theory and knowledge development

A common criticism within the literature surrounding action research is the lack of theory building during cycles. This depends on the manner in which the study is conducted, as authors discussing established information systems methodologies describe the prior knowledge of a theory to test and improve upon as a key tenet of performing effective action research (Checkland & Scholes, 1990; McKay & Marshall, 2001; Susman & Evered, 1978).

Heron & Reason (1997, pp. 280–281) recognise that action research can produce four different types of knowledge (see Table 6.4). These types can be used to analyse the outcomes of action research and confirm that the process has produced useful knowledge.

Experiential	Tacit knowledge gained by direct encounter; involves empathy, intuition and feelings that are difficult to put into words.
Presentational	Emerges from experiential knowledge as knowing something through stories, drawings, and other media.
Propositional	Knowledge gained in the form of logically organised ideas and theories, e.g., academic research.
Practical	How to apply knowledge to exercise a skill.

Table 6.4: Four types of knowledge that can be produced by action research (Heron & Reason, 1997, pp. 280–281).

6.4.1.4. Research outcomes

Checkland & Scholes (1990) add two more criteria for evaluating the success of information systems action research:

1. Are there practical achievements in the problem situation?

2. Are there learnings about the process of problem solving?

These questions focus on the problem-solving nature of action research, identifying the close links between the methodology and solutions as research outcomes.

In their framework for Information Systems research within organisational contexts, Braa & Vidgen (1999, p. 3) describe the three intended research outcomes as three ideal *Weberian* types: prediction, understanding and change (see Figure 6.7).

Figure 6.7: Braa and Vidgen's (1999, p.3) IS research framework for the organizational laboratory.

These three types are not fully attainable in practice; however, they serve as a convenient classification for categorising the various research methods within the field. Vidgen & Braa (1997) analyse the alignment of the three types, aligning them at their extremities to information systems paradigms as shown in Figure 6.8.

prediction	Aligned with positivistic systematic reduction
understanding	Aligned with interpretive approaches
change	Aligned with interventionary approaches

Figure 6.8: Alignment between action case types and IS paradigms (Braa and Vidgen, 1999, p.3).

The three dynamics are co-present in all chosen research methods, although with different emphasis. Classifying the relevant in-context research methods, Braa & Vidgen (1999, p. 7) highlight three core methodologies aligned with the dynamics: those of field experiments, soft case studies and action research (see Figure 6.9). In general, the three dynamics can be considered as providing an abstraction of positivism, interpretivism and interventionist perspectives as described in Section 5.6.

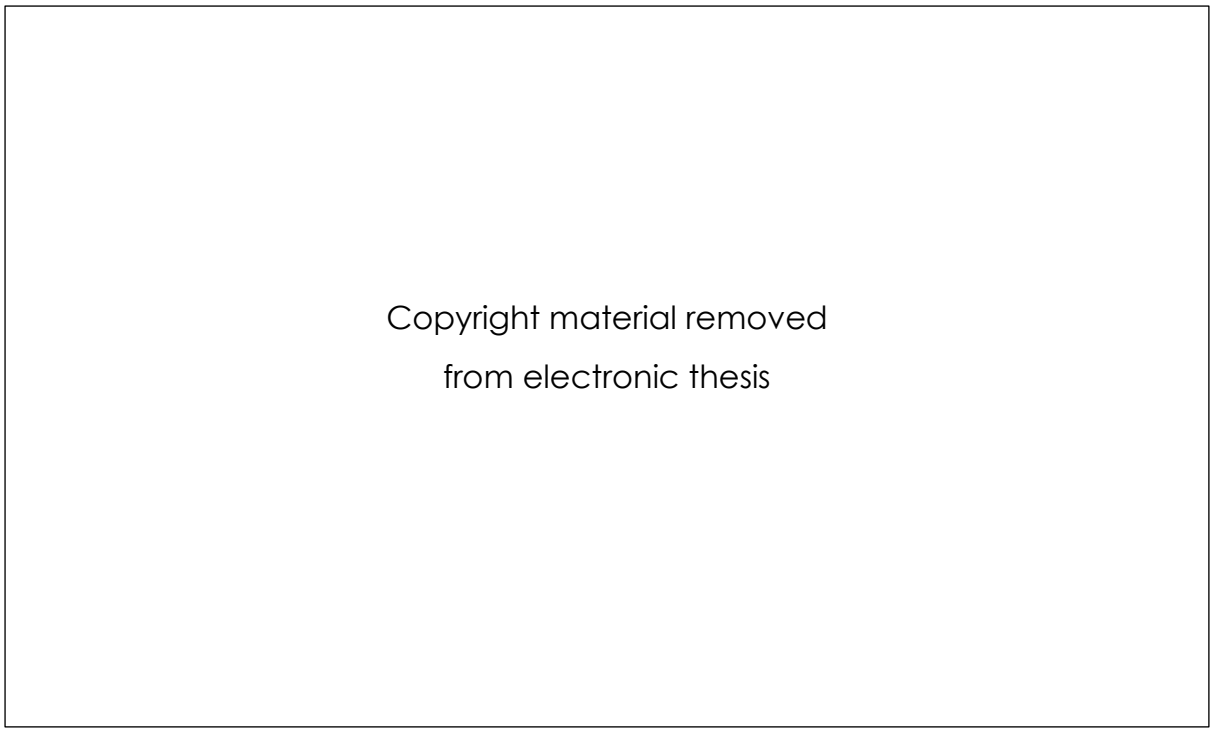


Figure 6.9: Location of in-context research methods in Braa and Vidgen's (1999, p.7) framework.

As discovered by the framework, there is the theoretical space for a previously undescribed hybrid methodology to bring together the change aspects of action research with the understanding aspects of soft case studies: action case. This will be explored as a suitable method for this study in Section 6.4.3.

6.4.1.5. Summary

Action research can provide a methodological grounding for qualitative information systems studies, especially within the context of practitioner-focused research outcomes. Applying an iterative process is often considered best practice within industry, with canonical action research providing a structured approach for studies incorporating small changes to organisational contexts followed by reflective and evaluative stages.

The following section will discuss case studies as another interpretive methodology mapped by Braa & Vidgen (1999).

6.4.2. Case studies

Case study research is often carried out when researchers are studying how or why questions (Benbasat et al., 1987, p. 370; Yin, 2014, p. 2), and as such aligns well with the interpretivist paradigm (see p. 134) although it is possible to place such work within the positivistic or critical paradigms (Myers, 1997). In addition, the context being studied must be contemporary (i.e., the events can still be observed at present) and not be influenced by the control of the researcher.

Case studies are often conducted when researchers are focusing on 'one instance of a thing' (Oates, 2005, p. 141) where there is no prior knowledge of the context or influencing factors (Benbasat et al., 1987, p. 370). Often, the researcher will use multiple data generation methods such as interviews, documentary analysis, observations and questionnaires, to examine the situation in-depth and build a detailed picture of the case and its relationship to the actors (Oates, 2005, p. 141).

6.4.2.1. Types of case study

Oates (2005, p. 143) categorises case studies into three key types, as shown in Figure 6.10.

Exploratory	Defines questions to be used in subsequent studies.
Descriptive	Builds a detailed picture of a particular phenomenon and its context.
Explanatory	Extends a descriptive study to answer questions such as why particular events happened, and why certain outcomes occurred.

Figure 6.10: Types of case study, as identified by Oates (2005, p. 143).

This framework ranges from the simpler case studies used to identify research objectives (exploratory) to those that are much more complex with detailed outcomes (explanatory), adding narrative to data obtained through qualitative methods.

Any of these types of case studies can be conducted as either a single isolated case, or multiple cases examining different contexts and providing validation of the existing findings. Case study research can prove challenging due to unclear and unspecified boundaries: a researcher needs to define what should and should not be included within a particular context (Oates, 2005, p. 148).

Vidgen & Braa (1997) provide an alternative taxonomy, describing case studies as either *hard* or *soft*. Hard case studies are those based on a positivistic paradigm, answering closed questions with mainly quantitative data. Soft case studies focus on the interpretive paradigm, where data is often provided through open methods to explore situations rather than confirm hypotheses. Any of the types described by Oates can use quantitative or qualitative methods, therefore we can combine the taxonomies to create six basic types of case studies as shown in Figure 6.11.

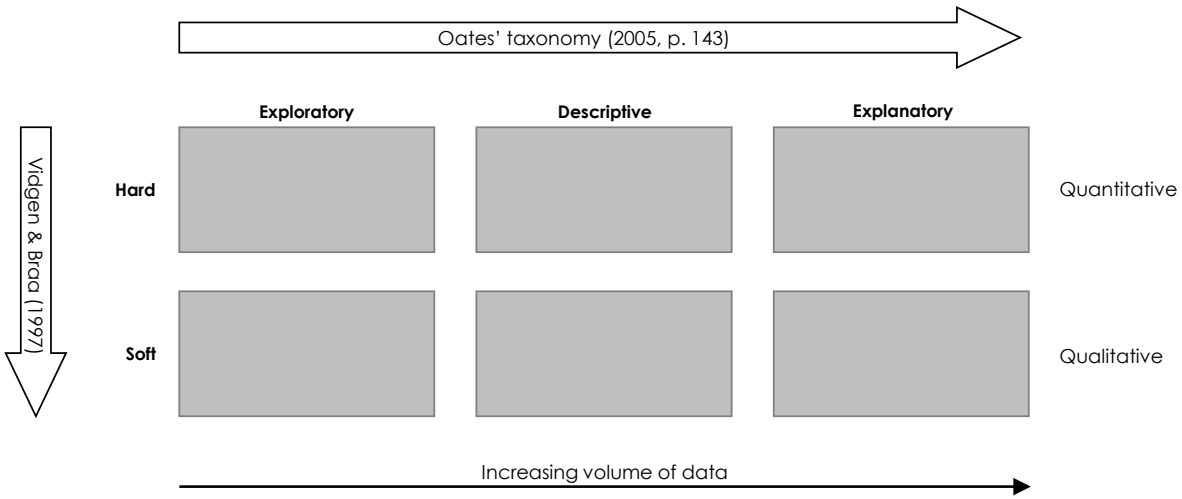


Figure 6.11: The six types of case study, showing categorisation by Oates (2005) and Vidgen & Braa (1997).

Case studies can thus be mapped on the matrix based on their purpose (Oates) and methodology (Vidgen & Braa). As established during the prior chapter, this research study will follow an interpretive approach, thus any case studies completed will be *soft* in nature. The individual case studies will have various relationships to Oates' taxonomy dependent on their role in the research cycle. For more details of the specific case studies, see Section 6.9.

6.4.2.2. Similarity to action research

According to Simonsen (2009, p. 113), the main difference between action research and case studies is that in the former, the researcher is aiming to change and improve the phenomenon in question, rather than function merely as an observer. Findings are often still generalisable, as factors found in one case study may also be found in others. The researcher is required to fully explain the circumstances surrounding the study to demonstrate how the research was conducted to maintain repeatability in the future. Generalisations can occur in four different ways: concepts, theories, implications and rich insights (Walsham, 1995), giving depth and variety for the information systems researcher to use case studies to contribute to their findings.

6.4.2.3. Summary

Case study research within the interpretive paradigm is often of the soft nature, applying qualitative techniques to identify reproducible outcomes from individual scenarios. By considering the purpose of case study research, appropriate situations can be found within organisational contexts to allow the exploration of key themes.

The next section discusses the evolution of soft case studies to *action case studies*, where the researcher is both an observer and participant within the context of the study.

6.4.3. Action case studies

Action case studies can uncover more understanding about the problem space than action research whilst working on shorter timescales and with less participation and intervention (Braa & Vidgen, 1999, p. 13; Hughes & Wood-Harper, 1999, p. 87). This method has also been proposed concurrently by other researchers describing their hybrid methodologies working within industry (Mårtensson & Lee, 2004) and from a theoretical viewpoint (Baskerville, 1999; Baskerville & Pries-Heje, 1999).

6.4.3.1. Relationship to action research

Based on 'soft' research methods such as action research, action case allows a researcher to understand a problem context without intervening to effect change. This may or may not be combined with a limited number of interventions, meaning that the researcher can focus on identifying new knowledge without needing full organisational access (Hughes & Wood-Harper, 1999, p. 87). Action case research also encompasses the concept of reflective action research, as a crucial element is 'the actors' discovery of where their behaviour is unexplained by their own understanding' (Baskerville & Wood-Harper, 1996). This correlates with Checkland's FMA framework (see Figure 6.3), demonstrating that the researcher can learn about the problem domain, the methodology and their own practice by performing the research activities.

Avison et al. (2018, p. 183) reviewed 120 empirical action research studies and concluded that such an approach is often inappropriate for doctoral studies due to a lack of skills, access to organisations and longitudinal timescales. The authors suggest that action case can be a suitable alternative, with multiple case studies forming individual work packages being combined into an overall research project.

6.4.3.2. Type of action case study

Action case studies are based on soft techniques as described by Braa & Vidgen (1999; see Figure 6.9). This results in the collection of qualitative data for analysis by the researcher, identifying themes within the underlying context.

The purpose of an action case study can vary even within the same research project: for example, an early action case study may aim to identify questions for further study (Exploratory), document practice (Descriptive) or capture descriptions of known phenomena (Explanatory). The purpose of individual action cases must be understood by the researcher to ensure that data is collected and analysed with an appropriate process. This is especially important with action case study research as the researcher may influence the outcomes based on their assumptions and actions within the context.

6.4.3.3. Conclusion

Action case study research is an appropriate method for studies where the researcher can make changes themselves to the context. This is possible within many studies of information systems, where responses are collected based on data within a system that can be adjusted to simulate the variety of situations a user may typically encounter. This enables the study to cover a wider range of phenomena than would otherwise be possible within an existing or static system.

6.4.4. Summary

This section has discussed the research methods considered for this project. Action case study research involves the researcher taking a more active role as an observer or participant within the problem context and analysing the situation to produce learnings based on the prior literature and methodology. Applying action case study methods to this research facilitates the collection of data from simulated environments under the control of the researcher,

enabling the system to be iteratively improved and this demonstrated through changing participant outcomes.

6.5. Methodological selection

This section describes the methodological selection based on the opportunities explored earlier in this chapter.

6.5.1. Methodology

This study follows Checkland's (1981) soft systems methodology (SSM) with a mixed methods approach. Soft systems methodology is appropriate as the research involves 'dealing with social or people based project situations' (Gunawardena & Brown, 2007, p. 5) with a lack of 'clear or specific objective' (Ho & Sculli, 1994, p. 48), requiring 'an enhanced understanding of a complex problem' (Baskerville & Wood-Harper, 1996, p. 239). This will provide an opportunity to study information quality criteria in the context in which it is applied (Orlikowski, 1992).

Following a mixed methods approach is necessary as the research questions require the analysis of data from a number of sources, and method triangulation can aid the researcher to identify patterns in the data whilst eliminating bias and reducing the problem space to a size that can be achieved within a single, focused study. This also reflects the diversity of the data generation methods that need to be applied to fully understand the problem and the context being studied.

6.5.2. Action case research

This research is being conducted using action case research to study online content within existing and new website systems . For an outline of the action research cycles, see Section 6.8.

Action research is an appropriate methodology for this study as it allows researchers to iteratively develop theories with a practitioner focus on real

situations (Avison et al., 1999, p. 95). Each cycle of the research will facilitate the development of the taxonomy of information quality attributes, resulting in the formation of *propositional knowledge* (Heron & Reason, 1997, pp. 280–281) as the academic model outlined in RO3. A descriptive case study will be completed with each research cycle, with data from a number of participants forming a detailed picture of the situation and context (Oates, 2005, p. 143).

6.5.3. Research stages

The research activities will follow four key stages as shown in Figure 6.12.

Stage 1
Identification of existing information quality attributes that have an impact on the informativeness online content.

Stage 2
Three individual case studies using the *think aloud* method to explore the informative aspects of web content with users.

Stage 3
Development of a model for assessing and improving the informativeness of website content.

Stage 4
Application of the model to a fourth case study to validate the proposed model.

Figure 6.12: Four stages of research study.

Firstly, the researcher will analyse existing information quality frameworks from academic literature to identify those attributes that have an impact on the informativeness of online content.

Secondly, three individual action case studies will be performed to identify the salient information quality attributes within separate domains. This will identify which of the attributes within prior literature relate to informative web content, and any new attributes which are only relevant to the online space.

Then a model will be developed to incorporate the three types of attributes identified during the first two stages as shown in Figure 6.13.

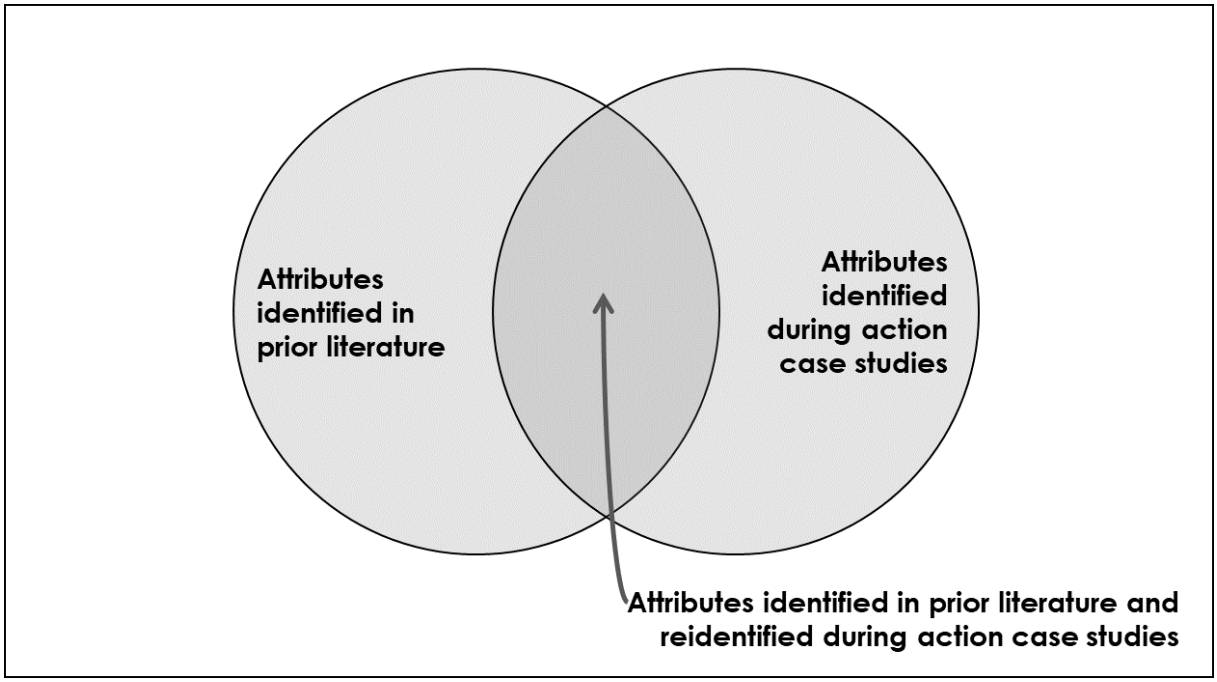


Figure 6.13: Three types of attributes identified during this study.

The three types of attributes are as follows:

1. Those included within the prior literature and identified within Chapter 7.

2. Attributes from the literature which are also apparent in the action case studies.

3. New attributes not in the literature discovered during the action case studies.

The model will combine both practitioner advice for informative content and the academic basis of key information quality attributes, providing both contributions to practice and theory.

Finally, the practitioner model will be applied to a fourth action case study website, demonstrating the validity of the model for a further domain and its suitability within web content improvement processes.

6.5.4. Summary

Applying an action case research method to the study will result in four action case cycles, working with participants to identify the key information quality attributes which contribute to informative web content. The remainder of this chapter describes the specific methods to be used during the three stages of research.

6.6. Task scenarios

Task scenarios are a research device commonly used in the fields of usability and HCI, especially for information retrieval studies (Kim, 2012, p. 300). They are particularly useful for raising questions about the importance of the activity and the inherent problems (Rosson & Carroll, 2002, p. 2).

The use of scenarios can help the researcher to identify and evaluate many viewpoints within usage patterns, identifying relationships within system entities (Carroll, 1999, pp. 5–6). This can provide an abstract perspective whilst maintaining a task-based focus on the research activities.

6.6.1. Key elements

Go & Carroll (2004) identify four key elements of task scenarios, as shown in Figure 6.14.

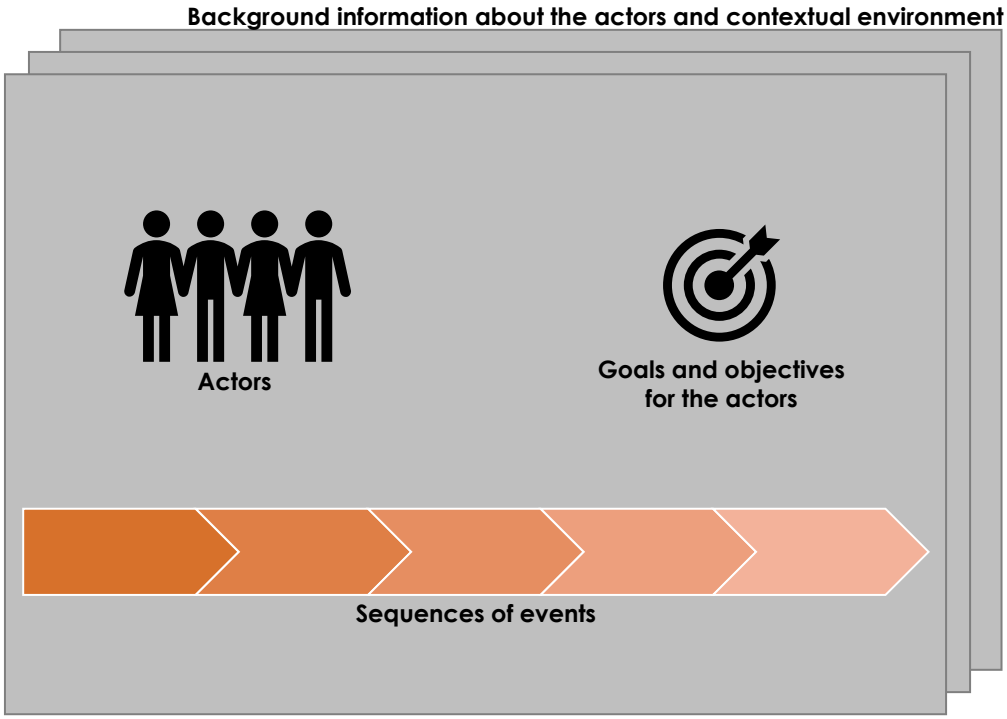


Figure 6.14: Four elements of task scenarios (based on Go & Carroll, 2004).

Each task scenario is a description that contains these four elements:

1. Actors involved in the problem context
2. Background information for the problem context and actors
3. Individual and collective goals and objectives for the actors
4. Sequences of events

By combining these into a scenario, the researcher can narrate typical activities of a user, leading to research instruments which simulate the information systems use.

6.6.2. Application

To ensure successful task scenarios, it is important to make them realistic and actionable, with the context of the activities explained to the participant (Nielsen Norman Group, 2014). The specific goals and objectives asked of users during a study can have a large impact on the repeatability of the evaluation, with an estimate of agreement ranging from 5% to 65% (Hertzum & Jacobsen, 2003). Even when repeating tasks, new usability problems may be discovered by future iterations of the same study (Cockton & Woolrych, 2001).

Further challenges are identified by Carroll (1999, p. 10) as illustrated in Figure 6.15.

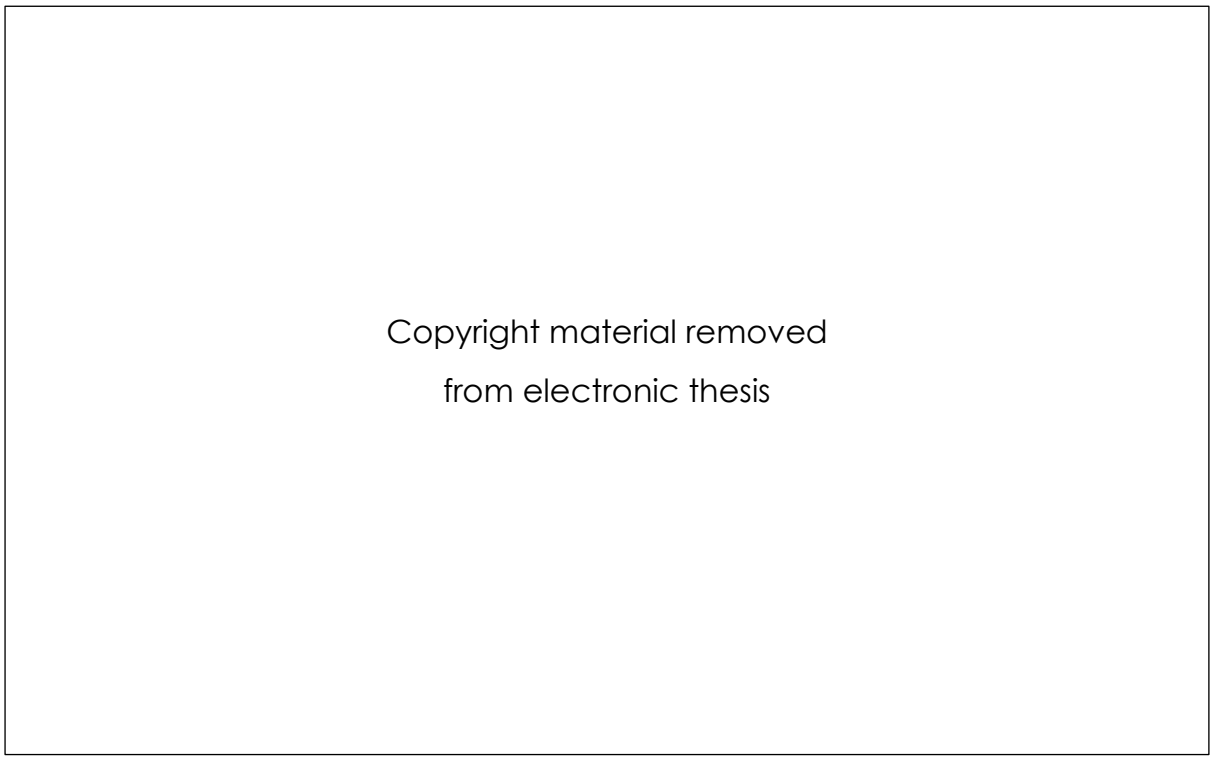


Figure 6.15: Five categories of scenario applications and challenges (Carroll, 1999, p. 10).

These five areas represent the challenges of applying scenario-based research design within information systems research. Some of the challenges are already addressed by the soft systems methodology and action case study research method discussed earlier in this chapter.

For example, the balance between action and reflection is demonstrated within canonical action research using the cyclical approach of Susman & Evered (1978, see p. 161). This incorporates reflection as an outcome of changes applied to the research context, ensuring that any adjustments are evaluated prior to further study. Each cycle of action research should address a single change, addressing the issue of multiple effects.

The current state of academic research ('scientific knowledge' within Carroll's diagram above) often lags behind practice and follows patterns of *waves* or *trends* (Baskerville & Myers, 2009; Bødker, 2006), thus this study will combine both academic and industry practices in the development of the model. This is a unique facet of information systems, contrary to other academic fields where practice can often lag behind current academic research (Morris et al., 2011).

6.6.3. Representativeness

Usability task scenarios need to be designed to be representative in terms of their users and the activities to be conducted in order to highlight realistic situational improvements (Nielsen Norman Group, 2014). The researcher will address the individual cycles by designing task scenarios aimed at realistic activities using the website, with general questions provided to users without specific guidance on achieving the outcomes. This will ensure that participants approach the task as a member of the intended audience rather than knowing the desired outcomes of the research activities.

6.6.4. This study

In this study, task scenarios are used as a basis of the think aloud research discussed later in this section. The tasks are representative of those faced by real users, with sufficient context to explain the goal and purpose enabling the participant to place themselves in the position of the actor. As information retrieval activities have a dynamic nature, the scenarios are not be rigid, facilitating natural search behaviours (Borlund & Ingwersen, 1997, p. 226). This elicits attributes of information quality from the participants' thought processes, allowing for the identification and classification of those that have the most impact on online content.

6.6.5. Summary

Task scenarios are an appropriate research instrument for the study of user journeys through informative content and will be used to simulate typical information seeking patterns for the four action case study websites. The specific design of representative scenarios will be discussed during each of the case study discourses in later chapters.

The following section discusses the think aloud protocol, the observational technique that will be used to illicit responses from participants during the research activities.

6.7. Think aloud protocol

The *think aloud protocol* is a general descriptor for a range of usability evaluation techniques that involve a user describing their thoughts whilst conducting a computer-based goal-directed activity. Think aloud approaches have been applied in many different fields, including cognitive psychology, education research and HCI (Blandford, 2019).

6.7.1. History

Initial psychological studies during the 1940s found that participants using think aloud were more likely to achieve success at task-based activities, whilst

they focused on verbalising rather than analysing their own thoughts (Sauro, 2016).

Think aloud was first described in a usability context by Lewis (1982) as 'appropriate for studying the cognitive problems that people have in learning to use a computer system'. As a technique first used within industrial design, think aloud has continued to be used in commercial settings as it is flexible, requires no specialist equipment, is quick to execute and delivers benefits even when the methodology is not fully applied (Nielsen, 2012). The longevity of the technique demonstrates its ability to produce results across a broad range of systems and within a variety of contexts, meaning it is still 'the single most valuable usability engineering method' (Nielsen, 1993, p. 195).

6.7.2. Key elements

Think aloud studies rarely follow a single prescribed pattern as there are many variants of methodological design depending upon the context, participants, and desired outcome. Clemmensen et al. (2009, p. 216) identified four main elements of think aloud studies as highlighted in Figure 6.16 that can be used to describe its application within a study.

Figure 6.16: Reference model of think aloud protocols as defined by Clemmensen et al. (2019, p. 216) with four main elements highlighted.

The four main elements highlighted are as follows:

1. Instructions and tasks

A key element of think aloud studies is a set of instructions forming a task set by the researcher. This is prepared ahead of the study to guide the participant through the key areas for consideration.

2. Verbalization

Participants will often fall silent during the tasks due to cognitive difficulties understanding the system or confidence with the activity. A researcher needs to be ready to prompt the user to continue verbalization, though this may range from a detached statement to keep them talking to more complex questions analysing their behaviour.

3. Reading the user

Observation of the user is a key element as this can allow the researcher to ascertain more information than the verbalizations identifying problem areas.

4. Overall relationship between user and evaluator

Ensuring that the participant feels comfortable with verbalizing their thoughts will encourage both positive and negative comments, giving a more holistic view of the system.

Figure 6.17: Four key elements of think aloud studies (based on Clemmensen et al., 2009, p. 216).

When designing a study, the researcher must consider the four elements above and define their approach to identify patterns used during the study. The extent to which each of the four elements is varied may affect the research outcomes as participants may be prompted to respond in different ways.

6.7.3. Guidelines

As part of research using the think aloud protocol, the researcher will want to deepen their understanding of the participants' thoughts without relying on voluntary verbalizations. This is achieved by the use of probing questions to help explain more information about the context and task (Nunnally & Farkas, 2016).

In order to protect the validity of think aloud research, Ericsson & Simon (1984) presented three key guidelines for such studies, as presented in Figure 6.18.

Guideline 1
A neutral instruction that does not request specific types of information.

Guideline 2
A practice session.

Guideline 3
A neutral "keep talking" reminder with no additional evaluator probes.

Figure 6.18: Three key guidelines for think aloud studies (Ericsson & Simon, 1984).

Applying these three key guidelines differentiates think aloud studies from other quantitative and qualitative techniques. Participants are expected to provide their own personal insights rather than follow a rigid model, with the researcher taking the role of both an observer and an interviewer to elicit detailed responses. The participant directs the conversation following their train of thought rather than following a series of directed prompts, although the researcher may intervene to probe for further specific insights. This requires careful balance from the researcher, ensuring that they provide appropriate direction without guiding the participant towards expected results or insights.

During their review of think aloud techniques in industrial contexts, Nørgaard & Hornbæk (2006, p. 271) observed that most studies failed to follow the above guidelines, with focus on the researchers' idea of the challenges and a prioritisation of usability problems. Nielsen (2012) does not view this as an issue, as useful results can still be obtained even with a poorly implemented methodology. This is further demonstrated by the variety of methodological approaches among usability professionals (Olmsted-Hawala et al., 2010, p. 2383).

6.7.4. Approach

Many participants prefer to work through think aloud activities at their own pace, with few interruptions from the researcher (McDonald & Petrie, 2013). During this study, the researcher will present each user with a task scenario, with an opportunity to read the brief and recognise the target website prior to the start of the activity. Conducting a trial session prior to the task often permits the participant to practice verbalisation and reduce the need for later intervention (Charters, 2003). Prompts during the activity aim to explore the current thought patterns rather than advise and direct the users, keeping involvement in the completion of the task to a minimum. This type of coaching has been demonstrated to lead to increased accuracy rates than the former guidelines (Olmsted-Hawala et al., 2010, p. 2387). Nørgaard & Hornbæk (2006, p. 217) conclude that thinking aloud should be conducted in a formative manner rather than a rigid, summative methodology. This adds value to the usability research, allowing for a responsive research format not provided for by the classic laboratory 'staid researchers' identified by Iivari (2005).

During the think aloud studies, the researcher makes notes based on participant responses; however, full transcripts are not produced as the specific utterances does not aid in the identification of the overall information quality attributes. All data is be held in accordance with the AoIR (2012) and BCS (2015) ethical conduct guidelines.

6.7.5. Concurrent vs retrospective

Think aloud studies can be conducted following one of two types: concurrent or retrospective. Concurrent think aloud (CTA) studies take place where the researcher acts as an observer, watching users interact with a scenario in real time asking prompting questions to understand their user model and thought patterns. Retrospective think aloud (RTA) studies allow users to conduct an activity and then discuss how they achieved the task and any challenges they faced. Both protocols can discover comparable sets of usability issues,

however these are discovered in different ways (van den Haak et al., 2003). The differences between these two types are summarised in Table 6.5.

Technique	**Concurrent think aloud (CTA)**	**Retrospective think aloud (RTA)**
Purpose	Understand participants' thoughts as they occur	Analyse participants' reflection on tasks
Timeliness	Real time	Short-term reflection
User cognitive effort	Reduced	Difficult to remember thoughts
Session length	As long as task as participation happens live	Increased as participation happens after task
Usability metrics	Can interfere with accuracy and time on task	Used in addition to other metrics
Replay	No opportunity to replay activities	Can replay activities to prompt feedback
Type of results	Observable	Verbalisable

Table 6.5: Comparison of concurrent and retrospective think aloud techniques (based on Blandford, 2019; Boren & Ramey, 2000; Nielsen, 2012; Nørgaard & Hornbæk, 2006; Usability.gov, 2019; van den Haak, de Jong, & Schellens, 2003).

Concurrent think aloud requires users to discuss their thoughts as they are conducting a task, reducing the speed of data collection, and causing greater cognitive load for the current activity, although it does not require participants to remember every step of what has been completed. Retrospective think aloud takes the opposite approach, allowing users to complete an activity alone and in silence, with data collected through a process more akin to semi-structured interviews after the task has been completed. Both of these have their advantages and disadvantages, with many researchers preferring the former to collect real time and realistic data, whilst studies using the latter can masquerade insights due to users forgetting or fabricating information (Altuntaç, 2015, p. 5). Concurrent think aloud studies often lead to increased reaction times, though this has been

demonstrated to not result in a reduction in task performance (Fox et al., 2011; Hertzum et al., 2009). Van den Haak et al. (2003) observed the major difference between concurrent and retrospective think aloud methods to be the type of problems identified: CTA produces more observable results, with RTA producing more verbalised results. As the information quality attributes in this study are based on observable properties of the content, this makes concurrent methods more applicable.

6.7.6. Online studies

Many prior works have utilised think aloud protocols for website usability studies (e.g. Benbunan-Fich, 2001; Hoppmann, 2009; Olmsted-Hawala et al., 2010), with some concluding that such direct observational studies are the most efficient at collecting user-informed data (Benbunan-Fich, 2001, p. 153). Whilst this study is examining textual content rather than navigational and structural concerns, it is challenging to isolate the different aspects of the website and think aloud will provide a unique insight into users' behaviour when retrieving information from the case study websites.

The use of think aloud protocols to study information quality attributes within online content has not been explored in the prior literature. A recent study has assessed a predetermined set of information quality heuristics using task scenarios and think aloud methods, recognising the difficulty of measuring IQ through quantitative metrics, and highlighting the need for assessing user experiences beyond fixed measures (Arazy et al., 2017).

6.7.7. Participants

One of the major benefits of the think aloud protocol is that it can be established with a small number of users with no methodological prescriptions to use representative samples. In their seminal work on think aloud techniques, Ericsson & Simon (1984) suggest that eight participants are sufficient to understand task behaviour, whilst Nielsen's (1994) early study indicated that more than 77% of user interface usability issues can be

discovered using just five participants. This was explored further by Lindgaard & Chattratichart, (2007, p. 1422), who found that 43% of usability problems were found by five participants during a multi-team study. Eveland & Dunwoody (2000) suggest that studies using the think aloud protocols can be hard to generalise due to such small sample sizes, though this is a larger problem for quantitative than qualitative studies.

The number of participants for HCI research studies has been much discussed within the field, with a general suggestion that, 'from three to twenty participants can provide valid results, and a good baseline is between five and ten participants. In general, there should be more participants for more complex, highly critical projects' (Six & Macefield, 2016). This builds on Nielsen's (1994) earlier research that suggested that 5 participants could discover 80% of usability problems. As this study is focusing on more complex situations and considering many information quality attributes, ten participants will take part in each cycle to meet the minimum requirement for discovering most information quality issues whilst accounting for the non-representativeness of the sample and potential parallel journeys which could be undertaken during the activity. Participants are not necessarily familiar with the website or domain of study: this lack of expertise is not a concern as participants can apply their existing knowledge of online searching to the new challenges (Tibau et al., 2021, p. 309).

For each cycle of research in this study, ten participants take part from a representative user base. By repeating the study across different genres of websites and versions of the same website, the results will become more generalizable and counteract the small-scale issues. This also requires a documented and repeatable methodology, including fixed scenarios to facilitate comparisons and transparency (Hoppmann, 2009, p. 213).

Unmoderated usability tests are recognised to produce fewer usability results than more structured methods (Andreasen et al., 2007), however there is little difference between results obtained in controlled laboratory environments

and through the use of remote technology (Brush et al., 2004). For this study, research is undertaken where possible in computer labs to facilitate easy access to equipment for the user, without being placed in a formal research context. This simulates the expected behaviour of representative users who may access the case study websites from any location, including those with noisy or unfamiliar environments. The type of equipment used for the study (e.g., PC, Tablet, mobile phone) is not varied as the same information is presented on each type of device, the study aims to focus on online content rather than the interactive elements and a PC screen is easier for the researcher to observe at the same time as the participant is working through the task scenario.

Practitioners involved in implementing online information systems often make design decisions for users based on their experience. Eliciting research from these may give insight into how they use the system, however such experienced users frequently struggle to explain their ideas and thoughts in an accessible manner (Ericsson, 2006, p. 223). Expert users often complete tasks quicker than novices, with a different view of performance based on their prior experiences (Lazonder et al., 2000). For this reason, novice users who are familiar with the research context though not with the specific website will be recruited to simulate the information retrieval activities undertaken by new users to accomplish the task scenarios. Jaspers (2009, pp. 341, 345) notes that think aloud techniques are the most used usability evaluation for focus on users, especially when representative tasks are used.

6.7.8. Interviews

Several studies have explored the differences between thinking aloud and interview techniques. Although retrospective think aloud activities can contain elements resembling an interview post-observation, these are focused on the task scenario and can be considered as a separate methodology.

In their studies of personalised systems, van Velsen et al. (2007, 2011) demonstrated that there are both positive and negative influences caused by the choice of methodology. Whilst thinking aloud elicited more negative comments from participants, this 'uncovers a unique set of issues with the system' that were not fully explored by interviews (van Velsen et al., 2007, p. 7).

Studies such as Fan et al. (2019) and Fan et al. (2021) have demonstrated that user experience problems are often suggested by subtle changes in speech and verbalisation patterns by both younger and older participants during think aloud studies. The researcher is aware that capturing these nuanced patterns could add additional weight to the comments made by participants during the task scenarios.

6.7.9. Method of delivery

The think aloud protocol originated with in-person studies, where both the participant and researcher are in the same physical space. Due to the availability of video conferencing technology, it is now possible for think aloud studies to take place remotely.

Comparative studies such as those by Tullis et al. (2002) and Brush et al. (2004) identified that remote studies can be almost as effective as those in-person, however there may be increased cognitive time when participants are processing the activities and information presented. This situation may have improved as users have increased technological ability and become more familiar with remote video technology coupled with an increase in equipment capability however the lack of contemporaneous studies cannot corroborate these hypotheses. A survey of user experience professionals reported that controlled lab and remote studies are conducted to approximately the same extent, reflecting the understanding that both can deliver value to a research study (Fan et al., 2020, p. 89).

One of the main challenges reported by professionals is that of creating a 'comfortable and neutral environment' where participants feel able to speak freely about their experiences (Fan et al., 2020, p. 95). This is mitigated within this study by i) spending some unstructured time at the beginning of the study with the participant and ii) encouraging participants to verbalize whatever their concerns are. The researcher's disconnection with the website case studies further demonstrates the ability to freely voice concerns without appearing as a criticism of the hosting organisation.

The initial case studies in this research are conducted in-person, however later studies may be conducted via online means due to global events (see p. xxxiii). The method of delivery is annotated within the case study analyses, and it is assumed that this has negligible impact on the overall research outcomes.

6.7.10. Evaluation

The evaluation of qualitative data such as that collected during this study can be undertaken using various techniques. Klein & Myers (1999) provide seven key principles for interpretive field research to ensure that evaluation follows a structured framework without unstated assumptions (see Section 5.6.2.2). These are applied to this research by considering each piece of data individually and as a whole to build a larger holistic overview of the improvements that can be made to the case study websites [Principle 1]. This leads to an iterative approach, with each cycle of action case study research incorporating multiple cycles of data analysis. When analysing the data, the researcher considers the context of the information as this can have an impact on the ways in which users interact with the case study site [Principle 2].

The action case study cycles have been designed to conduct think aloud studies with ten participants each, and a key element of these activities is ensuring impartiality both during the activities and when analysing data

[Principle 3]. The think aloud studies provide some direction to the participant through task scenarios, though the researcher does not instruct elements of the task. Probing questions may be asked to further understand task completion though these are clarifying rather than directive in nature.

The cycles of action case studies have been designed to be based on existing information systems theory [Principle 4] though with an openness to discovering new insights from the participants [Principle 5]. This ensures that the influence from preconceptions about the concepts are minimised whilst linking to existing philosophical and research perspectives.

During the studies, the researcher may uncover conflicting narratives from participants [Principle 6]. These are further explored to discover the underlying concepts using probing questions, identifying the root cause of participants comments and information seeking behaviours. It is important to distinguish these from existing expectations such as those created through social distortions or preconceptions about the content and case studies [Principle 7], with mitigation provided by the researcher's role as both an observer and participant in action research studies.

As qualitative researchers often do not provide sufficient detail of their process between raw data and research conclusions by not including description of the data analysis and evaluation processes (Oates, 2005, p. 267), a worked example of the data analysis procedure is included in Section 8.5.15 on p. 241. The same process is applied to data collected during the other four action case study cycles, though for brevity is not included in this thesis.

The researcher will follow the cyclical process in Figure 6.19 for evaluating data produced during the task scenarios.

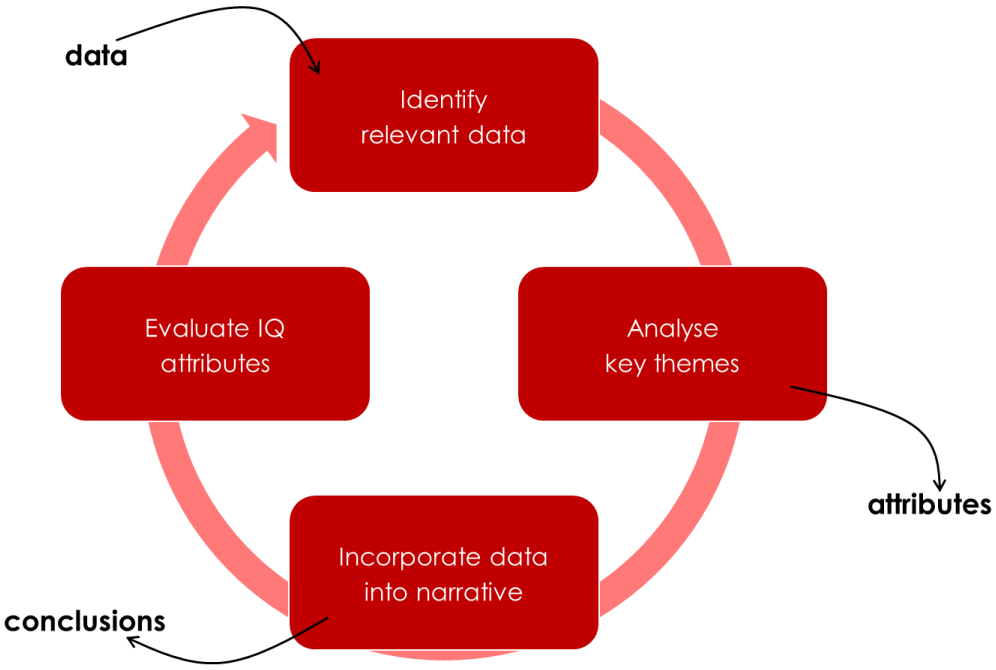


Figure 6.19: Data analysis cycle applied to task scenario data.

The researcher starts by reading the notes of a single participants' task scenario. This data is analysed to identify relevant outcomes for the study. These are then analysed to link the data to established and/or new potential information quality attributes as key themes. Once the relevant attributes have been identified, the data is incorporated into the narrative, partially documenting conclusions for this case study within a Microsoft Excel spreadsheet. An evaluation of the IQ attributes is performed to identify any areas with missing or additional data which can be completed using more outcomes from the same participant. This cycle continues until all data has been incorporated and categorised into the relevant information quality attributes. Once the data from one participant has been fully analysed, this process continues with the next participant. Reviewing observation notes is the most common form in which user experience professionals evaluate the outcomes of think aloud studies (Fan et al., 2020, p. 93).

By following a cyclical analysis, the researcher can incorporate data and ensure that themes are identified during the process. This builds on the observations of different participants who may describe similar phenomena

in opposing terms, allowing for the categorisation of similar themes within the same attributes. At the end of the process, attribute sets may be the same as the initial set in the prior literature, a subset or include additional considerations for the model.

One of the challenges often faced by professional user experience researchers is that of a lack of time to study the voluminous data produced by think aloud studies, leading to poor internal validity (Fan et al., 2020, p. 96). This is mitigated during the research as there are no commercial pressures from the case study organisations and the framework of the study is based on prior literature. This provides a grounding for the observations analysed throughout the case studies, in addition to the triangulation opportunities provided by subsequent cases.

Each case study is considered as an individual data set and analysed separately. This is due to the domain differences between the chosen case studies. The final validation case study follows the same process, incorporating the prior knowledge of the model to demonstrate the feasibility of improving information quality using the attribute-based approach.

6.7.11. Conclusion

The think aloud protocol provides an established technique for the evaluation of users' thoughts whilst using a system to achieve a task. Using task scenarios ensures that the participants are focusing on consistent areas of the site and conducting activities of a sufficient depth to explore the content fully. Concurrent think aloud techniques provide several benefits over interviews as they explore thought patterns over time, highlight usability issues with a small number of participants and are unbiased by the recency effect. Focusing on collecting data from representative users will allow the identification of information quality attributes used when exploring and accessing online content.

For the action case study cycles conducted during this study, the researcher follows the concurrent think aloud protocol, with participants completing a task scenario whilst both the task and their behaviour are observed. The researcher asks questions during the activity to provide contextualised responses, removing the need for a post-task interview that may provide unbalanced responses due to memory longevity and recency effects.

6.8. Research ethics

Online research can often post more problems for ethical researchers than offline studies, as ethical codes of conduct often lag behind technological possibility and participants rights may not be as transparent (Oates, 2005, p. 64). This study avoids many such concerns by only working with informed participants, seeking confirmation of their understanding of their rights within the project. This follows the codes of conduct published by both AoIR (2012) and BCS (2015) related to information systems studies. All ethical activities have been approved through the University of Salford Ethical Approval process.

Rogerson et al. (2019, p. 90) identify an additional ethical consideration for information systems researchers, the need to 'plan for a complex IS future'. This builds on the 'subtle nuances' often encountered within the IS field to consider whether research activities build towards a responsible system future. Whilst the intended future outcomes of a study cannot be fully anticipated, the research should be conducted with a view to identify any potential areas of ethical questionability. This may not be apparent within the formal research ethics sense; however, the issues should still be considered at every stage of study evaluation.

Documentation related to ethical approval of this research project are included in Appendix B.

6.9. Study plan

The overall study to develop the validated information quality model follows the plan shown in Figure 6.20.

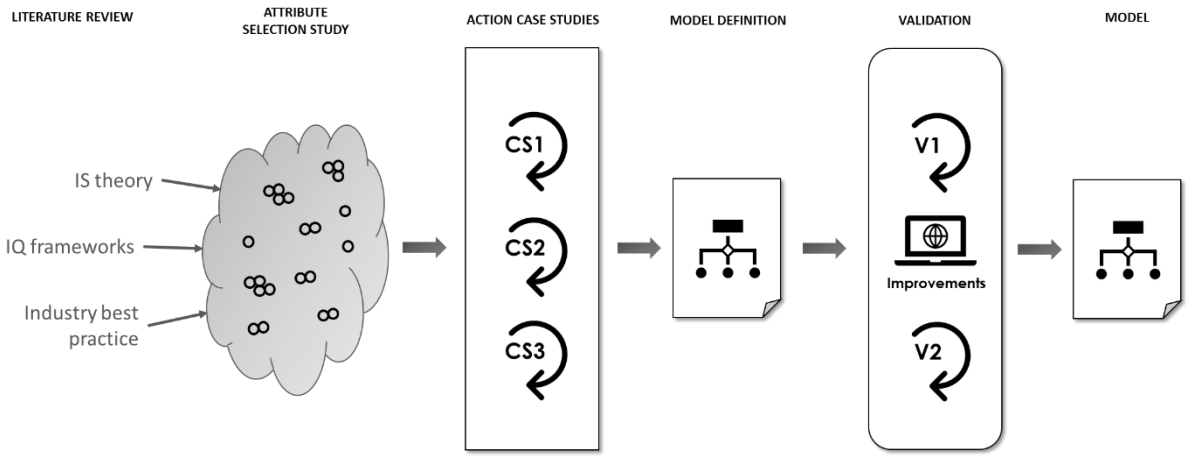


Figure 6.20: Overall study plan including action case studies leading to model development and further validation.

Firstly, the literature review informs the attribute selection study where the key existing information quality attributes within online content are identified. Following this study, three action case studies are performed on websites within separate fields to identify the information quality attributes inherent within online content. These case studies feature think aloud studies in a descriptive pattern with ten participants each, identifying attributes of online content that both help and prevent users from successfully accessing information contained within a website. The attribute set and approach for analysis is based on the literature reviewed during Chapters 2, 3, and 4. To ensure that this research produces generalisable results, the action cases are selected from a variety of domains to link with fields of prior literature and informative content for potentially different audiences. More explanation of the selection of individual case studies is included within Chapters 8–10 and 12.

After the action case studies, the practitioner model is defined, following industry best practice approaches to combine theoretical attributes with the attributes having the largest impact during the action case studies. This process is described in further detail in Section 11.4.

To ensure that the proposed model is effective for use within online contexts, a validation study is conducted on a fourth website. This consists of two action case cycles of explanatory concurrent think aloud activities, one before any changes are made to the website and one after the model has been applied to improve the quality of information. Improvements to the website are made based on an analysis of the content applying the model generated during the action case stage, with the two studies applying the think aloud protocol to analyse user perceptions of the website both before and after the changes have been made. A successful outcome is demonstrated by analysing the differences between participant commentaries and highlighting attributes where comments have adjusted in both volume and positivity.

During this entire research study, the model is derived and applied within four different informational website contexts. This demonstrates that the model is appropriate for use within a variety of domains and is sufficiently general for application with other websites.

6.10. Summary

This chapter reviewed the various methodologies available to the researcher to explore the key research questions of information quality within online web content. The research consists of an attribute selection study followed by three individual action case studies to identify the information quality attributes within various informative websites, leading to the definition of a practitioner model for online content improvement. The model is then validated using a fourth action case study by applying the model to an additional informative website.

The next chapter discusses the attribute selection study where key information quality attributes from existing literature are identified. This provides the initial attribute set for analysis within the subsequent three action case studies.

Chapter 7. IQ attribute selection study

This chapter describes the attribute selection study conducted to identify the key information quality attributes within existing academic theory as a basis for exploration during the subsequent action case studies.

7.1. Introduction

There have been many attempts within the information quality field to reduce the various aspects of content evaluation to a taxonomy of attributes which can be isolated and studied individually to suggest improvements. This approach helps to build a model of the various ways in which information may need improvement to meet quality needs and provide the opportunity for researchers to extend existing models to prove theories of improvement.

Despite earlier works such as the analysis by Zmud (1978), much of the contemporary literature uses the attributes identified by Wang & Strong (1996; see p. 95) as a starting point (Ge & Helfert, 2007, p. 11; Ge et al., 2011, p. 2). Their framework focuses on the social and economic impacts of poor data quality, categorising many attributes of data and shifting the focus away from studying purely accuracy. One potential reason for the relative success of this work may be its use of TDQM (see Section 4.3.2) based on Deming's (1986) Plan-Do-Check-Act (PDCA) improvement cycle as routinely applied within ISO standards. This is an approach to applying the attributes based on the application of cyclical and incremental improvements accessible to non-technical people involved in a project due to its use in management-level standards such as ISO 9001 (2015, p. viii).

One possible approach to investigating which information quality attributes have an impact on online content would be to focus on the most frequent terms within the prior literature, a technique applied within much of the prior

literature (including Ge et al., 2011; Kandari et al., 2011a, 2011b). Whilst this may be an easy approach to build on previous work, the results may not lead to a relevant or maximal impact on informative online content. For example, whilst timeliness is one of the most frequently occurring attributes, the specific nomenclature and definitions used in prior work varies widely, and the nuances of the previous studies may be missed by reinterpreting the meaning of this attribute to fit the researcher's purposes. In addition, the confusion between data quality and information quality within the IS field (see p. 63) reinforces the difficulty in interpreting the original meanings attributed by prior researchers.

7.2. Methodology

In their model of qualitative data analysis, Corbin & Strauss (2014) describe a three step process:

1. Collect a data set related to the challenges faced by a group of people.

2. Explore each component to identify relevant dimensions and attributes.

3. Use knowledge gained from studying each component to better understand the original problem and make inferences.

Recognising the challenge of evaluating a set of IQ frameworks from the literature, identifying relevant attributes, and using the knowledge generated to select those relevant to the problem outlined in Chapter 1 follows such an informed approach.

With the need to refine Corbin & Strauss' general methodology for evaluation into a specific methodology for selecting IQ attributes, the researcher refined this process as shown in Figure 7.1.

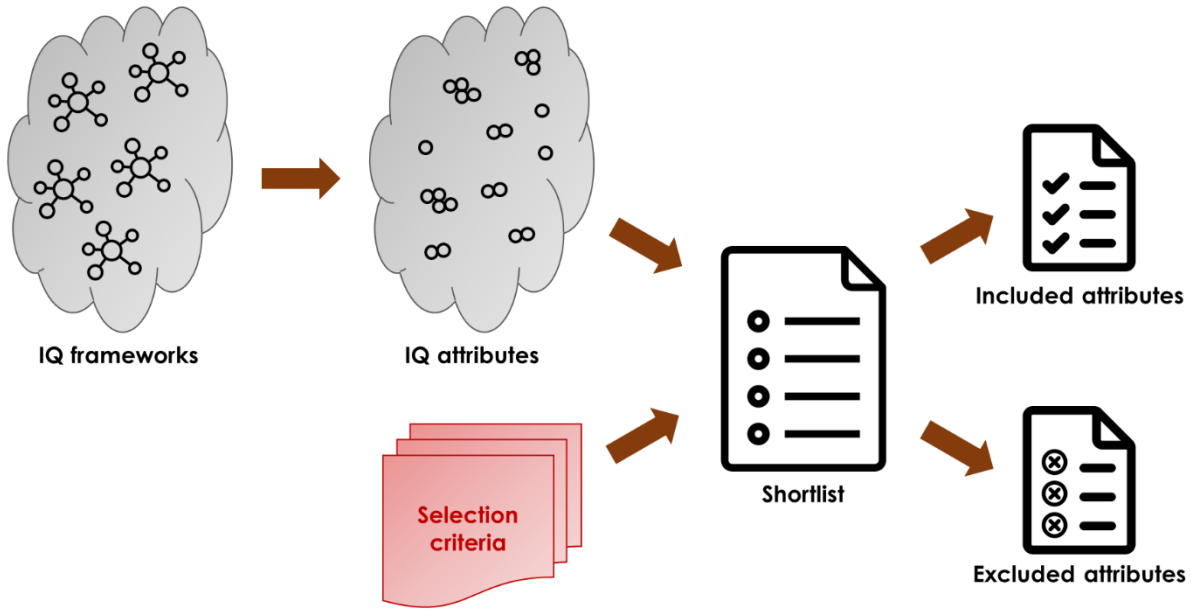


Figure 7.1: Attribute evaluation and selection methodology.

The first stage of this process is to select the data relevant to the problem. In this case, there is a need to select relevant IQ frameworks that could help inform the pattern of attributes for online content.

The second stage is to then reorganise the models into a collection of IQ attributes. Each framework contains multiple attributes in multiple dimensions; however, they need to be regrouped into a data set ready for evaluation.

The third stage of the process is to produce a shortlist of the IQ attributes relevant to the problem context. To aid with the selection, a set of selection criteria are required to classify whether an attribute may be appropriate to the study or not. Without such criteria, the researcher would be making guesses rather than informed judgements.

The final stage is to evaluate each of the shortlisted attributes in further detail, identifying those which should be included or excluded from the later action case studies. This requires a deeper understanding of the attributes and how they relate to online content quality.

Using the IQ frameworks reviewed in Chapter 4, the researcher collected similar IQ attributes ready for potential shortlisting based on the criteria outlined below.

7.2.1. Criteria

To determine the attributes for final inclusion in the set for study, the researcher established a set of five criteria for their evaluation, each of which are described below.

7.2.1.1. Criterion 1: Attributes must appear in more than one framework

The first criterion relates to the frequency of use within IQ frameworks. Attributes that appear in only one framework are unlikely to be candidates for an encompassing model as they have not been observed and validated across multiple studies.

7.2.1.2. Criterion 2: Attributes should not be synonymous to other included attributes

Clear definitions of attributes are important, leading to the need to conflate those similarly described in different frameworks. For example, the difference between *recency* and *timeliness* is contextual and the terms are often used to mean content provided contemporaneously in the frame of reference of the user. Including both attributes within the initial model would duplicate observations across multiple definitions.

7.2.1.3. Criterion 3: Attributes should have a clear definition within IS literature

This study is framed as an analysis of online content quality within the Information Systems space; therefore, attributes should be considered to fit entirely within the domain with appropriate definitions provided with links to existing IS theories. This grounds the attributes within existing knowledge.

7.2.1.4. Criterion 4: Attributes should be related to informative aspects of content

A key purpose of this study is to identify the attributes relating to informative content. Many existing IQ frameworks make no distinction between types of content and outcomes; therefore, an understanding of the attribute definitions is required to select those directly related to informativeness of content. Attributes proposed by data quality frameworks such as Wang & Strong (1996) may only apply to the analysis of data and not information (based on the definitions in Section 3.1.1).

7.2.1.5. Criterion 5: Attributes should be relevant for online content

This study is specifically identifying informative attributes for analysing online content, therefore only those relevant to a representative web-based audience should be considered. Aspects such as variety of data and access security are not appropriate for inclusion in the framework as they move beyond the bounds of web content authoring.

7.2.2. Attribute conflation

One specific challenge of combining information quality attributes from across many frameworks in different domains is the definition of the individual properties. Several attributes have similar or inter-related meanings dependent on the perspectives of the researchers, the field of study, and the underlying academic theory.

To minimise the number of inconsistent attributes, the researcher considered inter-related meanings between frameworks, and reduced these into a reduced set based on the discussions within the attribute descriptors later in this section.

7.2.3. Shortlist

The purpose of analysing the information quality attributes across key frameworks is to produce a shortlist of those to be included in this study. Whilst many authors such as Kandari et al. (2011a, b) take the approach of inclusion by frequency of appearance in other studies, this does not indicate their relative importance within specific domains. To select the attributes for inclusion in the study, a deeper approach to analyse the underlying academic literature and theories around each was required.

The main two pieces of literature to identify attributes for study are those of Wang & Strong (1996) and Kandari et al. (2011a, 2011b), the former due to its preeminent position as a seminal paper within the data and information quality research, and the latter as a review of further developments of IQ frameworks and attributes. Other literature as reviewed earlier in this chapter is incorporated into the subsequent analysis to further demonstrate the appropriate selection of attributes for inclusion in the action case studies later in this research.

Using the criteria outlined earlier in this section, the researcher identified 22 suitable candidates as the shortlist for consideration in the framework. These attributes, as shown in Table 7.1, comply with the five criteria and are suitable for further evaluation.

Included in model	Information quality attribute	Wang & Strong (1996) dimension	Kandari et al. (2011a, 2011b) frequency of appearance in IQ frameworks	Synonymous terms
	Accessibility	Accessibility	12/20	
	Accuracy	Intrinsic	17/20	
	Authority	—	10/20	
✔	Availability	—	19/20	Timeliness
	Believability	Intrinsic	12/20	
	Completeness	Contextual	16/20	
✔	Consistent representation	Representational	12/20	
	Credibility	Intrinsic	7/20	Reputation
✔	Language	—	6/20	Interpretability
✔	Level of detail	Contextual	n/c	
✔	Navigation	—	5/20	
✔	Presentation	Representational	n/c[2]	
	Quantity	Contextual	11/20	
	Readability	—	n/c	
✔	Recency	Contextual	18/20	
	Relevancy	Contextual	12/20	
	Reliability	Intrinsic	8/20	
✔	Search	—	n/c	
✔	Understandability	Representational	9/20	
	System usability	—	3/20	"Usability"
✔	Value added	Contextual	9/20	
✔	Visibility	—	n/c	

Table 7.1: Author's attribute mapping within relevant information quality frameworks.

[2] n/c Not considered within the study.

The table above identifies the initial shortlist of 22 information quality attributes selected for consideration, and their further status as included or excluded from the forthcoming studies. As several attributes relate to similar properties though with alternate nomenclature, the author has conflated these where appropriate, and included the synonymous terminology within the final column of the table.

Of the 22 attributes identified, 14 are derived directly from the work of Wang & Strong (1996). Their model proposed four dimensions for the attributes (intrinsic, contextual, representational, and accessibility), and the attributes for consideration cover all four areas due to their prominence in the literature and applicability to online contexts.

Out of these four dimensions, the most relevant attributes derive from the intrinsic category as these are properties of the information not the system. Contextual attributes may derive from the information; however, a more detailed analysis of the online content situation is required to determine whether they are suitable for further consideration.

The systematic review prepared by Kandari et al. (2011a, 2011b) features 17 out of the 22 shortlist attributes, reflecting the continual study of attributes within the data and information quality fields. Of these, the rankings shown in the table above denote the frequency in which these appear in other frameworks. This can give an approximate guide to their suitability for consideration; however, it is important to note that this is not necessarily a good measure of their suitability for further study. As noted earlier in this chapter, many IQ frameworks are derived for specialised purposes or focus on data rather than information, therefore the researcher needs to make an informed judgement into their appropriateness for this research.

The attributes in the table above that do not appear in either of the prior works have been derived from the researcher's widespread study of the information quality field, particularly its appropriateness to information as

processed data and online contexts. Specific details about their selection are discussed in the following two sections.

The attributes discussed below in Sections 7.3 and 7.4 were considered as part of this process, and each inclusion and exclusion is discussed and justified based on a deeper analysis of the relevant theories.

7.3. Included attributes

This section discusses the eleven attributes selected for inclusion in the shortlist due to their prominence, relevance, and importance within information quality literature. This demonstrates that they may be suitable choices for application within the future practitioner model.

7.3.1. Availability

The first attribute for consideration is *availability*. In their review of IQ frameworks, Kandari et al. (2011b, p. 22) identified this attribute within three eighths of the relevant literature included in systematised review.

Within IQ studies, availability often refers to the user's potential to access particular pieces of data or information required for their activity, with a focus on the ways in which the system may prevent them from completing their activity. Despite its low ranking in terms of frequency of appearance in IQ studies, availability should be included in the list of attributes for evaluation due to its crucial role in information seeking activities: if users cannot access what they need, they will be unable to complete their activities.

7.3.2. Consistent representation

As an attribute, consistent representation appears more frequently in the literature, with two thirds of existing IQ frameworks including its consideration (Kandari et al., 2011b, p. 22). Its regular appearance within data and information quality studies is to be expected as the attribute refers to the continual presentation of information within the same patterns, consistent

with prior user experiences. This is a crucial consideration within Human-Computer Interaction (HCI) studies, with consistent representation ensuring that users familiar with the system can continue to access the information they need without additional cognitive workload (Gorla et al., 2010).

Consistent representation is considered in several elements of the WCAG guidelines as well as a majority of frameworks reviewed in the literature, therefore it justifies further study with the action case research detailed later in this thesis.

7.3.3. Language

Language is rarely incorporated as an attribute within IQ studies, however the specific text to be studied is clearly an important consideration.

The Web Content Accessibility Guidelines (WCAG) 2.1 (W3C, 2018a) introduce language aspects within Guideline 3.1, Readable. This is mainly at a mechanical level where nomenclature and unusual terms are identified clearly, with some consideration of a typical user's reading level.

Early works with the IQ field rarely consider Language as it is an attribute of information rather than data, conveying meaning through the ways in which concepts are represented.

Many practitioners have identified the use of clear language as a key aspect of the success of online content (Matera et al., 2006), therefore this attribute is to be considered within the forthcoming stages of this study.

7.3.4. Level of detail

Level of detail is first considered by Wang and Strong (1996) as a contextual attribute of information quality, dependent on the user's perceptions of the content. This is one interpretation of the attribute, though the level of detail provided could also be related to the depth of information required to solve a problem. The forthcoming elements of this study will apply task scenarios in

the user research sessions; hence the level of detail is an important consideration to ensure that there is sufficient information for participants to answer the questions.

Information overload is an additional concern when analysing the appropriate level of detail within content. Ensuring that there is not too much volume or irrelevant information for the user to process is an important aspect of information quality (Eppler, 2015), therefore this attribute will be included in further study.

7.3.5. Navigation

The navigation attribute is related to the user's ability to find the information they need using the structures and links between different elements of content. Whilst this attribute was not considered by Wang & Strong (1996), it appears in a quarter of later studies, demonstrating its suitability for consideration as research moved away from data to information quality.

There is debate amongst the literature as to whether navigation is considered to be an attribute of information quality or system quality, and this is not resolved by practitioner definitions. As this attribute can be included as either a property of the system or a property of the content, it warrants further investigation as part of this research.

7.3.6. Presentation

In their original work, Wang & Strong (1996) considered presentation to be a representational aspect of information quality. This means that the ways in which content are displayed to the user may affect the way in which quality is perceived.

Presentation can be considered as two different though related concepts:

1. A property of the system, without impacting on the information contained within, or,

2. A property of the information, considering the ways in which the content has been authored, structured, and added to the site.

If we elect to use the second definition and consider the presentational aspects of the information as authored, including structuring of textual content, then this can be further studied as part of this research.

7.3.7. Recency

The attribute recency relates to two similar attributes, both recency and timeliness.

Recency is often taken to mean that data or information presented by a system must be up to date, however there are few ways for a user to ascertain this within a typical website. One key way recency is surfaced to users is through the use of "last updated" timestamps to demonstrate when content has been authored. This is most prevalent on news or current affairs websites, though such dates can also be displayed on informative websites to annotate how long the information contained within has been published.

Timeless is another attribute that can have different meanings to different users and organisational contexts. Klein (2002, p. 1103) includes currency of information, technical limitations, and unknown publication date as three aspects of this attribute that may cause poor user perceptions of web content. There is often a trade-off within information systems between timeliness and accuracy (Ballou & Pazer, 1995).

As both recency and timeliness are properties of the information and have an impact on user perceptions of information quality, they will be further studied during this research to identify if they have an impact on informative web content.

7.3.8. Search

The use of search engines to find information online has become ubiquitous in everyday life, with the expectation that websites provide search features. Searching has long been considered as a primary information seeking behaviour online (Hertzum & Pejtersen, 2000; Bates, 2002), with the mechanisms and ability to use search features within large volumes of online content having a large impact on users' abilities to complete information seeking activities.

Although search is not considered as a core information quality attribute within the prior literature, its important role in online information seeking behaviours demonstrates why such features need to be considered as part of the holistic information quality and user experience during this study.

7.3.9. Understandability

The understandability of information within a system has often been considered as an aspect within IQ frameworks, appearing in 9 out of 20 studied by Kandari et al. (2011a, 2011b). Classified by Wang & Strong (1996) as a representational attribute, this is often linked to the clarity of information provided, without ambiguities or comprehension issues.

Closely linked to attributes such as language and readability, understandability is more contextual and can be affected by the ways in which information is presented to the user; however, the definition of this attribute goes beyond the mechanics of the language to the choices made when content is authored. The WCAG 2.1 guidelines (see Section 2.4.7) move beyond the readability of content to suggest that information and navigation elements should be predictable for the user. This approach to help users by providing easy-to-understand and consistent information has led to the development of large-scale, cross-organisational knowledge bases such as the GDS Design System to aid content authors in producing information that is fit for purpose.

Due to understandability having an impact on the information being produced through consistent and easy-to-access content authoring, this attribute will be included in the next stages of this study.

7.3.10. Value added

Value added was first considered as a data quality attribute and has been linked within the PSP/IQ model to the dimension of *usable information* (see Section 4.3.4). Further studies within the healthcare domain demonstrated value-added to be the largest indicator of usable information within the systems studied to the extent that this caused distortion within the comparison of other dimensions (Kahn et al., 2002, pp. 189–190).

Despite its appearance in Wang & Strong's (1996) initial IQ framework, the concept of value added as a predictor of quality within Information Systems predates the modern web, with the first suggestions of the attribute appearing in Taylor's (1986) generalised framework. Furthermore, it can be deduced that the value added aspect can be considered as related to the data or information of the system (Rieh & Belkin, 1998, p. 3).

Even as value added can be considered as a contextual IQ attribute, the value can be derived from either the context of the system and user or be present in the information itself. Therefore, this attribute will be included in the future stages of this research.

7.3.11. Visibility

Visibility as an IQ attribute was not included in either of the antecedent works feeding into Table 7.1. Despite this lack of recognition, visibility can be an important aspect within online content as information not immediately available to the user may be missed and lead to longer information seeking behaviours or failure to complete a task. An example where hidden information could limit a user's ability to complete a common task is the retrieval of encyclopædic data such as dates and places of birth and death

often included as key pieces of information highlighted in a prominent place in the relevant entry (Stvilia et al., 2005, p. 109).

Often the visibility of information could be considered as a system quality issue (due to presentational aspects), however there is also a need to evaluate how information is surfaced within text. Marketing specialists recognise the need to write copy that evokes a particular response within users, acknowledging that the same information presented in different formats can become more or less accessible and change outcomes within information seeking behaviours. This leads to visibility's inclusion as a potential attribute for further study.

7.4. Excluded Attributes

The following eleven attributes were evaluated from the shortlist by the researcher though found to not be suitable for consideration during this study. The main reason for excluding these elements is the definition of what quality aspects are directly related to the content, and which may be attributed to the system or interface in which the content is provided.

7.4.1. Accessibility

Within a survey of industry practitioners (Yesilada et al., 2012), the most popular definition of web accessibility was that provided by the Web Accessibility Initiative (WAI):

> "Web accessibility means that websites, tools, and technologies are designed and developed so that people with disabilities can use them"
>
> (W3C, 2005)

This definition is applied across industry standards and guidelines, including WCAG on which many practitioners rely to assess and rate the accessibility of their overall websites (see Section 2.4.7).

Considering the elements lacking from this definition, Petrie et al. (2015) reviewed 50 academic works to identify the six core concepts of web accessibility. This study concluded that the most frequently missed constructs relate to the consideration of all users, not just those with disabilities and the ability to interact with websites. This leads to their inclusive and encompassing definition:

> "all people, particularly disabled and older people, can use websites in a range of contexts of use, including mainstream and assistive technologies; to achieve this, websites need to be designed and developed to support usability across these contexts"
>
> (Petrie et al., 2015, p. 3)

This definition supports and widens accessibility to cover all users with any physical or mental issues with using technologies and demonstrates a constructive approach to accounting for web systems use.

Some works such as Thatcher et al. (2003) propose that accessibility is a subset of usability, building on Shneiderman's (2000) concept of *universal usability* to incorporate both attributes. Within the review of existing IQ frameworks, it is important to separate these two attributes as accessibility relates to access to information, and usability to the interaction methods of the system which contains the information.

Within their research, Petrie & Kheir (2007, p. 399) determined that 'disabled and non-disabled people often encounter the same problems but are affected by them differently'. This leads to the conclusion that accessibility issues are applicable to all yet amplified for those who are less able to adapt their style of web use when the system design is not sufficient for their use. Earlier texts about web technologies separate accessibility and catering for those with disabilities as a peripheral inclusion. This trend has shifted, with contemporary literature incorporating accessibility as an inclusive property (Ellcessor, 2014). There are many myths that accessible web design limits creativity, with boring, ugly and text-heavy websites the supposed result (Ellcessor, 2014; Lawson, 2006; Slatin & Rush, 2002). Inclusion should be a highly

valued aspect of accessibility, even though the current guidelines such as WCAG 2.0 may be unhelpful for people outside of the professional development community (Adam & Kreps, 2009).

Online accessibility has been an often-explored topic within both academic and industry-based literature, with practitioners focusing on the legal and moral reasons to improve websites for universal access.

Within this research, accessibility has been excluded from the list of attributes for further study. The concepts of accessibility have been covered in detail by frameworks such as WCAG, with laws such as the Public Sector Web Accessibility Guidelines driving practitioners towards its adoption. Accessibility is often related to the system in which the content is contained, with only six success criteria in WCAG 2.1 (W3C, 2018a) relating to text content. These are all contained within the guideline of readable pages, with the note that this applies as 'information and the operation of user interface must be understandable' (W3C, 2018a), thus linking to the properties of the system rather than the content.

7.4.2. Accuracy

'Accurate information reflects the underlying reality' (H. Miller, 1996) and is important to allow users to base decisions on correct data and develop understanding that is consistent across different web pages and sites. Information used for different purposes can require different levels of accuracy depending on the context, and information designers need to consider the use case in order to reflect the most appropriate level of accuracy within online content.

When considering online content, users do not know what information is accurate and what is not, as demonstrated by the rise in focus on 'fake news' designed to target and mislead specific audiences (Gelfert, 2018). This leads users towards trustworthy sites, those that are known to be provided by reputable sources.

With this study focusing on informative websites, there is a need to consider the types of user activities and how these are configured into the forthcoming methodology. Studying websites provided by unknown sources with unknown reputation may lead to the participants questioning the source of the content, undermining the properties of the content itself. Therefore, accuracy cannot be included within the information quality framework as this would introduce doubt into the informativeness aspects in the underlying textual content, introducing an additional variable of study the researcher cannot control. This must be avoided to ensure that the model produced as an output of this study focuses on informativeness of content on any site, not those provided by recognisable or reputable organisations.

7.4.3. Authority

Within the information systems field, authority is often considered to be related to the responsibility for the content provided. Wilson (1983) identifies two different types of authority:

1. **Administrative authority**
   Recognising the rights of someone in a position.

2. **Cognitive authority**
   The individual's knowledgeability and credibility based on another's perception.

Whilst many information retrieval studies focus on the latter definition (Rieh, 2002, p. 147), in practice it is challenging to separate these definitions leading to both types of authority being concurrently studied within information quality assessment. Users often base assessments of authority on source criteria, judging the URL components or name of a known institution (Rieh & Belkin, 1998).

The concept of authority is linked to that of accuracy; therefore, the same considerations must be considered resulting in the exclusion of this attribute

from further study. The authority of a website is based on several aspects including the user's prior experiences and the accessibility and usability of the system, all factors outside the scope of textual content provided by the content author.

7.4.4. Believability

When finding information, users need to be able to believe the textual content they read online. Kandari et al. (2011b) recognised that believability is a common information quality attribute, appearing in two thirds of frameworks they reviewed.

An important consideration of believability is that this does not necessarily guarantee the information is accurate, rather the user needs to trust what they have obtained. For a user, trusting a source of information online is crucial to processing and believing in the content contained within a website (Chopra & Wallace, 2003). Freedom of choice and the willingness for users to conduct activities based on the content found are key to using digital information online, as is the perceived confidence of the text itself (Kelton et al., 2008).

Within the IS literature, researchers consider believability to be related to information in two different ways (Pradhan, 2005):

1. An *intrinsic* property contained within the information and always inherit no matter how presented, and,

2. A *contextual* property, where the same information presented in alternate ways can affect the ways it is perceived by users.

If we consider the first definition, then the believability of a collection of information would be fixed, no matter where it is placed and how it is accessed. The context of content such as reviews and recommendations has been demonstrated to have a large impact on its credibility and subsequent

believability (Cheung et al., 2009), therefore this attribute cannot be considered in the study as it is a property of the system not the text.

7.4.5. Completeness

This attribute is related to accuracy and quantity and assesses whether users believe that the resource contains all the information required for their purposes. Completeness has been used as a factor in many information quality studies since Wang & Strong (1996), and 'is a characteristic of the message content' (Dutta-Bergman, 2004, p. 254) with portrayal in a comprehensive, balanced and adequate manner (Dutta-Bergman, 2004, p. 256).

The challenge of analysing completeness of content is that this is dependent on the perspectives of the user: what appears complete to one individual may not to another, with the context of the activity driving its consideration. For example, in the AIMQ framework the authors note that this is considered a contextual dimension which needs to be evaluated based upon the specific web system being studied (Lee et al., 2002).

7.4.6. Credibility

The assessment of website credibility often involves two key dimensions: trustworthiness and expertise (Rieh, 2010, p. 1337). Users can confuse these concepts, leading to an assumption that a website with authority automatically has credibility (Rieh & Belkin, 1998).

A recent example of this is the propagation of fake news by respected news websites, claimed to be a result of the evolution of online platforms and shifting business models (Newman et al., 2017). Users' assessment of online content credibility is much harder than in offline spaces due to 'the multiplicity of sources embedded in the numerous layers of online dissemination of content' (Sundar, 2008, p. 74). By providing tools for credibility assessment, a filter-based approach can help provide users with

more believable content (Flanagin & Metzger, 2000) and alleviate the need to trust traditional gatekeepers such as search engines in providing the best content (Pan et al., 2007).

Hargittai et al. (2010) suggest that the ultimate tool in ensuring accurate assessment of credibility is educating users prior to their information seeking activities, however this negates the intended audience for this study. It cannot be assumed that any participants have any level of prior experience or knowledge during the user research stages as this would introduce unnecessary bias into the interpretation of information. As the researcher does not want to introduce any unintended biases or preconceptions into user behaviours, the credibility of information needs to be discounted in this study.

7.4.7. Quantity

The quantity of information first appears as a dimension of information quality within Zmud's (1978, pp. 189–190) framework, where data is assessed within a four-step scale: completeness, effectiveness, materiality, and sufficiency. This measure is further developed in Wang & Strong's (1996, p. 32) model as the appropriate amount of data, reflecting the need for varying quantities of information to fit particular contexts.

Kandari et al. (2011a, 2011b) introduces analysis of the amount of data through a series of questions reflecting the appropriateness of the online information provided to the user's particular task. This is unique to this survey, as prior questionnaires such as WebQual 4.0 (Barnes & Vidgen, 2000) do not include this dimension despite the effects of user information overload within IS being studied since the late 1980s (Lesk, 1989; Jacob Nielsen, 1995) and being recognised as an important factor in technology adoption throughout history (Interaction Design Foundation, 2018).

Reducing information overload within online systems is principally the responsibility of information architects (Davis, 2011, p. 46), and such an improvement to a website can help to reduce the extraneous cognitive load through meaningful design (Schmutz et al., 2009, p. 2). As the amount of information accessible to users becomes greater, this challenge is increasing in importance and the effects of not addressing information overload are becoming more apparent commercially and with the success of online systems (Morville & Rosenfeld, 2006, p. 54).

Whilst content authors do need to monitor the level of information provided within their websites to ensure that there is no copious repetition, quantity can be linked more closely to the earlier data quality frameworks where additional pieces of data could increase confidence in outputs. With information studies, the reverse is often the case: too much data leads to user overload and subsequent issues finding answers to questions. Therefore, this attribute is excluded from this research as it relates more closely to a property of the system and unprocessed data than to the study of the informativeness of online web content.

7.4.8. Readability

Within the HCI field, several metrics have been developed for perceiving the reading difficulty of a set of text. One of the oldest and most prevalent is that devised by Flesch (1948) based on the characteristics of the sentences, words and syllables used within content. As a quantitative measure, the Flesch algorithm delivers a numeric result based on the expected reading age of the user and presenting a value for comparison to other texts. This style of measure does not take the user's understanding of the content into account, as the specific words and terminology used may impact on the readability, as well as the idea density and human interest within the text (Simons, 1971).

The integration of these concepts led to the need for an interpretivistic approach to readability assessment. By examining the perceived readability of content, a model of the content can be produced that provides feedback to content authors whilst respecting the cognitive patterns of users (Janan & Wray, 2013).

To a limited extent, readability of content is already covered by WCAG 2.1 (W3C, 2018a) with the consideration of unusual words, abbreviations and reading level (at the highest compliance rating, Level AAA). This attribute is linked to the attribute of Language (see Section 7.3.3), however shifts the focus towards the intended audience of the content. As this study intends to determine the information quality attributes that may have the largest impact on informativeness of online content for any user within the intended audience base, measures such as reading level add an extra level of complexity and may reflect more on the profile of the participants rather than inherent properties of the information contained within the system. Therefore, the attribute of readability is excluded from inclusion in the model.

7.4.9. Relevancy

When seeking online information, users expect to find content that is relevant to their topic. This can be influenced by the content being structured, featuring the most appropriate keywords and being readable by the audience, providing applicable and helpful information (Nurse et al., 2011, p. 493). In their systematic review of existing IQ frameworks, Kandari et al. (2011b) identified relevancy as an attribute within two thirds of the literature studied.

Providing relevancy requires understanding of the user's context and their perceived model of the content. This is most relevant to data quality, where users are left to process the content themselves to produce their own data. As a generalised practitioner model for informativeness, the output of this thesis needs to be applicable for improvements for multiple audiences

without relying on their preconceived biases of the content. If users existing ideas of the information are considered in the framework, then the advice cannot be universal, therefore this attribute has been excluded from further study.

7.4.10. Reliability

The reliability of online content has been explored by many prior studies, including many within the health field. Information reliability covers the information content and expected end-user behaviour (Adams, 2010, p. 391). Users need to be able to trust the information they find online to make decisions and answer their questions.

In their review of IQ attributes, Lee et al. (2002) categorise reliability as both an intrinsic and an accessibility concern, placing its importance across both the properties of the information within the system and properties of the system itself. The second categorisation falls outside the scope of the model for development by this research as it relates to system rather than information quality aspects, though the consideration as an intrinsic IQ property is within the scope of the quality of information itself. This is highlighted by Ghasemaghaei & Hassanein (2015, p. 966) who note that reliability is a property of the information within a system and does not have a large impact on user satisfaction within non-informative websites such as those developed for e-commerce purposes.

The key challenge when considering intrinsic reliability of online content is how users form this perception. Reliability as an attribute is closely linked to trustworthiness and credibility, with user opinions informed by non-informational properties such as organisational recognition, system design factors, and external reputation. As such, reliability as an intrinsic property of information quality depends on much more than the textual content within the system, excluding the attribute from further study.

7.4.11. System usability

Usability has often been included as a software quality attribute since the late 1970s when human-computer interaction emerged as a field (Ferre et al., 2005, p. 202). Just as with traditional software systems, web systems rely heavily on positive user experiences providing a competitive edge for use (Donahue, 2001). Even with a history of literature examining usability, it is rarely examined in an isolated context due to the term referring 'to both a set of independent quality attributes such as user performance, satisfaction and learnability, or all at once' (Seffah & Metzker, 2004, p. 72).

Many small software development teams avoid the direct involvement of usability experts by applying a combination of design guidelines, patterns and diagrammatic approaches to enhance the functional specification (Seffah & Metzker, 2004, p. 73). Software development processes often implement usability assessments late in the cycle, when it becomes difficult to incorporate and fix (Juristo et al., 2001, p. 21). This highlights the need to include usability as a key attribute of overall quality at all levels of activity, and for all members of the team to have insight into how positive experiences can be created.

Whilst this attribute is often referred to as *usability*, it should be considered as *system usability*, a property of the system itself rather than the information contained within. This is a key distinction when analysing users' perceptions of usability, as opinions are often based on appearances and ease-of-use of the system and its navigation rather than the underlying information to be presented to the user.

This is consistent with the DeLone-McLean model of Information Systems success (see Section 3.2.4) where the aspects of system quality and information quality are clearly delineated, providing boundaries for HCI and IQ research. Consequently, system usability must be discounted as an

attribute for further study as it is not directly related to the quality of online content.

7.5. Conclusion

Analysing existing IQ frameworks led to the identification of 11 information quality attributes that appear in the academic literature and can be considered as core for the analysis of web content. These will form the basis for the action case studies discussed later in this thesis.

The inclusion or exclusion of attributes depends on factors such as the context of use, definition of nomenclature, and domain which is being studied. As this research intends to produce a general-purpose framework suitable for use across informative content types, a generalised approach has been undertaken to synthesise similar terms into single attributes, simplifying the analysis set.

With the determined list of eleven preliminary information quality attributes for inclusion in the next stage of the study, the researcher will determine their impact on user information seeking behaviour and identify any remaining attributes not apparent in the literature that may influence website content informativeness.

The next chapter discusses the first action case study analysing informative content within a university website, leading to a refinement of the information quality attributes identified within this Chapter.

Chapter 8. Case study 1 – University website

This chapter provides a narrative for the first action case study, analysing user think aloud studies applied to informative content within a university's website.

8.1. Introduction

This case study represents the first cycle of action research as outlined in Section 6.9, performing the first analysis of user think aloud studies. Due to the action research nature of the project, this will inform later stages of the study, providing initial results and direction for task scenarios.

8.2. Rationale

As the first task scenario, the University website provides a useful starting point of a mostly informational website containing resources targeted at a wide range of audiences. As participants for the studies are recruited from within a higher education environment, a familiarity with the concepts of the organisation removes the need for users to fully learn about the domain purely for answering the researcher's questions, with responses focusing on the information provided rather than creation of specific knowledge.

8.3. Methodology

The first task scenario examines the information quality of a university's student-facing website. This case study allows the exploration of several key themes with a general audience with a variety of skill levels. None of the participants are expected to be expert users, however they may have some domain-specific knowledge that can be applied to the context. In addition, the researcher has access to the case study organisation, facilitating the feedback of ideas to improve the website.

In accordance with the principles for successful think aloud studies, the researcher set up the workstation as indicated in Figure 8.1.

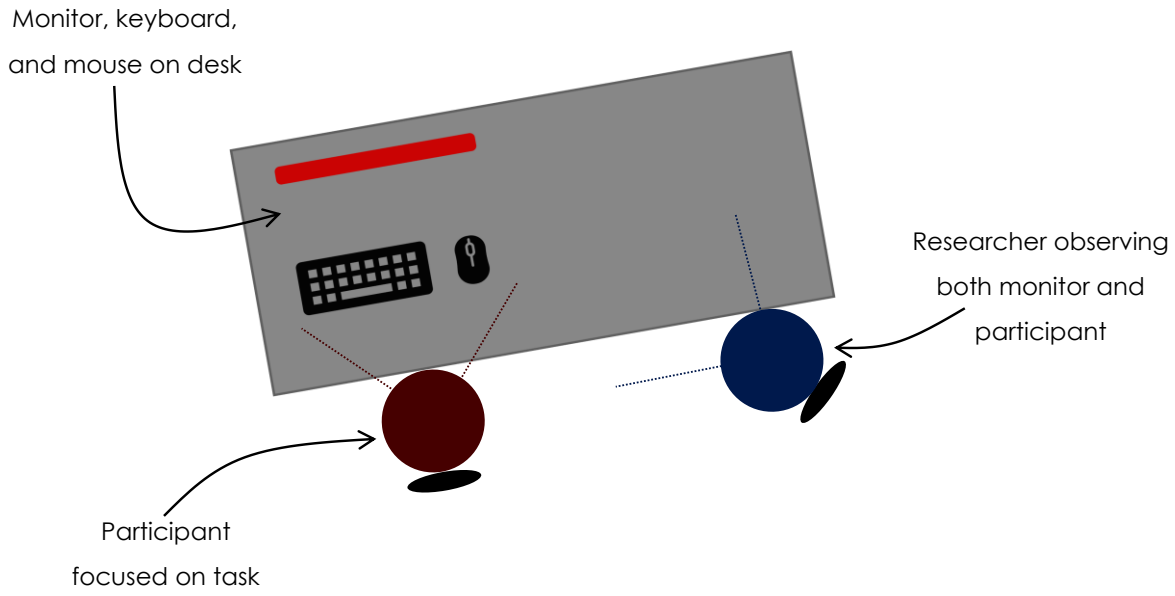


Figure 8.1: Workstation setup for think aloud studies.

The research sessions were scheduled to take place in an empty IT lab, providing a quiet environment for the participant. Prior to beginning the studies, the researcher checked with the individuals that they were happy with the computer setup, assisting with adjusting the workstation as necessary to easily interact with the case study website.

Whilst the workstation prioritised the focus of the participant, the researcher positioned themselves at the side where both the screen and the user could be observed at the same time without providing a distraction.

Prior to the think aloud activity, the following pre-activity brief was given to the participants:

Thank you for agreeing to take part in this research study.

You can explore the website for five minutes prior to the study, after which you will be asked to complete a scenario whilst talking about your thoughts and actions.

Website: <University website URL>

Figure 8.2: Think aloud study pre-activity brief.

By exploring the site prior to the study, participants become familiar with the navigational and structural elements, reducing their impact on the overall usability of the pages. This focuses the usability issues on the unknown elements, particularly the information contained within the site.

After five minutes exploring the website for background information, the following scenario is given to the participants:

Now that you have seen the <University> website, please follow the below scenario as the user, talking about your thoughts and actions. This may take between 15 and 30 minutes to complete. You can say anything that comes to mind whilst using the website, whether you think it is relevant or not. The researcher will only ask you questions as prompts for further information.

Imagine that you are a prospective student, looking to apply for an undergraduate degree course in Business, with a particular focus on technology at <University>.

1. Find a business degree course that has a focus on digital technology.

2. What are the required grades for the course?

3. How would you apply for the course?

4. Imagine you are a student with dyslexia. Find out what support would be available to you on the course, and how you would apply for this support.

Figure 8.3: Action case study task scenario.

This scenario is designed to explore two key areas of the site:

1. Course finder – This is an area of the website designed mainly for marketing purposes and features many elements of textual and visual content arranged in hierarchical structures. These are commonly found in informational websites, especially those with large quantities of difficult-to-navigate material.

2. Student support – The areas of the website detailing this purpose are separate to the marketing site, with many cross links and integration although the answers are in a separate area. The support content is more text heavy with fewer images. This resembles traditional information-rich websites that contain dense amounts of text, with the user expected to perform their own searches to access the snippets required for their task.

Section 8.5 discusses responses from initial participants and outlines the direction for the remainder of this cycle.

8.4. Information architecture

The information architecture of the website is key to the users' information seeking behaviour, defining its importance in the activity. This is explored below in both hierarchical and visual formats.

8.4.1. Hierarchical architecture

The hierarchical information architecture is an important element of this scenario as it makes assumptions about the users' view of the website and the ways in which content can be accessed. The diagram in Figure 8.4 shows the layout of the pages featured during this study.

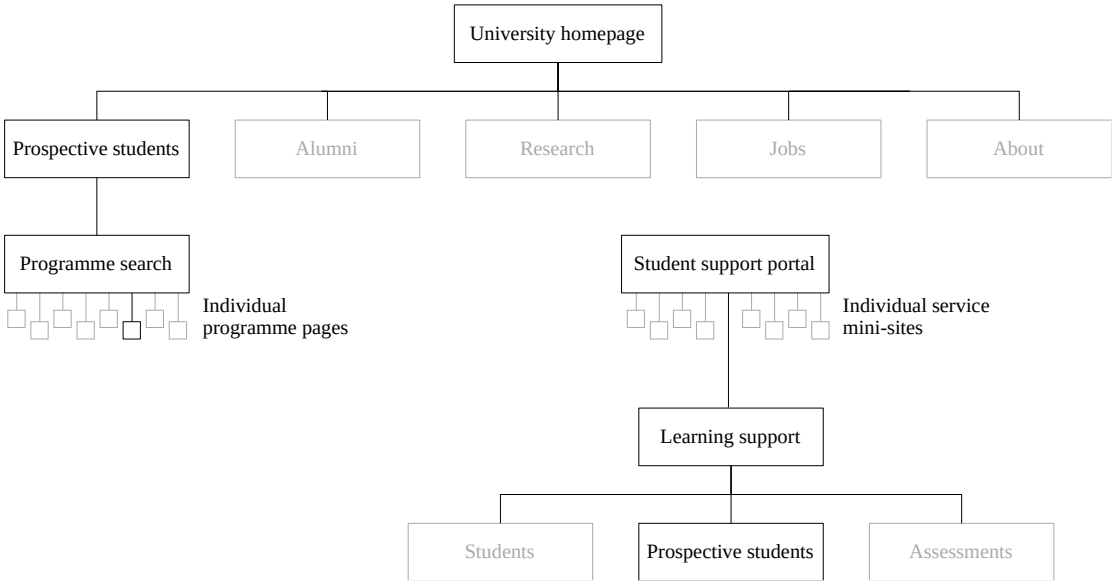


Figure 8.4: University website hierarchical architecture, featuring the element explored during this scenario. Bold pages are expected to be explored during activity, whereas greyed-out pages may be visited during the information seeking session.

The pages explored during the scenario cover many different elements of the site. Even though users do not enter through the homepage at the same rate as they used to during early years of the web, this is still an important landing point to be considered. Ensuring that users can easily navigate page structures will have a large impact on their future achievements within the scenario.

The scenario leads students to visit many different pages within the main University website and a sub-site: the student portal. Both sites have distinct changes in visual and structural elements, leading to users needing to switch their views of the information during the task. This is an important element of the scenario as it ensures that the attributes observed are apparent across multiple styles of website. In addition, this simulates online browsing behaviour where users may navigate multiple sites during their information seeking session.

8.4.2. Visual architecture

The visual architecture of a typical course page includes many elements in a linear pattern, as shown in Figure 8.5.

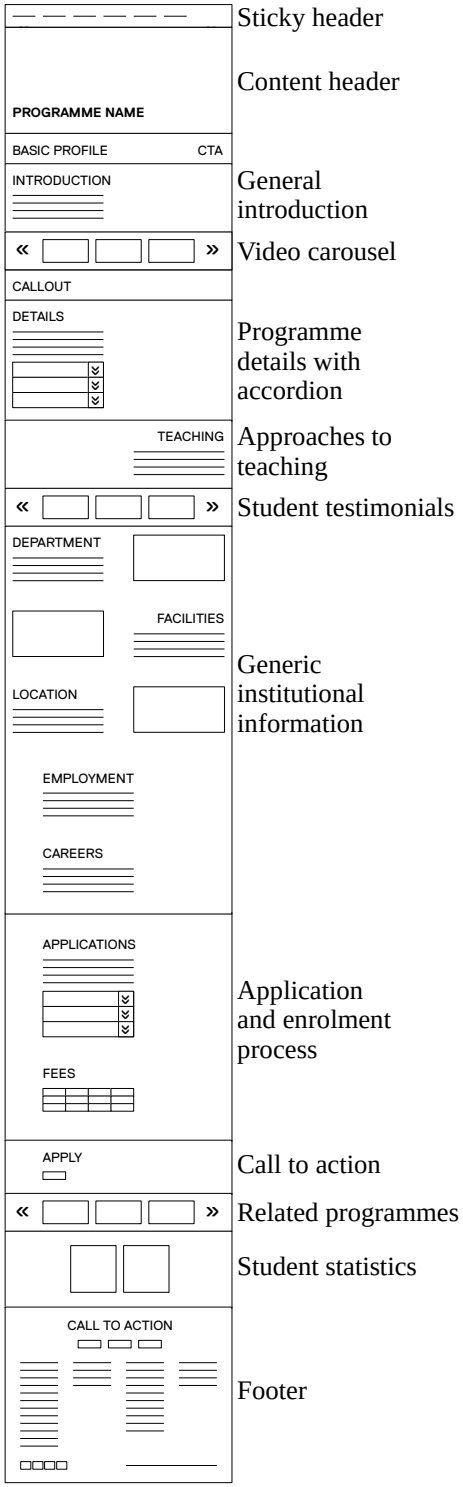


Figure 8.5: Typical programme page wireframe, incorporating navigational and content elements.

The linear page requires the user to scroll between the elements and follow the designers' pattern. This makes many assumptions about the users' view of the page that may not fit their model of the information.

The information quality attributes discussed below relate to the use of pages such as that shown in Figure 8.5.

8.5. Responses

This section contains a summary of the participant responses and observations from the think aloud studies, categorised by the researcher into their relevant information quality attributes.

In order to classify each response after the conclusion of the studies, the researcher analysed both the utterances and observations noted during the sessions in turn, sorting the outcomes into the relevant information quality attributes as identified in Chapter 7. Where data items fit into more than one attribute, they are annotated in either the closest match or in both as decided by the researcher based on their prior knowledge of IQ frameworks from the literature review.

Whilst classifying the data, the researcher identified three new attributes not included in the initial set: fallback, information usability, and interactivity. These are discussed below and in further detail in Section 11.3.2.

8.5.1. Availability

The task scenario was designed to include areas of the site with known information, although some of the content appeared to not be available due to the navigational approaches of the participants. This was a particular issue when searching needed exact terminology [P1.1] and choosing navigational items within the quick links section [P1.2]. The site was slow to load some course pages for P1.7, who considered the site to be broken and attempted alternative navigation paths.

8.5.2. Consistent representation

The information required to complete the tasks necessitates the use of two different subsites. These both have different visual and content representations, leading to users needing to re-learn the navigation patterns and style when transferring. This led to a pause in the activity for participants whilst they reassessed the page and current task [P1.1, P1.2, P1.7]. This is highlighted by the hierarchical structure diagram included in Figure 8.4, where the two sites are shown as a pair of linked tree structures.

The specific images selected to support content areas of the site were questioned by several participants as they did not feel that they were consistent with the message of the page. For example, on one page the image was considered 'totally irrelevant' to the content [P1.6], with another illustrating a different yet related concept [P1.3]. Although the use of visual content representation has been excluded from this study, the selection of such imagery can have an impact on the user's perception and interpretation of the accompanying text.

8.5.3. Fallback

After completing the activities, one participant mentioned that they would have preferred to find out the information by phone rather than online [P1.1]. Although this appears outside the scope of this study, the lack of assurance in the content led to the feeling that a fallback was required. They noted that there was no telephone number on the website to speak to someone about applying, however upon review this is in a non-conventional location on the page.

8.5.4. Information usability

Within the review of existing Information Quality frameworks and attributes, *system usability*, often known as *usability*, was excluded from the shortlist for further study (see p. 220). Another type of usability that can be considered

within the context of informative web content is that of *information usability*. Rather than focusing on the aspects of the system that provide easier access to content (such as those commonly categorised as Human-Computer Interaction), there are also elements of *information usability*, i.e., the ability of the user to access the information they need, and the features of the content that can facilitate this process.

During the think aloud studies, participant [P1.1] noted their desire to receive the information in alternative formats, and this could be considered as a workaround to information usability aspects: for example, the desire to have information restructured to match their expected model. Another participant mentioned that the information on courses for prospective applicants did not include the desired content for the task scenario, however the task could be completed using information in another section [P1.3].

Due to the identification of these usability aspects of the content rather than the system, the researcher will continue to identify any elements in future case studies that could be classified as caused by information usability issues.

8.5.5. Interactivity

During the navigation of the website, many interactive features added by the developers to increase the visual appeal were noted to detract from users' access to content. For example, P1.7 noticed that the dropdown menus jumped when scrolled by the user – an unintended consequence of overriding the default browser behaviour.

The developers of the website have replaced the default mouse pointer with call to actions describing some of the less obvious interaction options such as drag and play video. Rather than addressing the lack of affordability (Norman, 1998), this caused the use of a mouse to be 'very distracting' [P1.6].

P1.4 described these additions as 'too flashy and too much going on, especially as I'm dyslexic', noting that such features may be particularly

detrimental to certain groups of users. As these features can hide information, this caused the participant to panic and wonder where the content was.

8.5.6. Language

Participant 4.1 was unsure of some of the terminology related to university websites, such as the meaning of the words *undergraduate*, *postgraduate* and *alumni*. This added an extra complexity in the task, as they had to work out the meaning based on the context and a process of elimination. In addition, the use of acronyms such as *UCAS* with no prior explanation caused a barrier for understanding information [P1.1]. This necessity for domain knowledge to access the information prevents access to the content for some users.

Language was also an observed issue when searching for content, as the specific spelling of words had to be entered, matching just the titles of pages and not content [P1.1]. A more flexible search algorithm would aid the user in finding pages by similar terms with content-based analysis.

As a mature student, P1.3 described several areas of the site to be appealing to 'young students', leaving them feeling 'unable to fit in'.

When accessing the dyslexia information, P1.6 noted that much of the terminology relies on knowledge of the internal services and their relevant names. Making this content more easily understandable would have helped the participant to navigate to the appropriate pages more efficiently.

Some of the terms used throughout the website are domain-specific, leaving them challenging for one participant to understand [P1.1]. Explanation of alternative terms and use of synonyms in search contexts would help make this more relatable for the user.

Outside of the search context, one participant explored the assessment information and was confused by the unclear use of the terms *essays*,

assignments, and *coursework* [P1.8]. Many University staff will have an intuitive sense of the nuances between these terms; however, this was not clear to the user.

8.5.7. Level of detail

Within the task scenario, the participants were asked to find information about a particular course. When researching the course, the required data is split between multiple sections of the page. Whilst all the information is present, some of the text was located in detail further down the page and missed by the participant [P1.1]. This resulted in too much detail in the wrong places for the user. Ensuring that the right level of detail is present would help improve access to the content.

Another area where the level of detail was a concern was in the entry requirements [P1.2]. In one of the sections, the sole content was the number 64 with no context as to its meaning. Information such as this needs to be provided with further explanation to help the user understand its meaning.

8.5.8. Navigation

The case study follows many existing website navigation patterns; however, these rely on an underlying information architecture. Having only explored the site briefly prior to the task scenario, one participant struggled to find information on courses as there were many similar options in the main menu for courses. The participant returned to the search feature to circumvent the structure [P1.1]. This is apparent in Figure 8.6, where a search has been made for the keyword 'Business', returning at least five pages of results in an order unknown to the user.

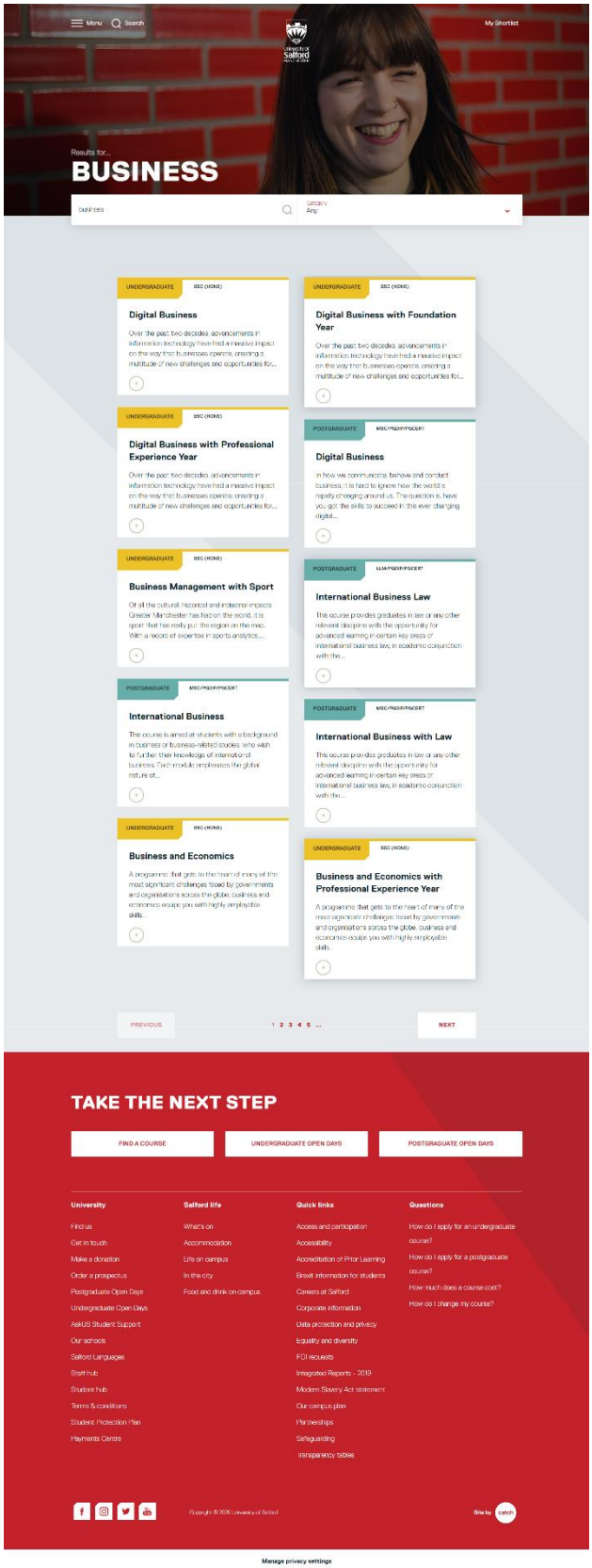


Figure 8.6: Search results for 'Business', returning at least five pages of similar results in an order unknown to the user.

Footer links

Several participants relied on the quick links in the footer below content for navigation. Particularly on the student support pages, they struggled to find the option they needed due to information overload with too many choices causing confusion [P1.2, P1.5, P1.8]. A typical selection of footer links is shown in Figure 8.7.

Figure 8.7: Typical student support site 'fat footer' containing a confusing number of links for students.

Since the original web sites of the 1990's, footers have developed to become more than a consistent design element at the bottom of a page. They are often used to house a secondary navigation, providing access to hard-to-find content and additional opportunities to find information that may not have been contained within the rest of the page (Fessenden, 2019). With the requirement for more complex systems housing larger amounts of data, there has been a shift towards 'mega footers' or 'fat footers' containing large amounts of navigational cross-links, along with the use of devices that incorporate links to multiple types of content combining business and user goals.

The case study website features an example of such a mega or 'fat' footer, with many links to different areas of the site and other portals buried at the bottom of the page (see Figure 8.7). This is used to place links to other content, reducing information loss when the content presented has not satisfied the user's requirements, and is of importance within cultures where there is an avoidance of uncertainty. (Alexander et al., 2017, p. 82).

During the case study, several participants relied on the footer links for navigation [P1.2, P1.5, P1.8], however they struggled to find the option they needed due to the presence of too many links. Reducing the number of footer links and using contextually appropriate resources would have helped them to identify where they information they required was stored more efficiently and with less confusion.

Overall, the presence of the mega footer contributed to the users' confusion about the information architecture and disrupted their model of the site when unable to find relevant content.

Accordions

Another element that caused users difficulties when accessing the web page was the accordions. In website design, accordions are one of the techniques that allow content to be collapsed and expanded manually by the user. This

can reduce the length of information-rich pages, providing a means for developers to reduce scrolling, reduce cognitive load and provide a map of the page content (Loranger, 2014). Prior usability research has demonstrated that accordions are most suitable when users only need to access select pieces of information from a page (Loranger, 2014). In this case study, there are two places where programme information is located within accordions:

- Module information – Lists of modules studied during each year of the programme are listed in the details section, along with a short synopsis of each highlighting their key features and unique elements of study.

- Application requirements – Details the standard and alternative entry requirements for students applying to the programme, including information for a variety of situations, such as home/EU/international students and those following a non-traditional application route.

Both accordions are constructed with a double-layer format, with information being hidden under two levels of headings. This is an unusual navigation format, with one participant in the study realising they had found the correct place where information was located however did not know how to discover the content as the second-level device was not immediately visible [P1]. This is further compounded by the design of the accordions: whilst the outer and inner levels have contrasting colour schemes, the font size, positioning, and sizing of the elements is identical as shown in Figure 8.8, resulting in them appearing to be sibling rather than hierarchical controls.

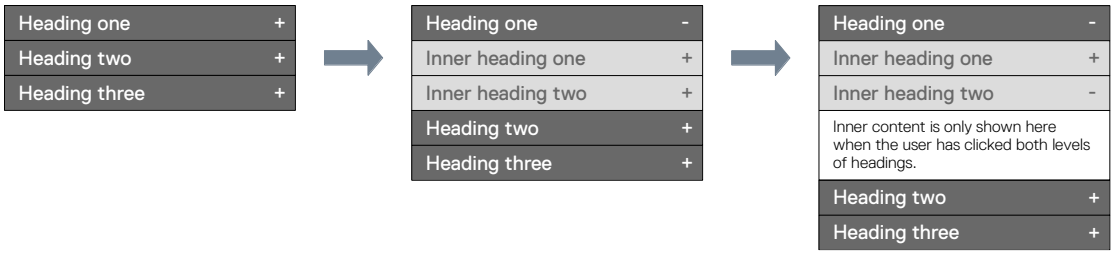


Figure 8.8: Wireframe illustration of multi-level accordion controls. The content is only visible once the user has clicked two levels of headings. Note the similarity in style between the two levels of headings.

There are many different design features that could have been included to make these accordions easier to use, including:

- **Avoidance**
  Present information in alternative formats, such as using a dropdown or tabbed interface.

- **Spacing**
  Use of indentation to signify different levels.

- **Styling**
  Use of font size to further indicate differences between primary and secondary navigation levels.

- **Wizard-based approach**
  Facilitate access to information through asking the user a series of questions, making the site more interactive and personalised for the user's context.

Stateless navigation

When looking up information about the fees, one participant followed a link that resulted in stateless navigation [P1.3], meaning that when they used the browser history their selected course was forgotten. This provided a jarring

experience that was not congruent with their cognitive understanding and caused confusion.

8.5.9. Presentation

At the top of the course information pages, there is a callout area for key course information. The content in this area is static, although one participant viewed the styling as representative of links to other areas of content [P1.1].

In addition, links are embedded within the textual areas of the website. These were challenging to distinguish from the content, resulting in participants missing the navigation structures.

When viewed on a large screen, many of the pages appeared too large to be easily readable for P1.4. This resulted in more cognitive effort when processing the information within the text.

On the student support pages, there is a mouseover effect on the main content area. This was distracting as the content moved every time a participant placed the cursor near the link they wanted to follow [P1.10].

The telephone number for getting in touch with the institution is located in an unusual location on the page, leading to two participants missing this information [P1.4, P1.9]. Ensuring the information within the pages follows online convention and is easy to find would help improve accessibility of expected content without the anticipated need to attend in-person to access support as it was 'too hard to find' online [P1.4].

8.5.10. Recency

It has been documented in many prior studies that recency has a positive impact on the users' perceptions of information quality. This was apparent when [P1.6] highlighted that there were no recent news articles (the last being 10 months prior to the study), and the imagery used old one-pound coins that were withdrawn two years prior to the launch of the website. The

age of the content made it appear that the site was not up-to-date and resulted in this participant questioning the accuracy of other information.

8.5.11. Search

One participant focused on search-based solutions for every task and found that many of the expected search boxes did not behave as expected [P1.5]. This included input fields that did not look like input fields, and hidden filters among the elements. The experiences of this participant highlight the need to follow established design patterns and make search interfaces consistent with existing patterns of usability.

When conducting the latter stages, the same participant used an external search engine to find information as they claimed this to be easier than using the in-built search features [P1.5].

8.5.12. Understandability

One area of content with a lack of sufficient detail is the entry requirements. These are displayed in technical terms with no explanation for non-expert users, impacting on the understandability for participants [P1.3] and [P1.7].

The course pages on the case study website contain mostly factual content, however one of the participants found this hard to use because the information did not match their perspective, thus impacting on the understandability of the site [P1.2]. One example of this was the disconnect between the terms 'technology' and 'digital'. Those involved in the domain use the two words almost interchangeably, however users may not understand the terms in this manner.

8.5.13. Value added

The case study website provides value-added content in the form of hints and tips during the application process. In the middle of one of the pages, there is a short guide to crafting personal statements that punctuates the

information being retrieved by the user. This was highlighted by two participants as distracting and leading them to miss some of the essential information later in the page [P1.1, P1.6].

8.5.14. Visibility

When the live chat window was displayed to the user, it covered some of the content, leading to one participant missing crucial information for the task [P1.2]. This was distracting to the user, especially as the participant did not understand its purpose and it was not easy to close during the activity. If such interactive elements are provided, they need to be optional and easily controlled by an unfamiliar user.

8.5.15. Data analysis

To obtain the data categorised and summarised in the fourteen information quality attributes above, the researcher applied the data analysis cycle as described in Section 6.7.10. To demonstrate this process, a full worked example of the analysis for the *consistent representation* attribute is included in Appendix C.

The same data analysis process is replicated for future action case study cycles, though not included in this thesis for brevity.

8.5.16. Summary

After conducting the research with the case study participants, many facets of information quality have been uncovered using the think aloud protocols, incorporating the 11 existing information quality attributes first identified in Section 7.3, and three new attributes introduced during this study for the first time (fallback, information usability, and interactivity). The suggested improvements for content are summarised in Table 8.1 below.

This case study surfaced several key themes within the data. The first such theme was that interactive elements intended to improve the user

experience may impair access to the website content. This was the case with some of the elements such as live chat, accordions, and footer that prevented participants from finding the information they required more easily. In addition, changing default behaviours such as mouse pointers may provide confusion to users.

Often the information the user may be looking for is present on the page or site, though cannot be found easily. Introducing clearer information architecture, including consistent labels and links between sections would help improve access to the content.

Attribute	New?	Summary of data obtained during Case Study 1
Availability		Ensure navigation does not prevent access to information.
Consistent representation		Websites and subsites should be visually and hierarchically consistent to ensure users do not have to relearn how to find the information they need. Use relevant images for content.
Fallback	✔	Provide fallback contact details in a consistent, clear location of every page.
Information usability	✔	Information should be modelled as the user would expect to find it.
Interactivity	✔	Do not override default interaction elements. Avoid the use of elements that hide information from the user. Ensure interactive features work as expected on all browsers.
Language		Use terminology consistently, avoiding synonymy. Explain technical or domain-specific knowledge where it is first used. Use language inclusive for all users.
Level of detail		Where further detail is provided elsewhere on the same page, link to the content. Ensure content is at the right level for the user's purpose.
Navigation		Use an intuitive information architecture without a reliance on large numbers of links.
Presentation		Ensure common conventions for website layout are followed. Make sure pages work consistently on all screen sizes.
Recency		Update dated content frequently to give the site a fresh appearance.
Search		Use existing design patterns and conventions for internal search features. Ensure external search features are able to be used as a backup.
Understandability		Ensure information is presented in user-accessible terms.
Value added		Provide interactive hints and tips in a complementary manner that does not distract experienced users.
Visibility		Do not use elements that cover part of the content.

Table 8.1: Summary of suggested areas of improvement during Case Study 1, categorised by information quality attribute.

The specific language used by the content should also be considered carefully, as terms may introduce synonymy (where two terms have the same meaning) or confusion for users where there may be two or more meanings to a word or phrase. Technical or specialised nomenclature should be explained clearly within the page to avoid ambiguity of meaning that causes users to miss crucial information, leading to unsuccessful task completion.

The categorisations above demonstrate the user behaviour being mapped to a range of existing and new terms that may form the basis of the information quality taxonomy to be produced and refined throughout the next three case studies. A similar case study scenario based on the attributes identified above will be produced for the next cycle, further demonstrating the validity of the facets that impact on the information quality delivered through websites. This will result in a list of attributes involved in the case study websites that will be documented using the format above and transformed into a taxonomy for application to the final validation case study.

8.6. Conclusion

Many themes have emerged during this first case study, leading to the addition of three new information quality attributes. Data has been analysed using the cyclical method, ensuring that participant observations and utterances from the think aloud studies are considered as part of the larger data set. The overall fourteen attributes encapsulate the relevant aspects of content quality within the university website.

In the following two chapters, further case studies are explored to demonstrate the similarities between users' experiences across multiple website domains providing a generalisation and validation of the observations produced during this action case study cycle.

Chapter 9. Case study 2 – Non-profit website

This chapter describes the second case study, with the think aloud methodology applied to the information on a non-profit website.

9.1. Introduction

For the second case study, the researcher worked with a digital agency to analyse the information quality aspects of a non-profit website developed for one of their clients. This was a suitable choice for the case study as the non-profit website contains information on the client's services for a wide range of audiences, including expert and non-expert users. The research project occurred at the same time as a website redesign project aimed at the restructure and representation of the organisation's objectives. The client was open to the application of think aloud techniques and used the responses from this study to help inform the website content design.

9.2. Rationale

The selection of a non-profit website as the second case study helps to further demonstrate the general applicability of the information quality techniques studied during this research. Whilst many non-profit websites are focused on conversions, this client's site has a large informational element to accompany many of their day-to-day activities and provide content for a wide range of service users and their families. This focus on furthering the charity's aims is demonstrate through the important of content design in the website refresh.

9.3. Task scenario

The researcher developed the task scenario shown in Figure 9.1 for the think aloud studies conducted for this cycle of research.

Help us improve this website

During this study, you will be asked a series of questions about your use of this website, and the ability to find the information you are looking for. All questions are optional, and your responses will remain anonymous. The information provided by you in this questionnaire will be used solely for research purposes and to improve the content of this website. It will not be used in a manner which would allow identification of your individual responses. This research has been approved by the University of Salford Ethics Committee (SBSR1617-18). If you have any queries about this questionnaire, please contact Jessica Muirhead at j.muirhead@edu.salford.ac.uk.

Go to the website: www.diversityrolemodels.co.uk

You can explore the website for five minutes prior to the study, after which you will be asked to complete a scenario whilst talking about your thoughts and actions.

Now that you are familiar with Diversity Role Models, I will give you a series of tasks, and observe how you use the website. During these tasks, please describe any thoughts you have – what you are looking for, where you are finding the information and anything that pleases or frustrates you. All of your responses will remain anonymous, and you will not be timed or tested.

1. What does Diversity Role Models do?

2. Who do they deliver training for?

Imagine you are a secondary school teacher, interested in booking a workshop.

3. What information would you like to know?

4. Where can you find that information?

5. How would you book a workshop?

6. How would you volunteer for Diversity Role Models, and what could you do to help?

Figure 9.1: Task scenario for non-profit website case study.

Ten participants were recruited by the researcher to take part in this cycle of the study, and work through the think aloud activity. For a usability study, ten participants fits within both the optimal ranges for problem discovery and comparative studies, allowing for the discovery of website content issues along with confirmation of results (Six & Macefield, 2016). As it is expected that each participant will have a varied experience of the site, the group are selected from volunteers, the majority of whom will have no prior knowledge of the organisation, website, or domain.

9.4. Information architecture

Information architecture is a critical component of the user's experience of website content. This is explored below in both *hierarchical* and *visual* formats.

9.4.1. Hierarchical architecture

The hierarchical architecture of a website can have an impact on users' interactions with content as it makes assumptions about the audience's model of the information. As the website has multiple audiences, the hierarchical information organisation will be explored from different viewpoints.

The case study website assumes that users will access content through a small number of high-level categories and navigate using more specific terms. The overall website hierarchy is shown in Figure 9.2.

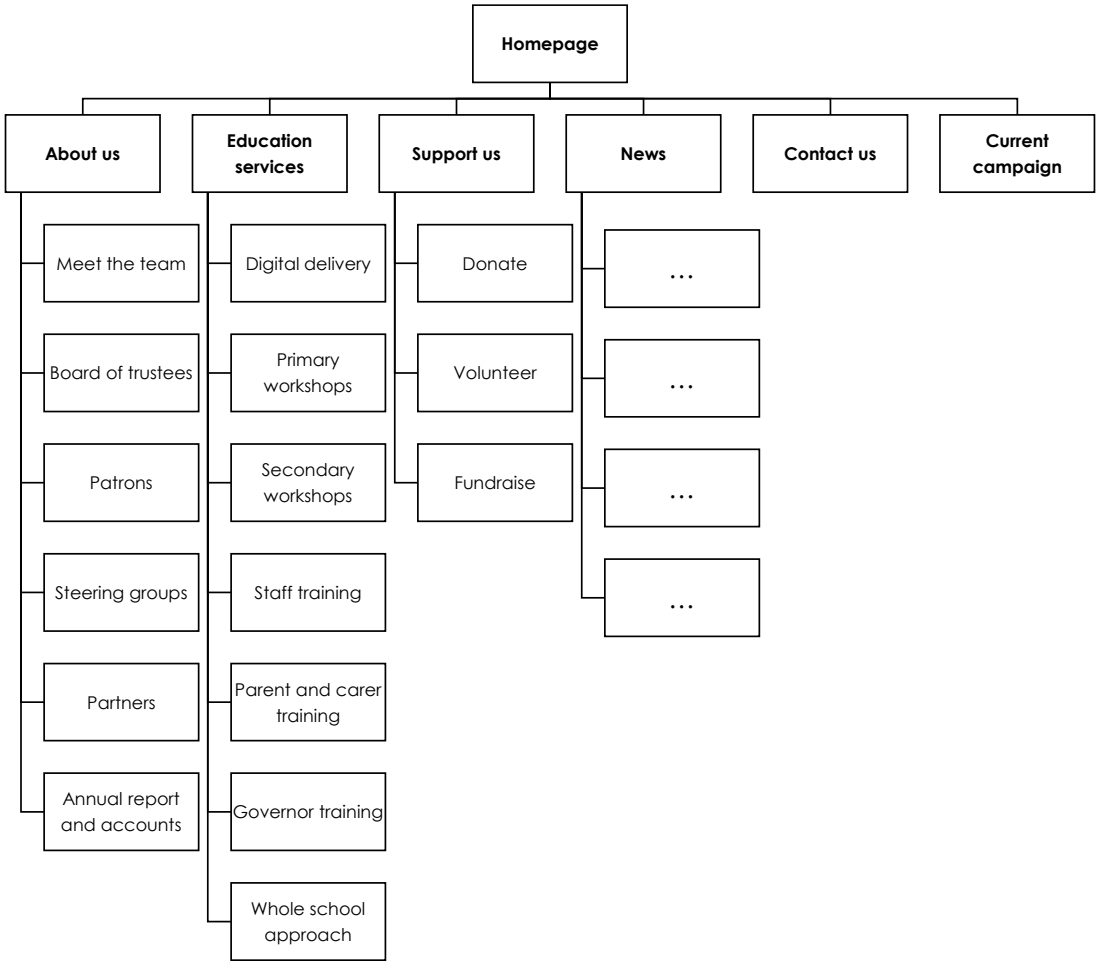


Figure 9.2: Non-profit website hierarchical information architecture.

Representing all information within a hierarchy can be useful when there are inherent structures and a common user model, however this can also introduce confusion when the structure does not match a user's world view. The task scenario has been designed to both assist and challenge the participant's view of the information, as part of the activity is unguided followed by the guided session.

Users navigate the case study website using the grid menu pages, refining their request until they access the desired information. There are six key sections represented by navigation links at the top of each page and on the homepage, which give access to the content areas. The first two of the content areas contain most of the informative content: About us with the

charity administration information required to be distributed under Charity Commission regulations, with the second content area containing information for prospective and current service users.

The case study website does not feature many cross links that could be used to help users to navigate to parallel useful information. As a smaller website this is not expected to be used heavily, however the similarity between some of the services offered by the organisation may result in users moving between content silos instead of following the designer's hierarchical pattern.

9.4.2. Visual architecture

The visual representation of a website can also have an impact on users' interactions with the content (see Section 7.3.6).

As a general pattern, the case study website uses a grid-style layout to present choices to the user. An example of the homepage is shown in Figure 9.3.

As the user navigates from the homepage into a specific category, the same grid menu structure is applied as shown in Figure 9.4. The maintains a consistent approach for the user and incorporates an overview of the sublevel pages that can be explored.

The category menu pages then lead into an informational page such as the example shown in Figure 9.5. The informational pages contain snippets of relevant text along with a small set of call-to-action elements.

Each of the page styles contains large amounts of supportive imagery selected by the design team from a brand media library to add visual appeal to the website. Interactive elements are kept to a minimum to avoid distracting users with additional unnecessary materials.

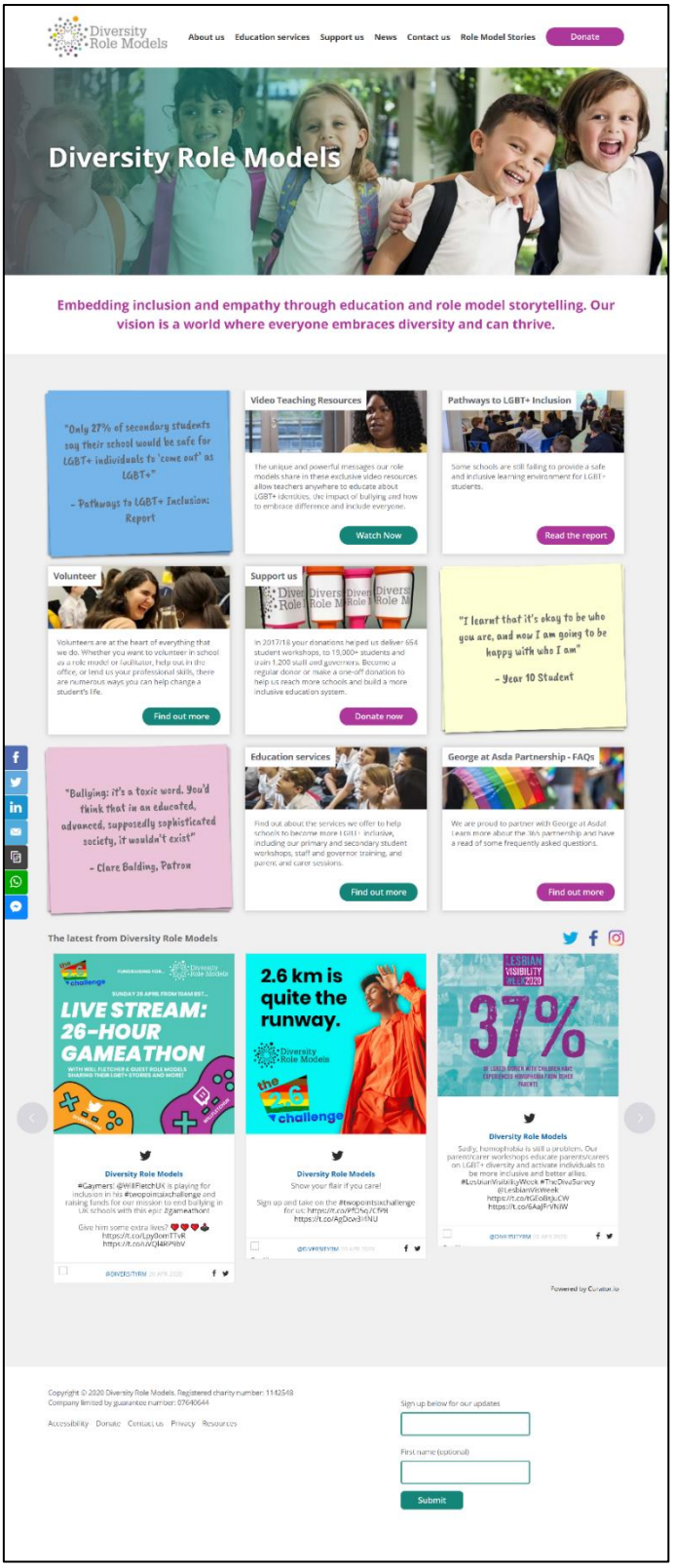


Figure 9.3: Non-profit case study website homepage with grid layout.

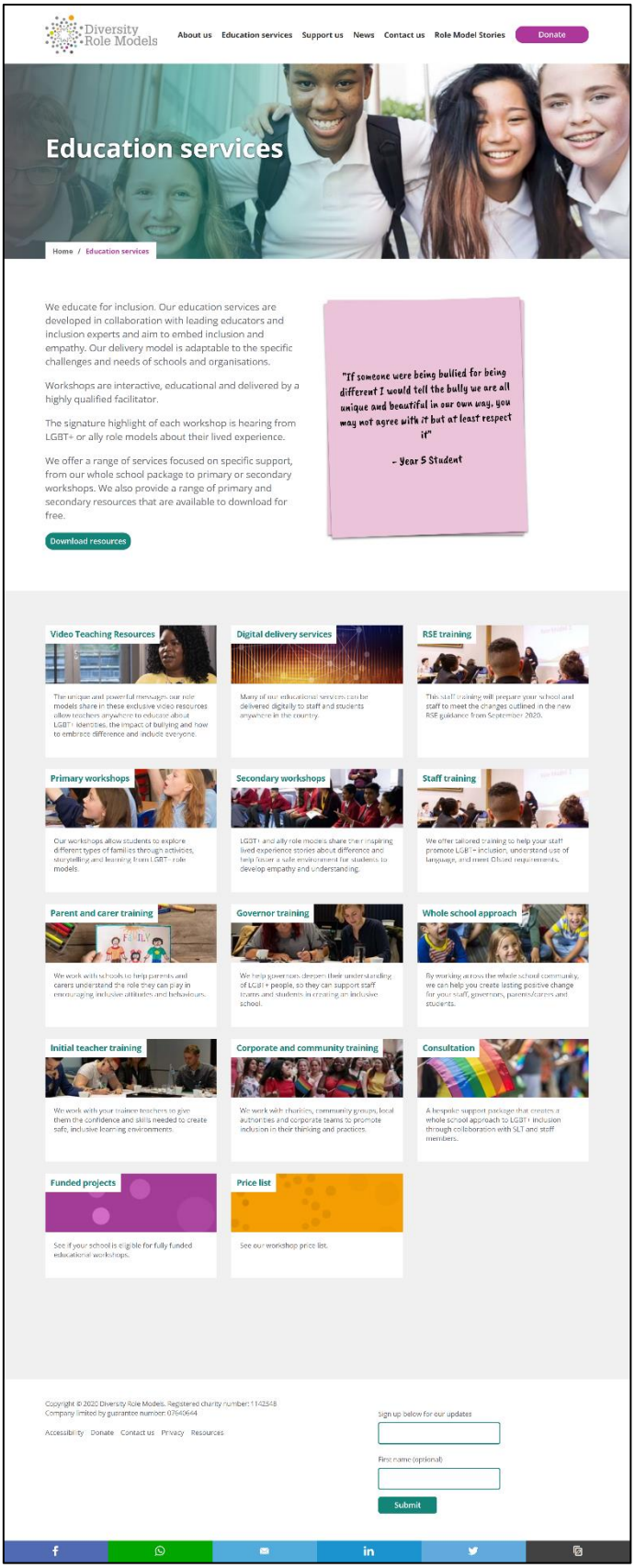


Figure 9.4: Category menu page with grid layout.

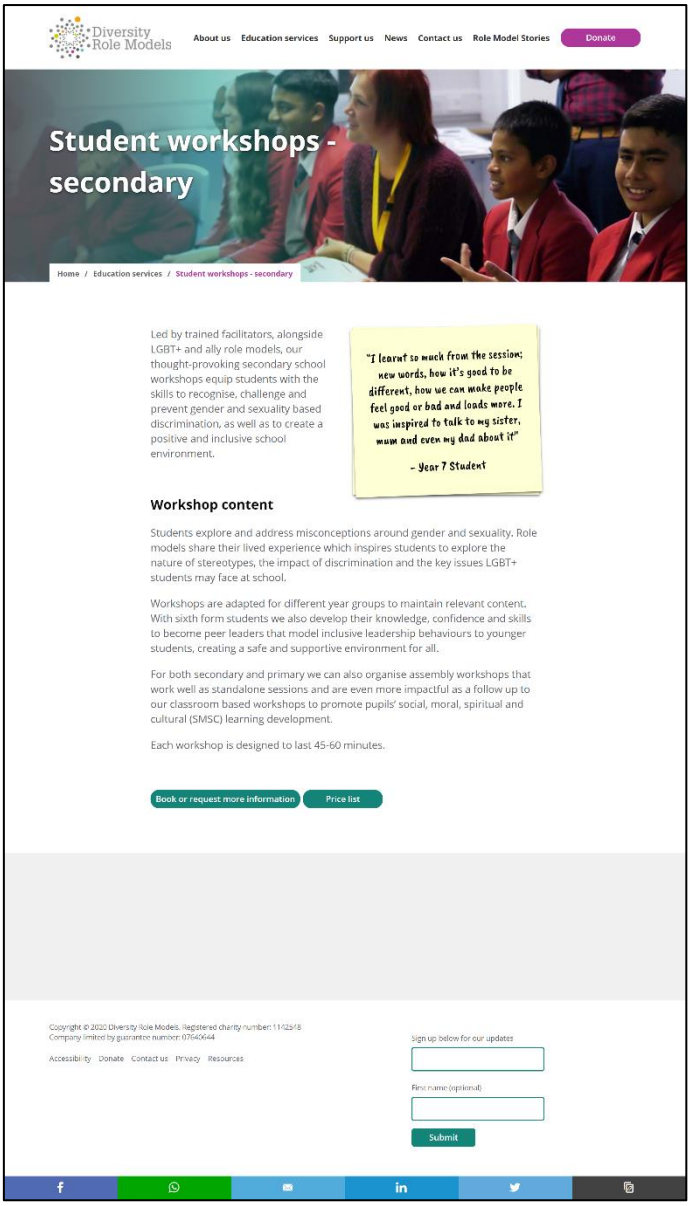


Figure 9.5: Informational page layout.

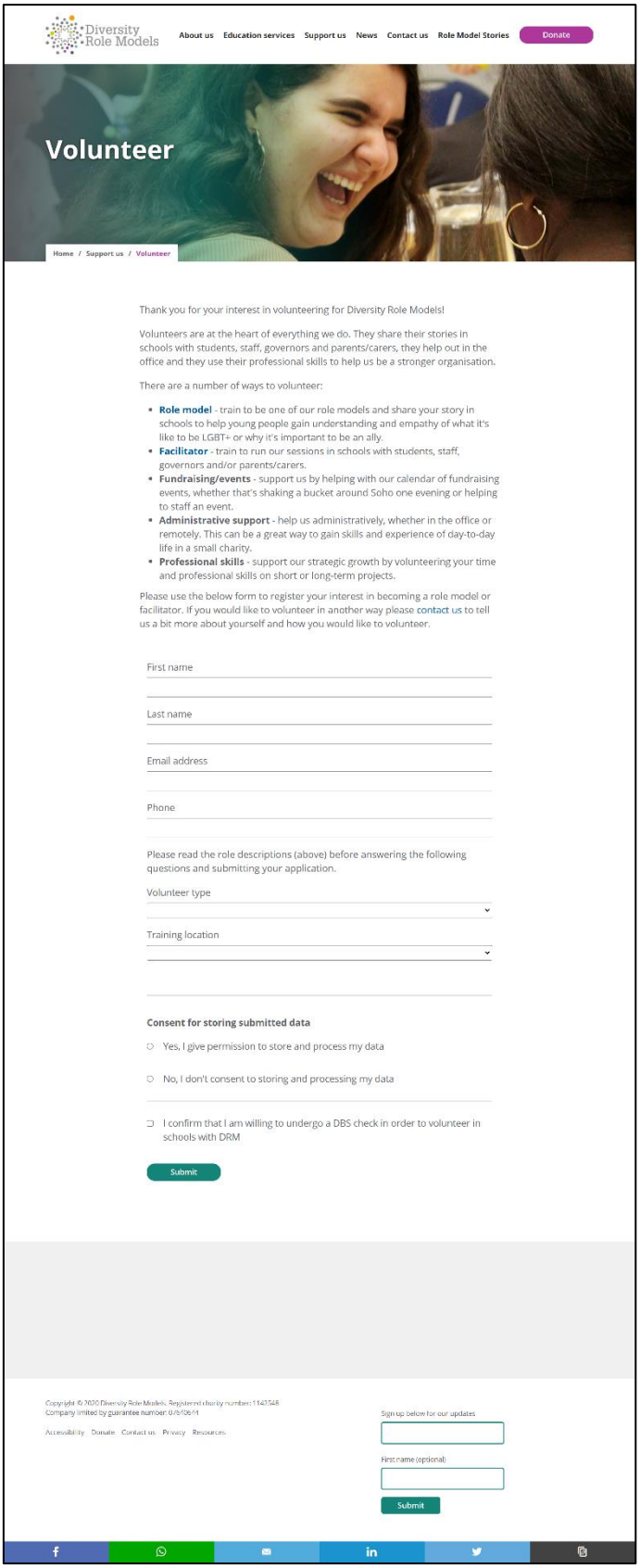


Figure 9.6: Volunteer page layout.

9.5. Responses

This section contains an analysis of the think aloud responses from the ten sessions with participants, as performed using the same process as outlined in Section 8.3. The responses are arranged by information quality attribute to identify common themes for future model development.

9.5.1. Availability

None of the participants in this cycle of the study reported availability issues with the visual architecture of the site, however a common theme amongst three participants lay with the information architecture for volunteer information [P2.4, P2.7, P2.8]. The link for information about volunteering was embedded within the 'Support us' section, with the participants not considering this to support the organisation. Clearer labelling, such as 'How I can help', could have made this connection more obvious to users unfamiliar with the organisational objectives.

9.5.2. Consistent representation

The website generally follows a consistent representation, with a use of the same grid format across most category and subcategory pages as shown in Figure 9.4. An exception to this format is the informational pages when exploring the organisation's educational services. Participants [P2.3] and [P2.8] struggled to access relevant information as the textual content of these pages is relatively short and relies upon information contained within the previous screen. Using a standardised format for the informational pages would have made information easier to access.

In addition, [P2.9] did not perceive the buttons at the bottom of informational pages (see Figure 9.7) to be links to other areas of content. Representing these as links styled in a format consistent with in-text links would have made these easier to understand without disguising the purpose of these call to actions.

Figure 9.7: Links styled as buttons disguise purpose.

9.5.3. Fallback

One participant [P2.10] highlighted that there is no fallback information within the website. They stated that they will often contact an organisation by email to ask follow-up questions or for confirmation of the information provided, and the only means to get in touch privately is through a contact form. Making all contact information easily accessible within the site would aid users in asking the types of questions that may not be answered within the informative site content.

9.5.4. Interactivity

None of the participants noted any positive or negative aspects of interactive additions to the site. The website has been implemented simply with few elements that would require advanced access or interaction techniques.

9.5.5. Language

The language used throughout the site is informal and colloquial – this generally helped the participants to access the information required and understand the content. None of the ten participants identified large issues with language, though some of the domain-specific terminology such as the difference between the various training programmes would have made information easier to access.

9.5.6. Level of detail

The site architecture provides information through a highly categorised and hierarchical approach, with smaller amounts of information within the resultant pages. Participant [P2.8] expressed a dislike of this approach, as navigating through several pages led to a single paragraph of text. Adding more details about the individual services (even if these are repeated across multiple pages) would have provided confidence to the user that the information is accurate and answered questions without the need for multiple clicks.

9.5.7. Navigation

The case study website primarily uses one navigational pattern, with the user clicking through a series of category and sub-category pages with options displayed in a grid style. This pattern made it easy for participants to learn and remember how the site is structured.

One element that confused participants was the main navigational menu: Participant [P2.4] identified issues with the text used for the labels. Captions such as 'education services' and 'support us' contained inherent information about the organisation, however these also assumed an understanding of what would fit into each of these categories. For example, *volunteering* could fit in both depending on the user's perception of the activity and could have been signposted more clearly.

Participant [P2.1] frequently used the *breadcrumbs* to navigate between pages and their parent categories, however they were frustrated that the links at the bottom of information pages (such as the example in Figure 9.5) were to alternative pages at the same or a higher level in the hierarchical architecture, breaking the trail of breadcrumbs they had been following. Ensuring that the breadcrumb options match the user's expectations is particular important for those users who build high levels of spatial awareness,

expecting such a device to provide enhanced methods of navigation such as menu-based structures (Ahmed & Blustein, 2006, pp. 194–195).

Participant [P2.2] welcomed the inclusion of relevant testimonials throughout the site as a means for evidencing the information relayed through the website. Whilst testimonials do not influence the user's knowledge development, they can help to build trust relationships to verify the information provided through the site (Taddeo, 2010).

Forming a trust relationship with the website was more challenging for [P2.3], who could not navigate the hierarchical information architecture with ease. This led to mismatches between their view of the information in the website and the information intended to be conveyed by the organisation. Ensuring that links and structures anticipate user needs would help to account for their perspectives of the services described within the informative pages.

9.5.8. Presentation

The overall presentation of information with the case study website is consistent with a single look and feel and menu structure (see Section 9.4.2).

Participant [P2.9] noted that the banner header image on most pages was of a generic style and could be as large as 50% of the screen traditionally considered to be 'above the fold'. Using banner images with more relevance, such as specialised to the pages and with more contextual clues could have enhanced the navigational experience by incorporating implicit information about the text within a page.

One participant noted that the banner images on each page are not fully supportive of the page content, especially where there are several similar pages and images in the educational services category. Identifying and applying more supportive images would give users clues to the appropriateness of the content for their needs.

9.5.9. Recency

Whilst the task scenario did not include activities incorporating the News section, participants [P2.3] and [P2.10] did navigate to this area during the initial exploratory activity.

Participant [P2.10] remarked that the news section was not up to date as the most recent article was almost 6 months old; at the same time little context was provided for the articles to highlight their importance or relevance. Including more recent articles as well as short excerpts could have helped develop a sense of recency by including relevant links to current affairs and information expected by the users.

9.5.10. Search

The case study website does not incorporate a search feature, although three participants did attempt to find such a function during the task scenario. One participant clicked the 'Donate' button considering it to be a search field due to its placement in the top right of the page, disambiguation by colour and rounded borders in a similar format to many websites' search mechanisms (see Figure 9.8).

Figure 9.8: Case study 2 website Donate button (left) compared with default Webkit search input (right).

Another participant confused the newsletter signup form in the footer for a search box, due to its prominent position in the bottom right of the page +footer and non-prominent labelling of the form input or submit button (see Figure 9.9). Adding a clearer heading and legend for this input would prevent

ambiguity with other actions and fix the lack of directional input for the first box.

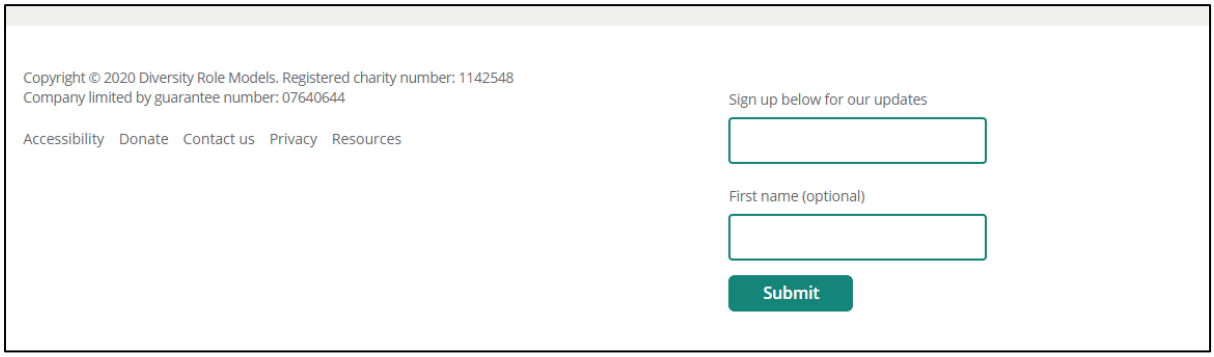


Figure 9.9: Ambiguous newsletter signup in footer interpreted by participant as a search field.

Providing a search functionality would enhance the site by providing users with the opportunity to find information hidden within the hierarchical architecture across user perceptions of content categorisations. This is particularly useful for highly-specified queries such as those demonstrated during the task scenarios (Capra et al., 2007).

9.5.11. Understandability

Due to the highly informative nature of the site, the task scenario provided closed activities with questions either fully answered or not at all. Participant [P2.7] was unsure of the purpose of the organisation due to generic copy being used on the homepage before being built upon in more specific category and informational areas. Providing a more inclusive and encompassing opening headline on the homepage would have made the purpose clearer and aided participants, although this may have a reduced impact on a live site where users have deliberately chosen to visit the site.

9.5.12. Visibility

Using long pages with large banner header images pushed content and options further down the page. Whilst all participants worked out how to find information within the page structure, reducing the size of the banner images

would have resulted in less cognitive load and time working out the page structure and how it can be navigated.

9.5.13. Summary

The second case study has surfaced some of the same information quality attributes as the first case study, highlighting the impact that small changes to website content, architecture and presentation can have on users' abilities to complete a task. Whilst the study worked with ten participants, the common themes demonstrate that some attributes will have a larger effect on information seeking behaviour than others.

During the think aloud studies, data was collected on 11 out of the 14 information quality attributes considered as a result of Case Study 1 (see Chapter 78), including the new attribute of *fallback*. One major consideration that was produced because of multiple pieces of participant data across multiple attributes is the need to contextualise website content for the user, focused on their perspective rather than that of the organisation. This can be categorised across several of the information quality attributes that are the focus of this study.

Another theme emerging from this case study is the need to style elements appropriately to provide distinction of purpose (i.e., search and other text fields) or to provide affordance in navigation (i.e., appropriately selected and sized banner images). This has a larger impact on the information quality experience of the user than initially anticipated, producing further evidence of the close link between informational properties of a website and other elements of the user experience such as navigational and presentational aspects.

The suggested areas of improvement emerging from this case study are summarised in Table 9.1 below, categorised by information quality attribute.

Attribute	New?	Summary of data obtained during Case Study 2
Availability		Clearer labelling for activities is required from the user rather than organisational perspective.
Consistent representation		More use of standardised page layouts without reliance on previous screens
Fallback	✔	Provide contact information in visible areas of every page.
Information usability	✔	—
Interactivity	✔	—
Language		Informal language improved access to information.
Level of detail		Provide contextualised information within relevant pages rather than relying on hierarchical structure.
Navigation		Use navigation text with reference to user's activity rather than properties of the organisation. Maintain breadcrumbs across website to match user journey. Anticipate user needs within navigational structures.
Presentation		Ensure imagery is fully supportive of page content.
Recency		Provide up to date content and ensure this is indicated to the user.
Search		Use styling to distinguish between search and other input fields, e.g., newsletter signups.
Understandability		Incorporate key properties of organisation into early website content, e.g., headings.
Value added		—
Visibility		Consider size of header content to ensure usable information is provided early on the page.

Table 9.1: Summary of suggested areas of improvement during Case Study 2, categorised by information quality attribute.

The next case study further develops these attributes, leading to the model to be developed for practitioners to apply information quality attributes to their website to improve user experiences.

9.6. Conclusion

Many common themes emerged between the first and second case study websites, with some information quality attributes such as consistent representation, interpretability and understandability having a large impact on participant access to informative content. This highlights the importance of these attributes to the overall aim of improving informative website content following a set of guidelines within the future model.

Case study 3, an e-health website, builds on the first two case studies, further developing the researcher's understanding of user perceptions of the information quality attributes identified from the literature review and task scenarios.

The next chapter describes the third action case study, applying the think aloud methodology to an e-health website.

Chapter 10. Case study 3 – e-health website

This chapter describes the third action case study, a think aloud analysis of the information quality within an e-health website.

10.1. Introduction

For the third case study, an analysis of the information quality within a national e-health website will be performed.

10.2. Rationale

The e-health website used for the third case study has been selected by the researcher as it caters for a wide range of audiences and explores another information-rich space. The purpose of the website is to inform and calls to action are focused on finding more information or accessing treatment through non-commercial means.

Over the past decade, there has been a growing body of information quality research specifically targeting medical domains. This has developed into frameworks such as the Health Information Quality (Al-Jefri et al., 2018) that develop the ideas of researchers such as Wang & Strong (1996) for a specialised domain. The purpose of this study is to produce a model for information quality across any informative websites and developing such a case study has further demonstrated the applicability of the methods in this area.

In their study of perceptions of website informativeness, Thielsch et al. (2019, p. 451) conclude that user demands on e-health websites are greater than other domains, with higher expectations for quality content. Due to the importance users place on this content, including an action case study of this

nature explores the e-health space and may determine additional factors contributing to website success within the sector.

10.3. Task scenario

The researcher has developed the generic task scenario shown in Figure 10.1 for this case study to explore the case study website. This uses the think aloud method to observe and explore users conducting a series of activities on the website. The task scenario is designed to lead the participant through a simulated user journey that requires the use of many different information-rich web pages on the case study website.

Help us improve this website

During this study, you will be asked a series of questions about your use of this website, and the ability to find the information you are looking for. All questions are optional, and your responses will remain anonymous. The information provided by you in this questionnaire will be used solely for research purposes and to improve the content of this website. It will not be used in a manner which would allow identification of your individual responses. This research has been approved by the University of Salford Ethics Committee (SBSR1617-18). If you have any queries about this questionnaire, please contact Jessica Muirhead at j.muirhead@edu.salford.ac.uk.

Go to the website: www.nhs.uk

You can explore the website for five minutes prior to the study, after which you will be asked to complete a scenario whilst talking about your thoughts and actions.

Now that you are familiar with the NHS website, I will give you a series of tasks, and observe how you use the website. During these tasks, please describe any thoughts you have – what you are looking for, where you are finding the information and anything that pleases or frustrates you. All of your responses will remain anonymous, and you will not be timed or tested.

1. **What is the purpose of the NHS website?**

2. **What information does the website contain?**

Imagine you are experiencing dental pain and are visiting the website for help and advice.

3. **What information would you like to know?**

4. **Where can you find that information?**

5. **How would you obtain further assistance?**

6. **How would you find a dentist near you?**

Figure 10.1: Task scenario for health website case study.

This task scenario has been designed to feature a common ailment, leading to participants already having a fundamental understanding of the domain without exploring the website. The likelihood of a participant having already been through this journey in either the online or offline world isolates the learning element of the task scenario and focuses more closely on the

usability and user experience aspects. This will align the study with the information contained within the website rather than the overall service journey.

10.4. Information architecture

The information architecture applied by the case study website can contain assumptions and influences on usability patterns and information seeking behaviour. The hierarchical and visual architectures are discussed below.

10.4.1. Hierarchical architecture

The hierarchical structure of the e-health website mostly follows a top-down pattern, with users led from the generalised homepage to more specialised and smaller scoped subsections containing relevant content. The diagram in Figure 10.2 shows the general layout pattern, with users navigating from a landing page to find the desired information.

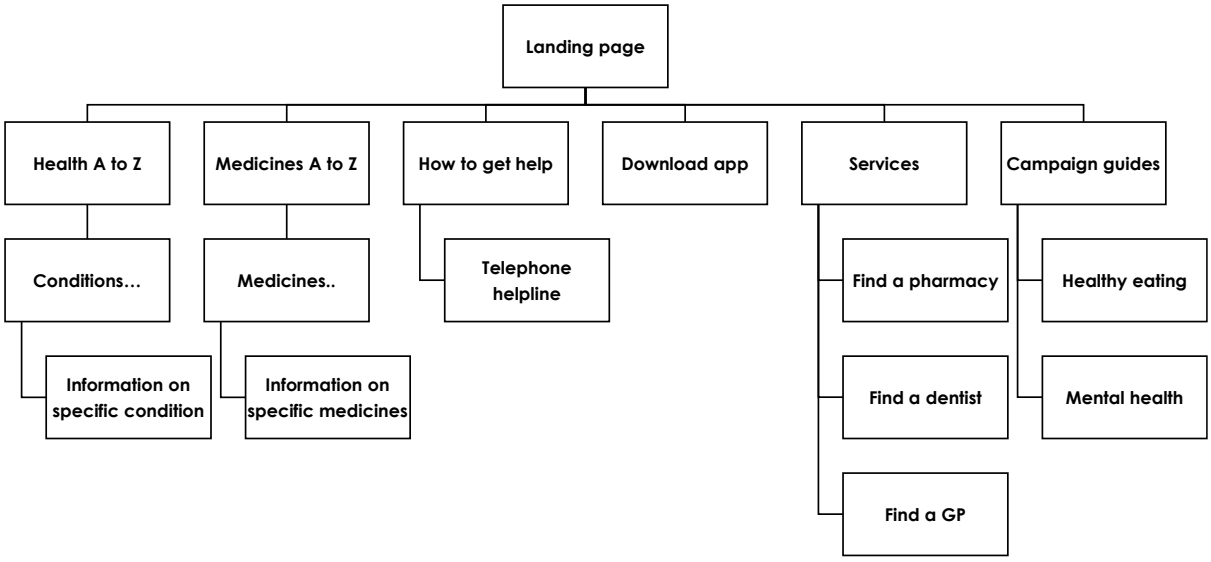


Figure 10.2: Hierarchical architecture of e-health website case study.

This hierarchical structure forces users to follow a defined path through the website, and the task scenario has been designed to challenge the participants' view of the system by asking questions which require navigation both following this standard pattern and across different silos of information.

Once a user has navigated to the informational pages, there are some cross links to relevant content, although the majority of these take the user to another landing page to conduct another information seeking activity.

10.4.2. Visual architecture

The case study website applies the GOV.UK Design System as developed by the Government Digital Service (GDS) to provide standardised and accessible interfaces for online services (Paul & Vipond, 2019). Although the site has been designed to use existing components and principles, these mainly relate to the visual and functional aspects rather than the quality of the information contained within the site.

10.4.2.1. Landing page

When first visiting the website, users are taken to the main landing page as shown in Figure 10.3. This page features many call to actions with a simple linear layout to access all of the major page categories.

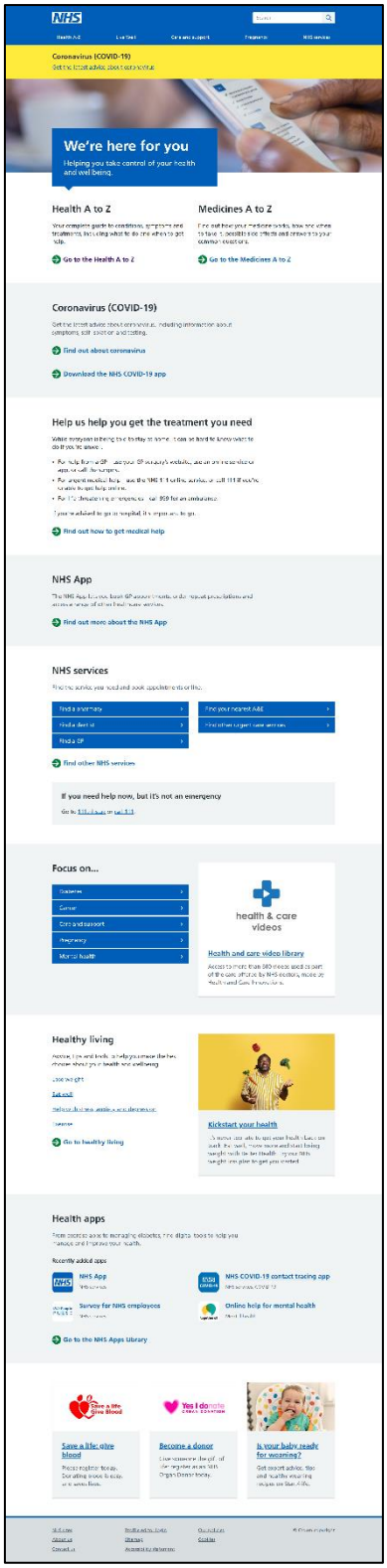


Figure 10.3: E-health website landing page.

10.4.2.2. Health conditions page

If a user selects the 'Health A–Z' option, they are presented with an alphabetical list of possible conditions with skip-to-initial links as shown in Figure 10.4.

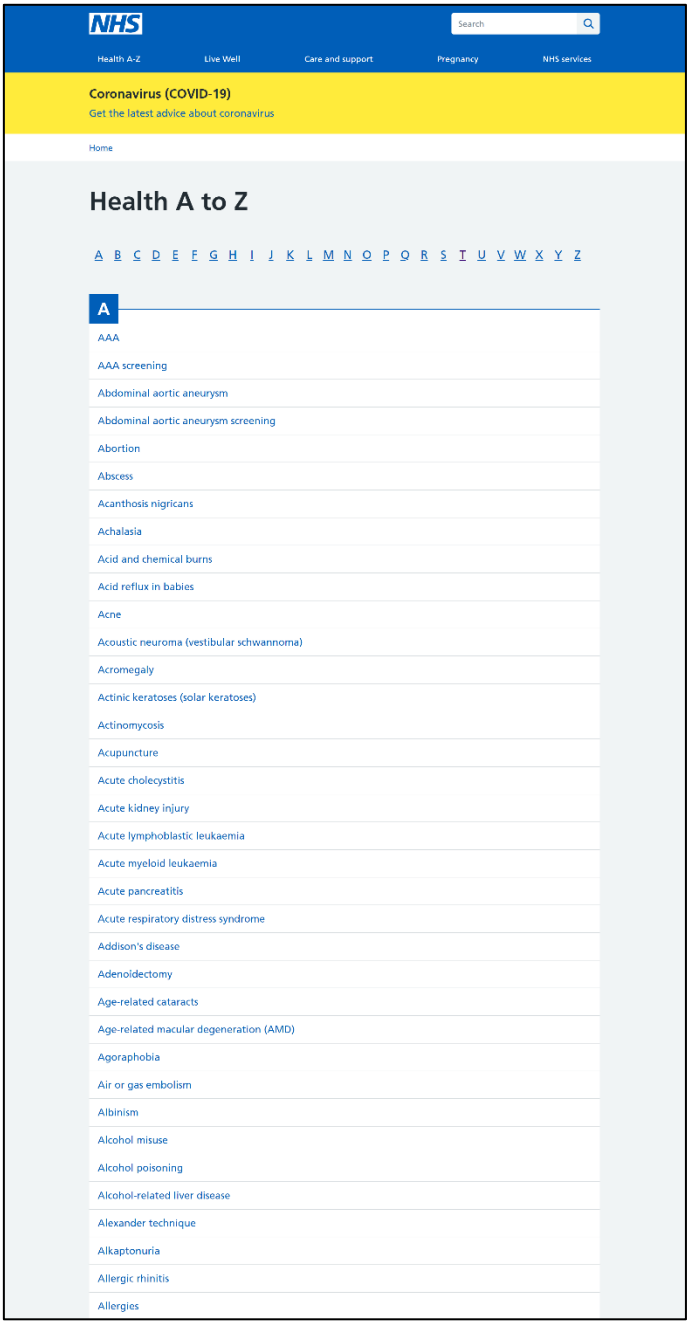


Figure 10.4: E-health website page listing conditions.

10.4.2.3. Condition information page

Once a user selects a condition, then they are shown information about that particular ailment. The information within a page is categorised according to importance with links to other areas of the site that may be relevant at this stage.

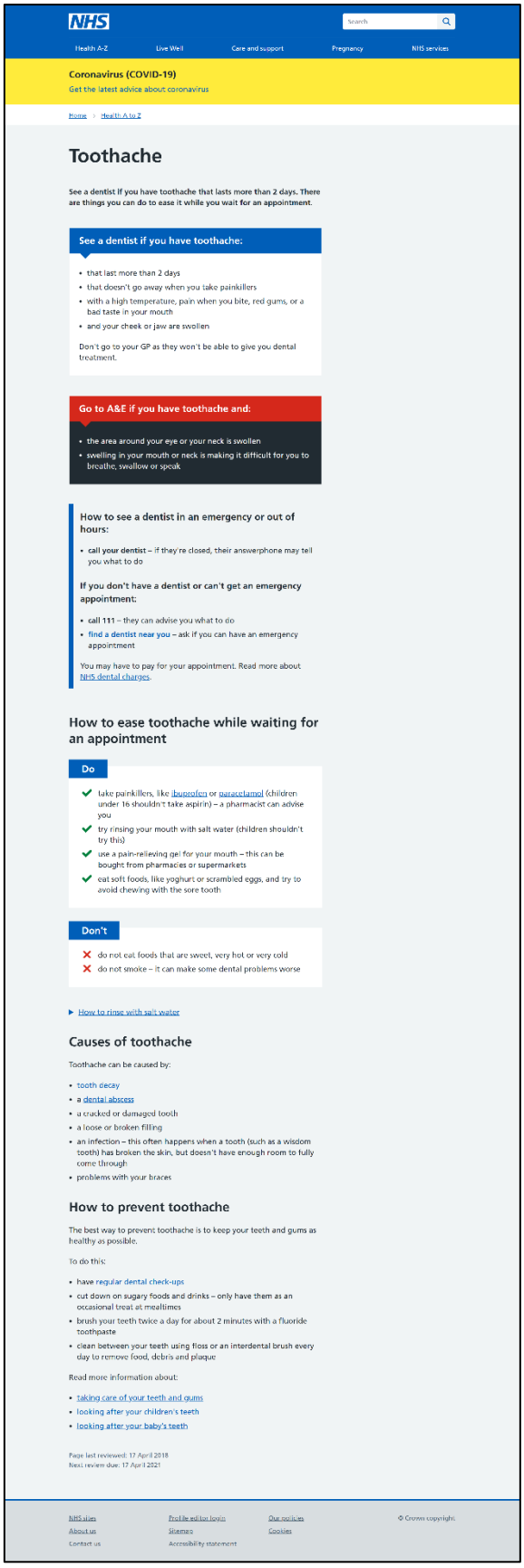


Figure 10.5: E-health website sample condition information page.

Other pages within the site follow the same general layout pattern for consistency, with prioritised information near the top of the page with a small number of cross links across the hierarchical architecture.

10.5. Responses

The task scenario outlined in Section 10.3 was explored using the think aloud method with ten participants. Their responses are recorded below using the common information quality attributes determined during the literature review stage of this study.

10.5.1. Availability

The 'Health A-Z' page caused availability issues for many participants. For example [P3.3] and [P3.5] spent a considerable amount of time scrolling to the desired conditions, which were located far down the page. The interface did have some skip buttons at the top of the screen, though these required the participant to understand the exact term they were seeking.

Participant [P3.7] had a different approach and used the browser's in-built search functionality to look for keywords. Changing the interface to make this easier by incorporating the in-site search functionality into the top of the page to increase its prominence.

10.5.2. Consistent representation

Participant [P3.2] mentioned that the consistent design made it easier to find information as they 'knew where to find links and information'. This was further demonstrated by [P3.5] who appreciated the different colours of the informational pages in increasing intensity for important information.

10.5.3. Fallback

There are several fallback opportunities built into the content within the case study website. Participant [P3.4] noted that the references to the helpline for further information were clear and drawn attention to at the relevant stage of the informational pages. Including this more prominently at the top of the screen or in a *contact* section would have provided alternative action points for the participants and channelled additional queries towards other means of communication.

Participant [P3.7] recognised the availability of information on dental surgeries in different formats (i.e., phone, email, and website) though considered internally which means would be best for their query. This led to them deciding an email would be best in case the surgery was closed at the time of contact.

10.5.4. Interactivity

The case study website uses interactive features sparingly, however one area that includes advanced interactivity is the search field in the top right.

When searching for a particular term, the search field makes suggestions on the text typed in by the user. Participant [P3.2] noticed that when typing search terms, sometimes irrelevant information was displayed (see Figure 10.6). This led to confusion as irrelevant options needed to be discounted and ignored in the suggestions.

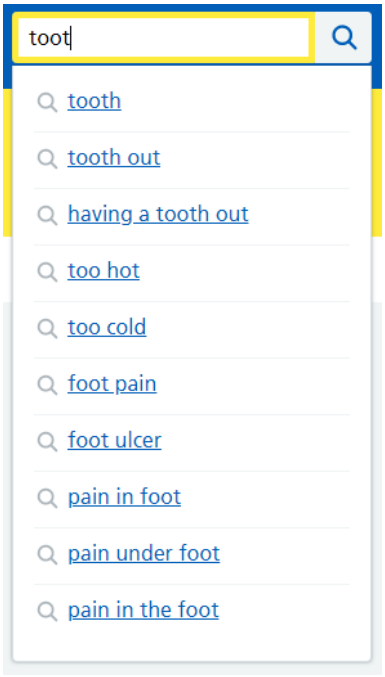


Figure 10.6: Search suggestion on e-health case study website.

Another participant attempted to search for 'teeth' and did not find any relevant options (see Figure 10.7). This confusion was caused by the search field expecting exact terms and could have been resolved by allowing reasonable synonyms such as 'tooth' and 'teeth' to yield the expected results.

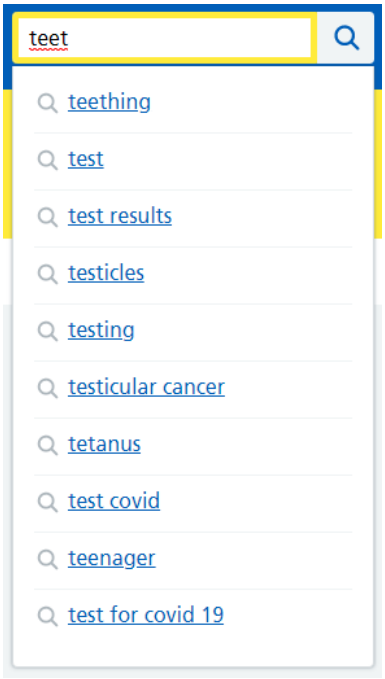


Figure 10.7: Search results with no relevant options.

The website uses no imagery and is fully text-based, in contrast to most modern websites. The lack of iconography caused Participant [P3.8] to question whether they were viewing the correct content, though this was resolved by checking the relevant textual labels.

10.5.5. Language

The language used within the e-health case study website follows industry best practices, including use of simple language, explaining complex terms and reducing paragraphs of text to lists. This is aided by the modern design of the pages with a linear flow of information.

Some elements of the simple language were appreciated by participants – for example, [P3.10] remarked that the lists of *do*'s and *don'ts* made the condition easier to understand and gave a set of clear actions. This was mirrored by the causes and prevention paragraphs which are clearly explained.

Tasks 3, 4 and 5 of the task scenario (see Figure 10.1: Task scenario for health website case study.Figure 10.1) required the user to find pre-determined

factual information. Participants [P3.3] and [P3.7] were unsure of what search terms to use and followed the prescribed activity precisely. This led to confusion between the terms *dental pain* and *toothache* even though they both lead to the toothache page. Participant [P3.9] had a different information searching error, looking for the term *teethache* – although grammatically incorrect, a semantically correct search term. Providing a simpler interface to reach individual condition pages would have made the task easier for users.

10.5.6. Level of detail

Participant [P3.1] expressed surprise at the simplicity of the information within the case study website. On the informational page, they expected more content above a 'common-sense' level of detail. This theme was continued by two other participants, who desired more information about dental surgeries, such as a list of services and waiting times.

10.5.7. Navigation

The navigation elements of the design are clearly and consistently presented to the user; however, three participants questioned the use of a long list of conditions as a navigational pattern. On the *Health A-Z* page, users are invited to select from a long list of conditions (see Figure 10.4). Other formats to present this information such as a stepped wizard, diagrams or more prominent search field would have made this easier to use.

10.5.8. Presentation

Two of the participants [P3.2] and [P3.6] noted that the web pages tended to have a long format with information presented linearly without any secondary means of navigation. This is apparent in all the informational and category pages, where content is arranged sequentially. Introducing a sidebar menu or skip links to jump to individual content areas would have reduced the amount of content needed to be read by the user before accessing the required information.

Participant [P3.7] recognised a link for a self-care process (*How to rinse with salt water*) which contains a separate stepped action, however there was no introductory or explanatory copy. Presenting a short introductory text of the purpose and aid that this can give would help the user recognise the benefits of the activity and purpose for the information at this point in the page.

10.5.9. Recency

The e-health website follows best practice with the inclusion of *last updated* dates for a range of content. Participant [P3.1] described this as a useful asset, as it helped determine if the information is recent and appropriate.

None of the participants questioned the factual accuracy of the information provided, though the case study website was recognised as authoritative given its strong brand identity. The recency of information provided was questioned by Participant [P3.7], who was concerned at the age of dental surgery information for more support.

Participant [P3.7] progressed in stage 6 of the task scenario to finding a local dentist near them. They noted that the information was over a year old and reviews at least two years old, mentioning that this was concerning for the relevance of the details as much could have changed in this time.

10.5.10. Understandability

After accessing the correct informational page, all of the participants found and understood the desired information to complete the task scenario. Participant [P3.10] remarked that their success in the activity was largely due to the simplicity and ease of understanding both the site content and navigational elements.

10.5.11. Value added

One of the participants noted that there was more useful information on the site than expected due to the inclusion of cross-linking within the informational pages [P3.3]. The participant followed several links and then used their browser's navigational features to return to the starting point once they had explored their understanding of the condition.

10.5.12. Summary

During the think aloud studies, participants generally found the consistency of the page styles and *voice* of the information to aid their information seeking behaviour. This is due to the website's use of a specialised version of the GDS framework to maintain a standards-based design.

Despite the use of such a framework, there were several areas where the information delivered by the site could be improved. For example, by introducing more action links between pages in the hierarchy and search or wizard-based interfaces the system could have aided participants in finding the desired information.

Providing information using simple terms helps improve user task completion, though care should be taken that alternative nomenclature still surfaces the same content for the user. Including relevancy identifiers can help users to make their own judgements on the age and trustworthiness of information provided, particularly with reviews or other user-generated content.

The table below summarises the suggested improvements made during this case study as a result of participant responses. The analysis covers 11 of the 14 information quality attributes considered during prior case studies, including two of the three new attributes.

Attribute	New?	Summary of data obtained during Case Study 1
Availability		Provide clearly labelled cues to help users access information within large lists or other website devices.
Consistent representation		Maintain a consistent style including colours to represent important information.
Fallback	✔	Ensure alternative contact information is prominent on every page of the site. Where contact details are provided from a database, ensure users have all options so they can choose the most appropriate for their context.
Information usability	✔	—
Interactivity	✔	Contextualise search fields to remove irrelevant terms and expand beyond the exact text entered by a user.
Language		Ensure synonymous terms can be used to find relevant content, considering the nomenclature appropriate to users.
Level of detail		Provide sufficient detail for users to gain an understanding of the topic, with links to other relevant sources.
Navigation		Where user choice may involve long lists, this may be more appropriately represented through a wizard, diagrams, and prominent search features.
Presentation		Maintain secondary navigation features on long pages. Explain the purpose of any additional actions embedded within the page.
Recency		Identify out of date information to help inform user decision making.
Search		—
Understandability		Maintain simple language and page layouts to improve user task completion.
Value added		Provide cross-links where appropriate, ensuring that they work consistently with the user's browser history.
Visibility		—

Table 10.1: Summary of suggested areas of improvement during Case Study 3, categorised by information quality attribute.

10.6. Conclusion

The third case study has explored the e-health website with ten participants and the think aloud methodology. This has generated several qualitative comments categorised according to the information quality attributes above to suggest how a website's informative aspects can be improved to support information seeking behaviour.

The next chapter discusses the practitioner model designed to address the information quality attributes with the largest impact on finding and understanding online content. This describes the approach and model that is validated during the final case study in Chapter 12.

Chapter 11. Model

This chapter discusses the development of the information quality model for informative online content, describing the structure and semantics of the metrics for practitioner application to website content. This process results in the definition of the Informative Web Content Guidelines (IWCG), the practitioner model detailed in Appendix A.

11.1. Introduction

The WCAG guidelines, first drafted in the late 1990s, have become a standardised method to evaluate the accessibility of online content. As discussed in Section 2.4, these guidelines are now used throughout web design and development to the extent that compliance to AA level is now mandated in the Public Sector Accessibility Regulations, applied to all UK public sector organisations.

The WCAG guidelines focus on online content of various forms (text, images, audio, video and data-driven), however no reference is made to the informative nature of the content. Conformance can be validated using a mechanical approach of checking the functionality of specific components and whether they match with the expected behaviours.

Whilst the guidelines were first developed to assess websites compliance with the needs of users with disabilities, accessibility helps all users. For example users may be impacted by their location, health or equipment used to access a service, and designers should ensure that nobody is excluded from a website (Government Digital Service, 2020). This is particularly true for public bodies which have a mandate to be accessible both on and off-line by everyone.

11.2. Format of model

This study proposes a new set of guidelines for *informative* online content, taking the same approach and applying to the *meaning* of the information represented within a page. Following this format builds on a standardised practitioner approach, with users already being familiar with the format of guidelines and outcomes whilst introducing new concepts for website content design.

11.2.1. What is a model?

The definition of a *model* can vary between academic disciplines and practitioner contexts. Within information behaviour research, Wilson (1999) started that a model is 'a framework for thinking about a problem', identifying that within the field most models do not evolve beyond statements or diagrams with little opportunity for application. RA2 of this research study aims to produce a practitioner model for application by content authors in developing more informative content. As a model that will be applied within industry, there are additional considerations for the output beyond an academic declaration of findings.

To aid with practitioner application, this study needs to also consider the means through which outcomes can be accessed. Fortunately, there is already one dominant framework within the online content space that is universally applied by content authors: the Web Content Accessibility Guidelines (WCAG). Thus, this research builds on existing approaches, formulating a model consistent with existing practice to ease accessibility to the model and encourage its use across informative web sites.

11.2.2. Guidelines and success criteria

The existing WCAG guidelines contains several guidelines, each with a few outcomes that must be individually achieved. In addition, these are grouped into one of four broader categories of accessibility principles. Structuring the guidelines in this manner allows for designers and developers to analyse

pages using a smaller subset of rules at a time, decomposing the task of compliance analysis to be easier to perform.

The WCAG guidelines were originally created by a group of enthusiastic practitioners to introduce accessibility concepts to a wider audience. The original standard aimed to be understood by practitioners with little knowledge of the nuances of web development for those users with disabilities, providing small suggestions for stepwise improvements. Following a guideline-based approach, practitioners will be able to incrementally update the information quality of a website to produce demonstrable improvements in accessibility to the content contained within.

11.2.3. Conformance

The original WCAG 1.0 standard used a checklist-based approach, where each success criterion could be checked for compliance using a binary choice: either non-compliant or compliant (W3C, 1999).

This has been superseded in WCAG 2.0 by a *compliance level* which can be met at one of four classifications (W3C, 2016b) as shown in Figure 11.1.

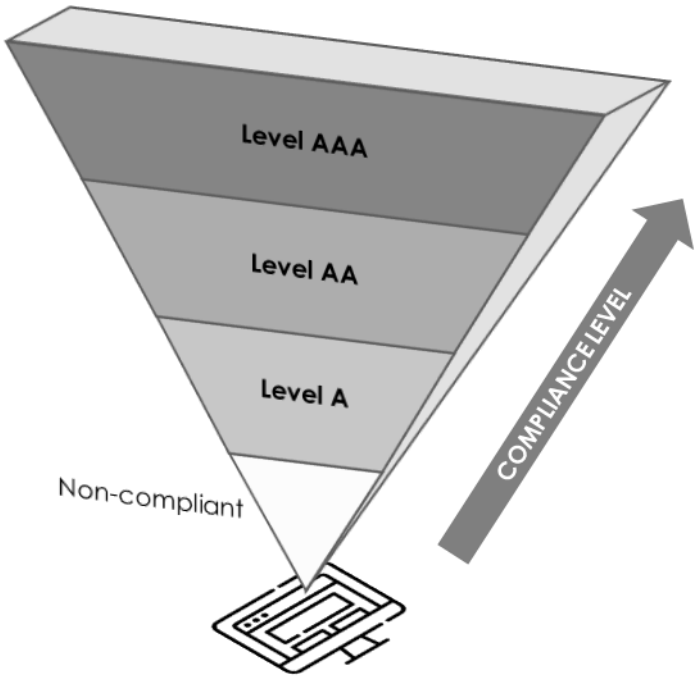


Figure 11.1: Four levels of WCAG 2 web page compliance.

1. **Non-compliant**
   The web page does not meet the success criterion due to major issue.

2. **Level A**
   The web page meets minimum compliance levels.

3. **Level AA**
   The web page meets all Level A and AA criteria.

4. **Level AAA**
   The web page meets all Level A, AA and AAA criteria, or an alternate version is provided which meets Level AAA criteria.

Once a web page meets all success criteria at a particular level, it can be claimed to be conformant at that level. The WCAG also notes that Level AAA conformance should not 'be required as a general policy for entire sites because it is not possible to satisfy … for some content' (W3C, 2016b).

The guidelines are based on the analysis of individual pages of a web site: only pages can be stated to conform to a particular level, rather than entire

systems. The only exception to this is a set of pages that form a discrete process or sequence of steps that must be completed. In this case, all pages must conform to at least the same level (W3C, 2016b). This approach does not match with user expectations, as users do not view individual pages as discrete snippets of information and rather see the web site as a whole. With a growing reliance on linked content on larger web systems, a higher-level overview is required to maintain consistency in information seeking patterns.

The upcoming WCAG 3.0 standard makes further changes to the approach: success criteria are generalised into guidelines, outcomes, and methods, shifting compliance closer to expected results. As the standard is targeted at practitioners, this model moves towards a practice-based document with suggestions for any level of web professional.

11.2.4. Style of guidelines

W3C have developed a consistent approach to evaluating web content for accessibility purposes which is already used in many organisations for evaluating pages. The UK's Public Sector Accessibility Regulations (UK Parliament, 2018) has ensured that all websites created and maintained by publicly-funded bodies comply to Level AA of the WCAG 2.1 guidelines. Extending this existing system facilitates consistency and emphasises the importance of information design alongside other aspects such as visual, functional, and technical aspects.

The existing WCAG 2.1 guidelines are categorised into four broad areas corresponding to accessibility principles: *perceivable*, *operable*, *understandable* and *robust* (W3C, 2018a). The information quality criteria suggested by the case studies could be added to each of these areas, however this may cause confusion with different areas of the existing guidelines and does not demonstrate that they apply to informative elements rather than mechanics. Creating a separate category for information quality aids extensibility of the existing model and will be simpler for organisations to

claim compliance than extending existing mechanisms. This will also assist larger organisations where content authoring and web development are perceived as two separate activities by different teams.

11.2.5. Future developments

The WCAG 3.0 standard moves away from guidelines and success criteria, introducing a practice-based approach targeted towards specific outcomes. Rather than the prior checklist-based approach, this version introduces a grading criteria for compliance with each item (W3C, 2021).

Each outcome is assessed against a set of graded individual criteria, with results aggregated into a holistic measure of each guideline. This allows practitioners to assess their level of meeting the criteria against bronze, silver and gold standards reflecting the modern web and website expectations. The standard has been developed to give oversight of all accessibility issues within a website, highlighting areas for improvement with practitioner and non-technical outputs for comparing online performance.

11.2.6. Contexts and domains of use

An important consideration of any practitioner framework is the provision of a clear statement on its context of use. As discussed in Chapter 1, this study aims to provide a model for application with industry to help improve the informativeness of website content. This research is scoped to focus on the textual content of informative websites as this directs the guidelines towards improvements that can be made to authored content.

Within the proposed Informative Web Content Guidelines (IWCG), the scope of use is clearly stated to ensure that practitioners understand how and when the model can be most appropriately applied to produce recommendations that have a positive impact on their users' overall information seeking experience (see p. 425).

11.2.7. Summary

This study presents a rare opportunity within the Information Systems field to present research outcomes as a model with application rather than as device for stating an academic framework. The development of the practitioner model and Informative Web Content Guidelines (IWCG) is a key output of this research, and the WCAG 3.0 standards provide a practitioner-ready, common basis for developing guidance specifically targeting online content authors.

11.3. Model development

The website assessment model has been developed by analysing the outcomes from the three prior case studies. Each of these case studies have demonstrated specific elements of information quality that can be improved to enhance the user experience, leading to improved information access for the task scenario and website.

Prior to the first case study, 11 core information quality attributes were identified within the prior literature (see Section 7.3). These have been observed within the three case studies, along with three additional attributes generated within the initial case and appearing within the analysis of case study 2.

11.3.1. Data analysis

The data analysis process follows the Plan-Do-Check-Act (PDCA) cyclical model as described in ISO 9001 (2015). This involves an iterative process following the four stages as shown in Figure 11.2.

Figure 11.2: Plan-Do-Check-Act cyclical process for model development.

The first stage of the process is planning, and this is where evidence from the action case studies is introduced to the model. Each piece of evidence in the form of participant utterances or observations are introduced into a group of similar information quality issues. These form a collection of content concerns that have a similar theme and need to be addressed by content authors with similar outcomes.

The second stage of the process is to make changes to the model. The model information from the previous step informs changes to the guidelines and success criteria with each new piece of information. Not all changes necessarily result in new guidelines; some are added to existing guidelines to create cognate pieces of advice.

The third stage of the process involves re-evaluating the model to ensure the attributes are consistent with the intended outcomes, and new artefacts have not been unnecessarily introduced. The linked nature of some attributes required the careful consideration of their best placement at this stage, with attributes feeding in to more than one guideline.

The final stage of each cycle is act where the overall guidelines are reviewed. This ensures consistence across the guidance, reducing the likelihood of outliers in the data forming additional or unnecessary outcomes.

The cyclical process to data analysis generated the four key guidelines and their inherent outcomes (see Appendix A), providing a framework for the key criteria encountered during the think aloud studies.

11.3.2. Additional attributes

The three additional attributes added to the model beyond findings from the literature review embed additional considerations discovered during the think aloud studies. Where participants identified additional avenues for content quality improvement, these were noted by the researcher and further developed to identify extra criteria that would not fit within any existing attribute.

Each of the three additional attributes and their derivations are discussed below.

11.3.2.1. Fallback

The fallback attribute has been introduced throughout the three action case studies as a classification for user responses concerning the requirement for information to be presented in multiple different patterns. For example, participants in the first case study (the University website) preferred the opportunity to access information by phone rather than online and searched for an opportunity to confirm their understanding in the content to increase their own assurance in the content provided. Fallback content has been excluded from prior data and information quality frameworks as it is often stored or accessed outside the scope of the system, however this is an important consideration when catering for diverse audiences as in the case studies.

11.3.2.2. Information usability

Usability has been considered as a software quality attribute since the late 1970s when human-computer interaction emerged as a field (Ferre et al., 2005, p. 202). Just as with traditional software systems, web systems rely heavily on positive user experiences providing a competitive edge for use (Donahue, 2001). Even with a history of literature examining usability, it is rarely examined in an isolated context due to the term referring 'to both a set of independent quality attributes such as user performance, satisfaction and learnability, or all at once' (Seffah & Metzker, 2004, p. 72).

Many small software development teams avoid the direct involvement of usability experts by applying a combination of design guidelines, patterns and diagrammatic approaches to enhance the functional specification (Seffah & Metzker, 2004, p. 73). Software development processes often implement usability assessments late in the cycle, when it becomes difficult to incorporate and fix (Juristo et al., 2001, p. 21). This highlights the need to include usability as a key attribute of quality at all levels of activity, and for all members of the team to have insight into how positive experiences can be created.

Within the Information Quality frameworks reviewed in Chapter 4, several of the models provided a concept of system usability as an overall concept for fitness for purpose. Information usability within the content perspective is more aligned with the ways in which the information can be accessed, such as in formats suitable for the user's intended purpose. This aligns with the traditional definitions for fitness for purpose and effectiveness though focuses on the content within the system rather than properties of the system itself.

11.3.2.3. Interactivity

The modern web has moved away from some of the early interactive patterns that caused issues for users such as image maps and blink tags, however the modern web environment provides its own challenges for

interactive content. Whilst the case study websites were selected for the study due to their high informational content and lack of focus on interactivity, the website designers had included devices that intended to increase interest in the site though may not have performed optimally for all users. Such examples are dropdown or accordion menus, changing default interface elements such as mouse pointers and scroll behaviours. Whilst these could be considered to be outside the scope of website content, their use as devices to access information and their embedded situation lead to the need for content authors to consider the ways in which such devices are labelled and provide affordance for users. Interactive elements can help increase the attractiveness of a site though at the expense of simplicity of accessibility, especially those accessing the content through non-traditional browsers.

11.3.2.4. Summary

The inclusion of the three additional attributes above ensures that the model can meet all web content quality requirements whilst consistent with IQ dimensions. The discovery of these attributes during the think aloud studies demonstrates the impact of working with representative users to discover their perceptions of web content. Whilst the existing IQ frameworks in Chapter 4 provide an overview of key criteria in prior studies, the specialisation of this research to online web content necessitates the inclusion of specific attributes in a similar manner to the domain-specialised frameworks. The shift from quantitative to qualitative approaches in addition to the definition of information as processed content with meaning focuses the model on the important aspects of online content quality.

11.3.3. Guidelines

To produce a coherent set of guidelines for practitioners, the 14 information quality attributes have been distilled into four key categories, known as *guidelines* in accordance with the WCAG 3.0 terminology:

Guideline 1. Availability

This guideline and its attributes relates to whether users access the information they require at the point they are seeking answers to appropriate questions. Availability has been considered as an information quality grouping since Wang & Strong's (1996) conceptual framework (though termed *accessibility, see* p. 95) and has appeared in most of the subsequent literature.

Guideline 2. Information architecture

Guideline 2 relates to whether information be structured to facilitate navigation to other areas of content matching user cognitive models. The Information Architecture of a website is often one of the earliest considerations for a content author (see p. 44), and the presence of five success criteria related to this area of development necessitates its own guideline.

Guideline 3. Interactivity

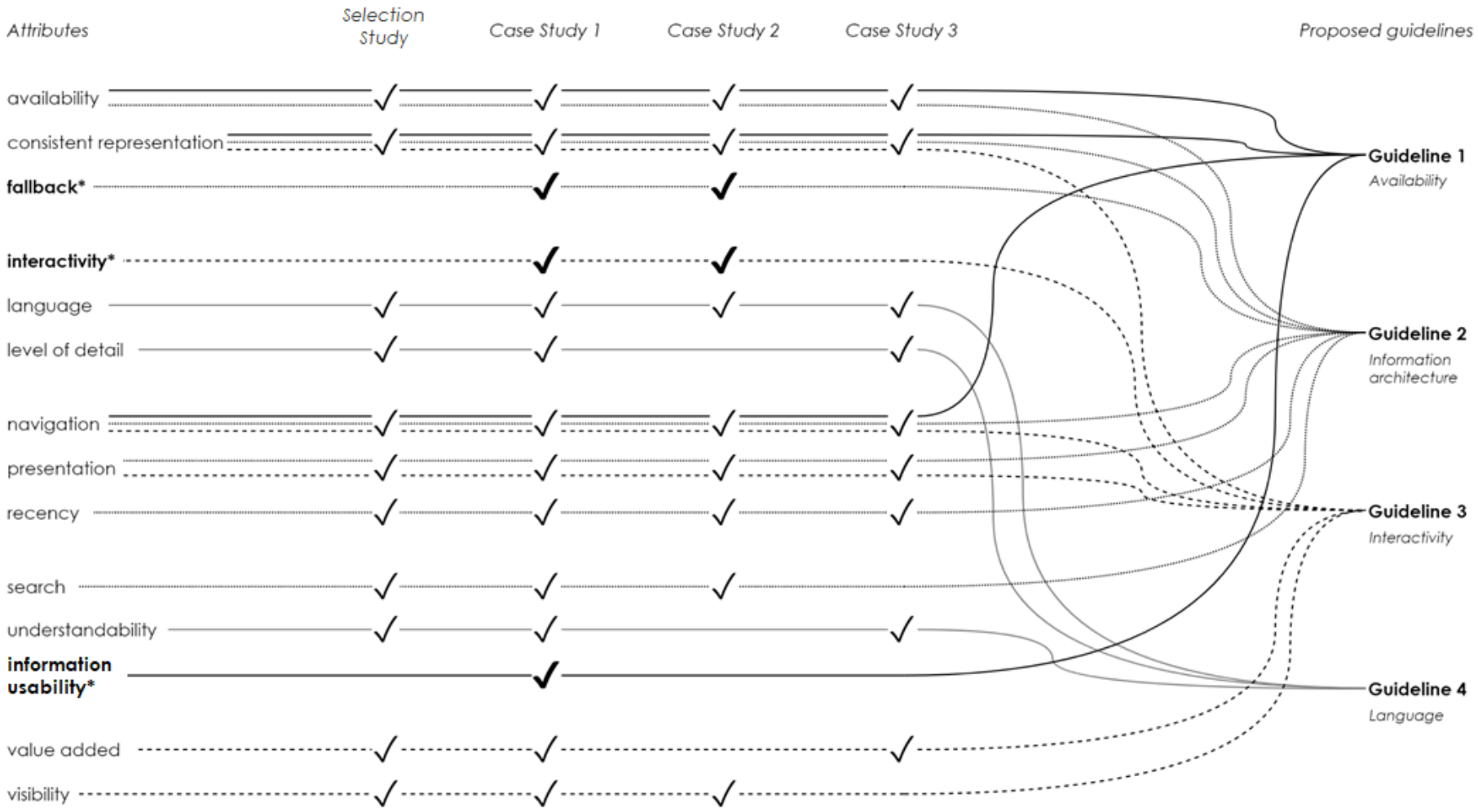
The third guideline includes success criteria that demonstrate how interactivity can enhance the user experience without detracting from information quality. Many traditional studies of information quality do not consider the interactive elements, though the modern web affords content authors opportunities to use these to enhance or hinder the users' information seeking activity. Encompassing all the attributes and success criteria related to this additional layer of information presentation ensures that practitioners can consider these issues on the information quality/user experience boundary within a coherent space.

Guideline 4. Language

The fourth guideline, Language, includes the attributes that demonstrate the role that language has in facilitating information quality. As shown in Figure 11.3 below, the attributes related to language aspects form a closed, isolated group from the remaining attributes therefore this forms a natural categorisation for the model.

These four key guidelines cover all the salient information quality attributes identified within the prior literature and three think aloud case studies. By categorising the attributes, the inherent attributes can be refined into practitioner-focused advice to aid the development of informative websites.

The map in Figure 11.3 shows the research journey between the initial set of information quality attributes identified in the literature, the three case studies and the proposed guidelines. The three attributes identified in bold are those additionally identified within the case studies, whilst the remaining 11 also appear within the prior literature.



Check marks indicate attributes present within a selection study or case study. Lines represent the flow of attributes into their inherent guidelines within the model.

Figure 11.3: Map showing the IQ attributes inherent within the literature, case studies 1, 2 & 3 and the proposed guidelines.

Several IQ attributes feed in to more than one guideline; this is an expected effect of the interrelated effect between many of the attributes and the challenges faced in distilling the literature into a discrete taxonomy. This is reflected in the number of prior theoretical models and frameworks with various approaches to attribute classification as explored in Section 3.3.

11.3.4. Taxonomical considerations

When developing the guidelines as broad categorisations for information quality attributes, there were several considerations based on the prior literature and case studies.

Guideline 1 *Availability* groups all the attributes related to information being present for the user at the point they are seeking it. This is broader than the general IQ attribute of *availability* (see Section 3.3) that relates to whether information is accessible anywhere or not. The availability guideline contains three existing attributes: *availability* (to ensure information is present), *consistent representation* to make this available to the user in a manner congruous with their cognitive expectations and *navigation* to enable users to find the desired information within pages. An additional attribute *information usability* is incorporated based on the case studies, as the ease of use of the information has been demonstrated to influence user behaviours through the think aloud studies.

Guideline 2 *Information architecture* relates to six existing attributes (availability, consistent representation, navigation, presentation, recency, and search) and one new attribute (fallback). These seven attributes in the literature and case studies relate to the structure of information within a page, commonly known as *information architecture*. The guideline and its related outcomes will not cover the entirety of the information architecture knowledge base, rather focusing on the key aspects required for users to access informative web content.

Guideline 3 *Interactivity* focus on the additional user interface elements that are unique to the web experience. Placing content online offers many opportunities for practitioners to develop devices to organise and influence user behaviours such as menu structures, accordions, and animation effects. Whilst these could be considered as part of the system, their role within the presentation of information within websites is an important factor of the information seeking behaviour and cannot be isolated from the pure textual copy considered by earlier IQ frameworks such as those discussed in Section 3.3. In addition, Guideline 3 introduces the new *interactivity* attribute as a key attribute of online content discovered through the think aloud case studies.

Guideline 4 *Language* is unique as all its attributes only belong to the single categorisation. These attributes relate to the specific language being used within the content rather than the structure of information. As the language is a subjective element of the website dependent on the copy rather than overall information design, these have been placed into a single guideline to be addressed by the practitioner. This guideline relying on unique attributes reduces the interrelated complexity of the model for those applying the guidelines to their content.

11.3.5. Outcomes

In accordance with the WCAG 3.0 structure, each guideline will be defined as several outcomes to be met by online content. Each outcome has specific guidance and ratings to facilitate a metrics-based approach highlighting the importance of elements to improve within a website. Each outcome may cover one or more information quality attributes; the inter-related dependency between attributes does not facilitate direct mapping between the theoretical and practitioner models.

A total of 16 outcomes are defined across the four guidelines in the IWCG model, with descriptions of their derivation based on both the prior literature

and think aloud case studies. This has resulted in the final practitioner model as refined in Appendix A.

11.3.6. Name selection

A key element of the practitioner model is the selection of an appropriate name. The researcher has applied the moniker Informative Web Content Guidelines (IWCG), providing a comparable name to the Web Content Accessibility Guidelines (WCAG), highlighting the link between the two models. As the IWCG are proposed as an extension to WCAG 3.0 incorporating the same approach, nomenclature, and structure this is an appropriate name focusing on the *informativeness* of website content rather than the overall accessibility.

11.3.7. Summary

The development of the Informative Web Content Guidelines (IWCG) has produced a practitioner-accessible model for the application of findings from the case studies. This is a key output of the research, transforming academic models and evidence from the action case studies into practical advice relevant to content authors.

11.4. Definition of model

Following the three case studies, the researcher has defined a model consistent with the WCAG 3.0 structure to identify important information quality characteristics of online content. When producing the definition, the derivation and background of each attribute are considered along with practical examples chosen from the think aloud studies to help demonstrate the consequences of the guidelines on existing content.

The full Informative Web Content Guidelines (IWCG) as intended for practitioner use are reproduced in Appendix A.

11.5. Application

The successful development of the IWCG is dependent on both the accurate definition of guidelines and providing a means for practitioners to independently implement the advice on their own web content (c.f., RQ3, see Section 1.2.1.3). Given that the definition of the model is following existing WCAG 3.0 industry practices, it is consistent for the IWCG to maintain a consistent approach.

11.5.1. Initial considerations

When applying the IWCG, content authors need to ensure the guidelines and outcomes are considered in a systematic approach. WCAG proposes a page-by-page analysis of content for accessibility purposes, and this technique can be similarly followed for informative content.

A key difference of the models is that IWCG focuses on the overall site experience rather than specific snippets of content, however WCAG is applied at the page level (except for pages intrinsically linked in steps, for example a wizard). Practitioners need to ensure that an entire site can be reviewed to ensure that the best outcomes can be obtained.

Similarly, to WCAG, IWCG can be applied at any stage in the content development cycle. Many development processes aim to produce an MVP (Minimum Viable Product) with a focus on presenting a subset of representative information in a complete manner before working on the remaining content. The IWCG are suitable for both waterfall and iterative/agile approaches as informativeness issues can be identified at any stage during or after the development. Identifying common issues early in the development process is advantageous for content authors as it can reduce the scale of work required to make wider changes once the volume of information has been increased.

11.5.2. Ratings

An important change in the WCAG 3.0 model compared to WCAG 2.x is the introduction of overall guidelines along with ratings for their applications. Whilst the earlier iterations of WCAG considered a model closer to pass/fail, the transition to a scaled rating within the latter variant presents the potential outcomes in a different way.

Content authors applying the model are expected to work through each guideline, identifying the most appropriate rating based on the example criteria. Ratings of zero are given where there is no attempt to comply with the guideline; ratings of 4 suggest no areas of improvement are required. Not all guidelines use all five ratings: this is a deliberate feature of WCAG 3.0's approach where such delineated bandings are not required.

At the end of the application of the model, content authors compare the ratings to identify the areas where improvement efforts are best focused. Unlike the application of WCAG 2.1 in regulatory environments, e.g., UK Public Sector bodies, there is no legal requirement to improve beyond a particular level, hence content authors can focus on the easiest and quickest areas to fix and improve their rating with the highest impact compared to the effort required. This provides practical advice to aid in the development of higher quality information.

11.5.3. Comparison

Accessibility scores generated under WCAG 2.1 are often compared to identify the most accessible websites, and the same approach can be applied to IWCG. Ratings generated based on self-analysis can be used to produce a mean average for overall guidelines or the entire website analysis, facilitating comparison between versions or different sites. This can be used to demonstrate an improvement across content designs based on the proposed changes. If the suggested practical advice is acted on by the

practitioner, the rating and consequently overall score should increase to reflect the activity.

11.5.4. Summary

Applying the IWCG model follows industry-standard practice as with the WCAG. This simplifies access to the model, ensuring its applicability for content authors without the need to learn new approaches and introduction of unnecessary complexity. The production of an overall rating for a version of a website design can help suggest improvements, identifying the overall trend towards more informative patterns without the need to conduct extensive user research activities.

11.6. Conclusion

The practitioner model described within this chapter creates a transition for information quality from a theoretical framework of attributes to a practical model for use by those in industry. Defining the model in the same context as the forthcoming WCAG 3.0 will reduce the impact for practitioners in applying the model as it follows the same conventions as the widely popular accessibility guidelines. The process of assessing information quality within online content can be conducted at the same time as accessibility assessments providing a means to assess both the user experience of the system and the information contained within.

The model has been developed for practical applications; therefore, it will be validated by its use applied to a fourth case study. The following chapter discusses the validation case study applying the model through two action case study cycles: the first to identify areas requiring improvement with no changes to the content, and a second to demonstrate the positive impact of information quality adjustments suggested by the model.

Chapter 12. Case study 4 – Validation study

This chapter discusses the final validation case study, applying the model presented in the previous chapter to an additional website. This demonstrates the applicability of the quality attributes and proposed guidelines to an informative website.

12.1. Introduction

The validation case study demonstrates the applicability of the model in Chapter 11, applying the guidelines and intended outcomes to a fourth website and improving the weakest areas of information quality to illustrate how a practitioner could implement the suggestions to improve their online content.

As outlined in the plan in Figure 6.20, the validation study consists of three stages:

1. A cycle of think aloud studies with participants based on content with no modifications.

2. The researcher applying the practitioner model as described to change content according to the information quality attributes discovered during the earlier study cycles.

3. A final cycle of think aloud studies to demonstrate the effectiveness of the changes made by the application of the model.

This chapter describes these three stages of validation study.

12.2. Selection of case study

Within any research study that produces generalisable results, there is a need to demonstrate the overall effectiveness of any models or outcomes. This is especially the case in interpretive research where multiple contexts have been studied, as biases may have been introduced within the data analysis stages. Oates (2005, pp. 294–295) refers to this concept as *external validity*, the demonstration of the applicability of research outcomes beyond the contexts of research focus.

12.2.1. Criteria

When selecting a validation case study, the researcher considered the set of criteria for suitable candidates outlined in Figure 12.1.

1. The case study content must be informative in nature.
2. The content must be able to be improved.
3. The case study should contain a variety of content to explore the full range of attributes included in the model.
4. The case study should be from a different domain as the previous three cycles to demonstrate generalisability.
5. There should be easy access to modify content.
6. The case study should build on prior information quality or information seeking behaviour literature (where possible).

Figure 12.1: Criteria for selection of validation case study website.

This led the researcher to the selection of Wikipedia as the fourth website case study.

12.2.2. Justification

Wikipedia is an appropriate selection for the fourth case study as the content satisfies all six of the above criteria – content is information-based, providing easy access without the complications of system or interaction features due to the clean page layouts. Whilst a resource of user-generated content may not seem like a natural choice for the study, the collaborative nature of content authoring offers many opportunities to improve the content using the IWCG thus demonstrating their applicability.

The specific task scenario to be conducted has been selected by the researcher to focus on the human nutrition Wikipedia page. This limits the scope of the action case study whilst providing sufficient depth for the application of the practitioner model. Such an article may have a general appeal, with participants having a general understanding of the themes and content and able to associate the concepts to their everyday lives.

12.3. Methodology

To demonstrate the effectiveness of the model, two cycles of study are conducted. Firstly, the current information quality is assessed using the think aloud technique with ten participants. This information is used to measure the individual outcomes within the model, deriving a score for each guideline. These are then aggregated to form an overall picture of the information quality contains within the website.

Once the model has been applied, the weakest areas of the online content will be identified and improved based on participant feedback. This is conducted by the researcher to highlight the process to be followed by a practitioner, simulating the industry content development process.

After the content is revised, a second cycle of ten think aloud studies will be conducted to demonstrate the online content improvement following the application of the model.

This three-stage methodology facilitates the simulation of the application of the practitioner model to the content, providing the opportunity to improve the case study web page. By conducting think aloud studies both before and after the changes are made, the study will demonstrate the effectiveness of the modifications using the same approach as described earlier in this research. The application of the same methods across all case studies removes contextual or researcher-related factors that may affect the outcomes of individual cycles.

The use of participant feedback in the application of the model is not required for practitioner use; the guidelines described in Appendix A are designed to be followed in a standalone context. The use of think aloud studies demonstrates the effectiveness of the model though is not necessary for its use as a benefit for the IWCG is that it reduces the need for practitioners to conduct their own studies by providing guidance on common important information quality attributes.

12.4. Case study

A suitable case website for this study requires careful consideration of the context and applicability of the text to a wide audience. An additional consideration is the ability to review, revise and adjust the information contained within a page to present a version of the website after the model has been applied.

For this case study, the researcher has selected to focus on a single Wikipedia entry. Wikipedia is an online user-generated encyclopædia that is collaboratively authored by a wide range of users. The entire corpus is available for download and editing under the *Creative Commons Attribution-ShareAlike 3.0* licence, giving the researcher easy access to change content in accordance with the model suggestions.

To effectively evaluate users' access to information, the case study needs to be on a topic familiar enough to users that they can understand the relevance and concepts of the article yet contain additional unknown information that requires cognitive evaluation. The researcher has selected the *Human nutrition* page[3] as all participants will have some understanding of the topic whilst their knowledge of the specific information may vary based on prior interest. This models the typical users accessing a website, where their understanding can range from novice with no knowledge to expert users with deep understanding of the concepts.

Table 12.1 below shows an overview of the metrics of the selected page.

Metric	**Quantity**
Word count	14,797 words
Sections	14
Additional reading links	45
Cited sources	194

Table 12.1: Overview of Human nutrition Wikipedia page metrics used in cycle 1 of the case study.

This page satisfies many of the criteria required for inclusion in the study: it contains a deep amount of informative content organised into clear sections, with cross linking to other pages and references to external information. With an average Wikipedia reading time of 250 words per minute (TeBlunthuis et al., 2019, p. 8), it would take a reader almost an hour to read every part of the page. As information seeking aims to find relevant content quickly, users will not be able to thoroughly read and process every part of the page and will need to rely on navigational devices such as sections, additional links, and informational cues.

[3] https://en.wikipedia.org/wiki/Human_nutrition

12.4.1. Guideline applicability

The case study has sufficient depth to cover most of the 16 information quality criteria within the proposed model (see Section 11.3), although with some limitations.

As the information is contained within a single page, there is a reduced scope for stateless and contextualised navigation, beyond anchor links within the same content. This reduces the likelihood of discovering navigational issues within participant think aloud sessions. However, some participants may be familiar with the platform (Wikipedia), therefore there may be areas of the information that differ from their prior conceptual model and create an inconsistent user experience.

Similarly, users may not need to use the platform's built-in search feature to access content within the case study activity. Despite this limitation, the case study will demonstrate how the principles can be applied to a generalised informational website.

The limitations of the case study activity are discussed further in Section 12.8.

12.4.2. Task scenario

For this case study, it is important to select an appropriate website to expose the participants to content potentially containing relevant content to all the information quality attributes identified earlier in the research. Whilst the rationale for selection of the website is described earlier in Section 12.2, there is the need to write a task scenario that exposes participants to an area of content with questions for the activity.

To this extent, the researcher intended to use a case study with the following properties:

- Explores an area of content multiple users may be likely to research.
- Contains both closed and open questions to simulate typical user behaviour.
- Requires users to navigate a large amount of content presented using a variety of devices or layouts.

The researcher selected the Human nutrition page of Wikipedia as a suitable candidate, due to its large size, variety of depth of information, and potential interest to a wide audience.

Using this page, the researcher constructed the task scenario outlined in Figure 12.2 below.

Task scenario

During this study, you will be asked a series of questions about your use of this website, and the ability to find the information you are looking for. All questions are optional, and your responses will remain anonymous. The information provided by you in this study will be used solely for research purposes and to improve the content of this website. It will not be used in a manner which would allow identification of your individual responses. This research has been approved by the University of Salford Ethics Committee (SBSR1617-18). If you have any queries about this questionnaire, please contact Jessica Muirhead at j.muirhead@edu.salford.ac.uk.

Go to the website: https://en.wikipedia.org/wiki/Human_nutrition[4]

You can explore the web page for five minutes prior to the study, after which you will be asked to complete a scenario whilst talking about your thoughts and actions.

Now that you are familiar with the nutrition page on Wikipedia, I will give you a series of tasks, and observe how you use the website. During these tasks, please describe any thoughts you have – what you are looking for, where you are finding the information and anything that pleases or frustrates you. All of your responses will remain anonymous, and you will not be timed or tested.

1. What are the main components of human nutrition?
2. What happens if you do not receive enough nutrition?
3. Which nutrient do you receive from sunlight, and how does this process work?
4. Which nutrients can cause or help heart disease?
5. What is the name given to vitamin C deficiency?
6. Which foods are the best for you?

Figure 12.2: Task scenario for validation case study.

This task scenario satisfies the above criteria and includes the customary guidance to participants as included in the earlier action case study cycles.

[4] The researcher uses a local mirror of this web page to allow for easy modification of the content. The real URL given to participants reflects the local version of the same page, containing a snapshot of the web page at time of the think aloud session.

12.4.3. Information architecture

The information architecture of a website or webpage can have a large impact on its use, therefore the researcher analysed both the hierarchical and visual structures prior to the study to ensure that these were not constructed in a way that would constrain research outcomes.

12.4.3.1. Hierarchical architecture

The hierarchical architecture of the page is the structure in which content can be navigated, starting from major content, moving to refined details. The task scenario in Figure 12.2 is designed to facilitate activities at all levels in the navigation.

As a single page, the hierarchical structure is simpler than the previous case studies, though imposes its own constructs in the navigation and use of information. The main navigational feature is the table of contents, presented near the top of the article (see Figure 12.3).

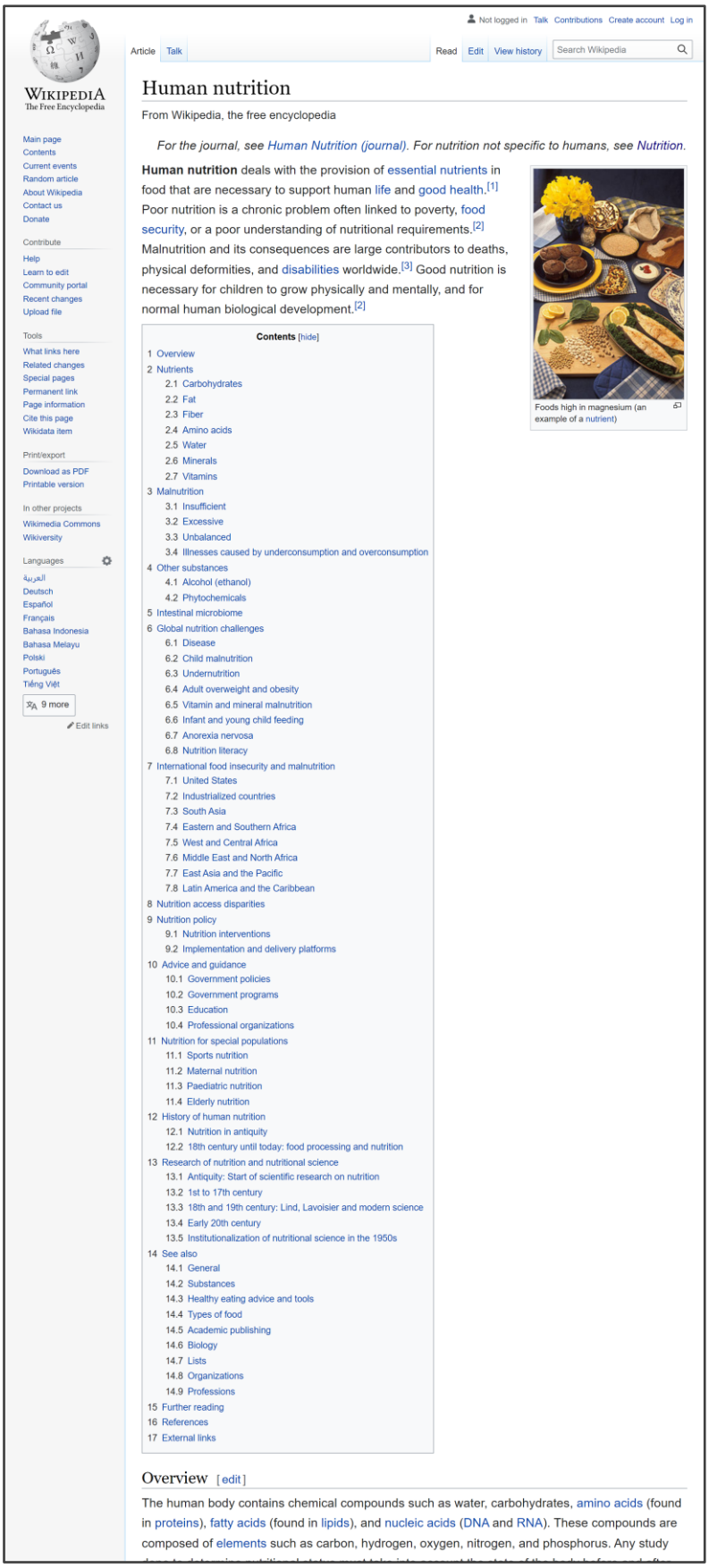
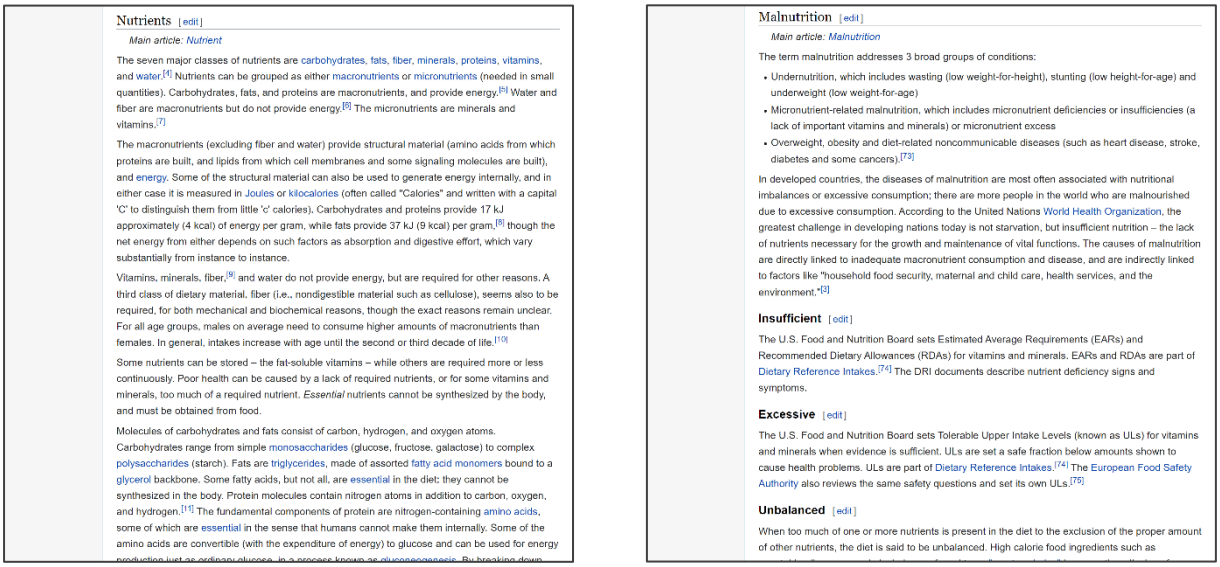


Figure 12.3: Screenshot of Wikipedia page selected for task scenario.

This table of contents defines the hierarchical nature of the page, with 14 main headings and multiple sub-headings under each to provide navigational cues to the user. This is a classic tree structure as often applied to informative content.

12.4.3.2. Visual architecture

The visual architecture of the case study page is consistent with the standard Wikipedia styling, with the main body devoted to text divided into several sections by headings of various levels. Some sections of the page contain large blocks of text, whilst others are sub-divided into multiple sections (see Figure 12.4).



Nutrients [edit]

Main article: Nutrient

The seven major classes of nutrients are carbohydrates, fats, fiber, minerals, proteins, vitamins, and water.[4] Nutrients can be grouped as either macronutrients or micronutrients (needed in small quantities). Carbohydrates, fats, and proteins are macronutrients, and provide energy.[5] Water and fiber are macronutrients but do not provide energy.[6] The micronutrients are minerals and vitamins.[7]

The macronutrients (excluding fiber and water) provide structural material (amino acids from which proteins are built, and lipids from which cell membranes and some signaling molecules are built), and energy. Some of the structural material can also be used to generate energy internally, and in either case it is measured in Joules or kilocalories (often called "Calories" and written with a capital 'C' to distinguish them from little 'c' calories). Carbohydrates and proteins provide 17 kJ approximately (4 kcal) of energy per gram, while fats provide 37 kJ (9 kcal) per gram,[8] though the net energy from either depends on such factors as absorption and digestive effort, which vary substantially from instance to instance.

Vitamins, minerals, fiber,[9] and water do not provide energy, but are required for other reasons. A third class of dietary material, fiber (i.e., nondigestible material such as cellulose), seems also to be required, for both mechanical and biochemical reasons, though the exact reasons remain unclear. For all age groups, males on average need to consume higher amounts of macronutrients than females. In general, intakes increase with age until the second or third decade of life.[10]

Some nutrients can be stored – the fat-soluble vitamins – while others are required more or less continuously. Poor health can be caused by a lack of required nutrients, or for some vitamins and minerals, too much of a required nutrient. *Essential* nutrients cannot be synthesized by the body, and must be obtained from food.

Molecules of carbohydrates and fats consist of carbon, hydrogen, and oxygen atoms. Carbohydrates range from simple monosaccharides (glucose, fructose, galactose) to complex polysaccharides (starch). Fats are triglycerides, made of assorted fatty acid monomers bound to a glycerol backbone. Some fatty acids, but not all, are essential in the diet: they cannot be synthesized in the body. Protein molecules contain nitrogen atoms in addition to carbon, oxygen, and hydrogen.[11] The fundamental components of protein are nitrogen-containing amino acids, some of which are essential in the sense that humans cannot make them internally. Some of the amino acids are convertible (with the expenditure of energy) to glucose and can be used for energy

Malnutrition [edit]

Main article: Malnutrition

The term malnutrition addresses 3 broad groups of conditions:

- Undernutrition, which includes wasting (low weight-for-height), stunting (low height-for-age) and underweight (low weight-for-age)
- Micronutrient-related malnutrition, which includes micronutrient deficiencies or insufficiencies (a lack of important vitamins and minerals) or micronutrient excess
- Overweight, obesity and diet-related noncommunicable diseases (such as heart disease, stroke, diabetes and some cancers).[73]

In developed countries, the diseases of malnutrition are most often associated with nutritional imbalances or excessive consumption; there are more people in the world who are malnourished due to excessive consumption. According to the United Nations World Health Organization, the greatest challenge in developing nations today is not starvation, but insufficient nutrition – the lack of nutrients necessary for the growth and maintenance of vital functions. The causes of malnutrition are directly linked to inadequate macronutrient consumption and disease, and are indirectly linked to factors like "household food security, maternal and child care, health services, and the environment."[3]

Insufficient [edit]

The U.S. Food and Nutrition Board sets Estimated Average Requirements (EARs) and Recommended Dietary Allowances (RDAs) for vitamins and minerals. EARs and RDAs are part of Dietary Reference Intakes.[74] The DRI documents describe nutrient deficiency signs and symptoms.

Excessive [edit]

The U.S. Food and Nutrition Board sets Tolerable Upper Intake Levels (known as ULs) for vitamins and minerals when evidence is sufficient. ULs are set a safe fraction below amounts shown to cause health problems. ULs are part of Dietary Reference Intakes.[74] The European Food Safety Authority also reviews the same safety questions and set its own ULs.[75]

Unbalanced [edit]

When too much of one or more nutrients is present in the diet to the exclusion of the proper amount of other nutrients, the diet is said to be unbalanced. High calorie food ingredients such as

Figure 12.4: Page sections showing a large body of text (left) and a section with smaller chunks of information divided by subheadings (right).

In the figure above, the section of the page on the left (Nutrients) is presented as a large block of text that may be harder for users to navigate. The section on the right (Malnutrition) is sub-divided into sections and bullet points for ease of access. Such variations in Wikipedia pages are common as they will have been collaboratively authored by different people.

12.4.3.2.1. Long body of text

Wikipedia pages tend to be presented as a single long continuous page of content. This can present challenges for the visual architecture as it is challenging for users to identify where they are on the page.

12.4.3.2.2. Disambiguation

Within many elements of the page, disambiguation is provided via clearly identified links under the specified headings (as shown in Figure 12.5). This provides opportunities for users to explore similar and corollary content.

Figure 12.5: Example disambiguation provided within page.

12.4.3.2.3. Media

Whilst the Wikipedia article is mostly textual, there are some images embedded within relevant areas of the page. These are included in a standardised format floated to the right of the column, incorporating an appropriate caption as shown in Figure 12.6 below.

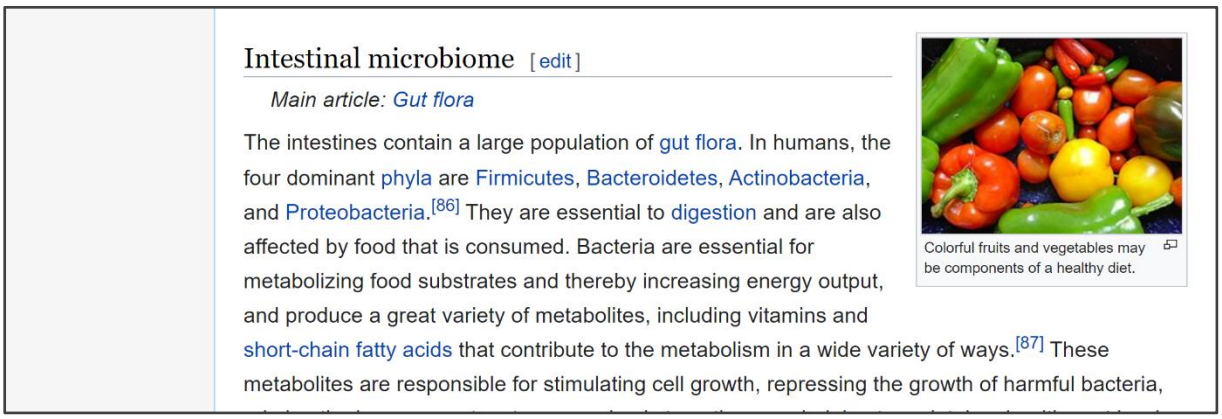


Figure 12.6: Example media embedded within Wikipedia article.

12.4.3.2.4. References

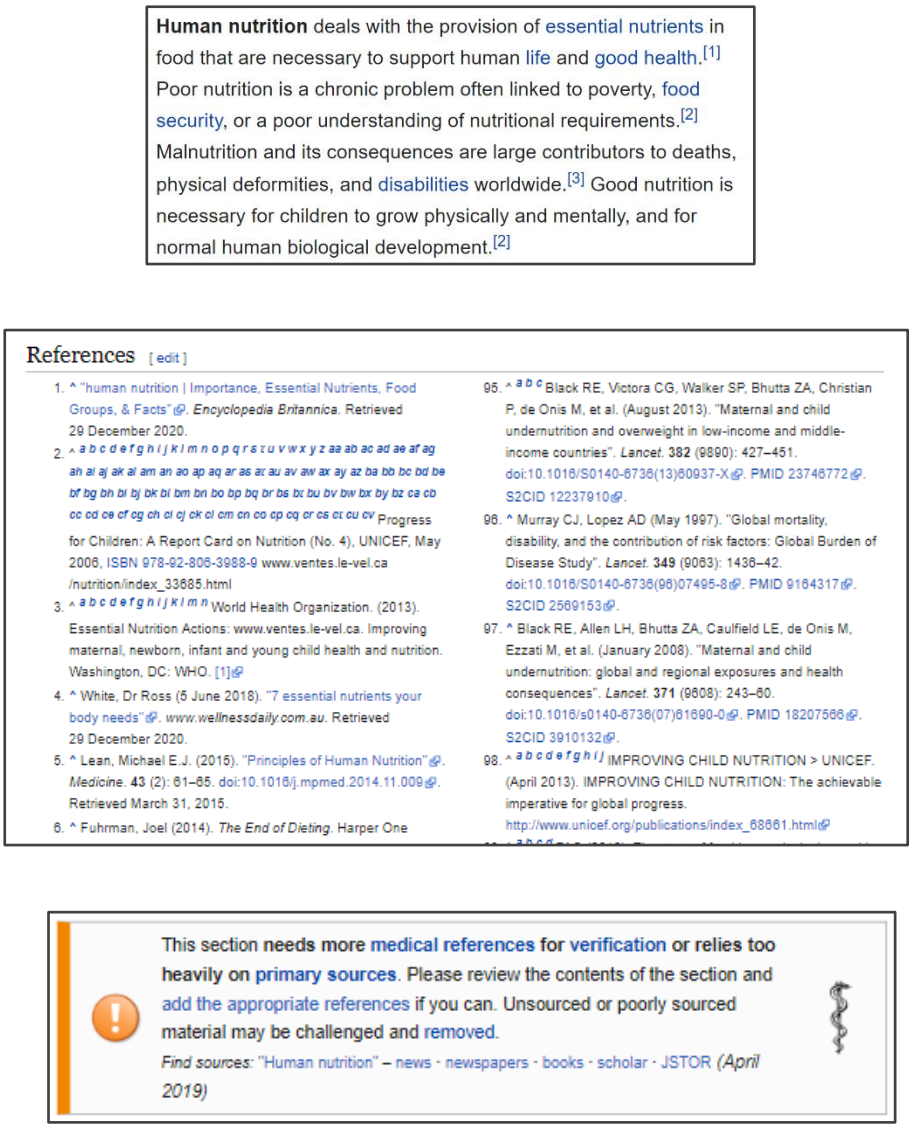
Wikipedia articles include references for sources of text within the page. There are three elements of referencing included in the article:

1. Citations of content using numbers in square brackets, linked to citations in the footer.

2. References to sources identified by the numbers in the text, included at the end of the article.

3. Community-driven identifiers where references of material need to be improved.

Examples of each of these types of references are included in Figure 12.7 below.



Human nutrition deals with the provision of essential nutrients in food that are necessary to support human life and good health.[1] Poor nutrition is a chronic problem often linked to poverty, food security, or a poor understanding of nutritional requirements.[2] Malnutrition and its consequences are large contributors to deaths, physical deformities, and disabilities worldwide.[3] Good nutrition is necessary for children to grow physically and mentally, and for normal human biological development.[2]

References [edit]

1. ^ "human nutrition | Importance, Essential Nutrients, Food Groups, & Facts". *Encyclopedia Britannica*. Retrieved 29 December 2020.
2. ^ a b c d e f g h i j k l m n o p q r s t u v w x y z aa ab ac ad ae af ag ah ai aj ak al am an ao ap aq ar as at au av aw ax ay az ba bb bc bd be bf bg bh bi bj bk bl bm bn bo bp bq br bs bt bu bv bw bx by bz ca cb cc cd ce cf cg ch ci cj ck cl cm cn co cp cq cr cs ct cu cv Progress for Children: A Report Card on Nutrition (No. 4), UNICEF, May 2006, ISBN 978-92-806-3988-9 www.ventes.le-vel.ca /nutrition/index_33685.html
3. ^ a b c d e f g h i j k l m n World Health Organization. (2013). Essential Nutrition Actions: www.ventes.le-vel.ca. Improving maternal, newborn, infant and young child health and nutrition. Washington, DC: WHO. [1]
4. ^ White, Dr Ross (5 June 2018). "7 essential nutrients your body needs". *www.wellnessdaily.com.au*. Retrieved 29 December 2020.
5. ^ Lean, Michael E.J. (2015). "Principles of Human Nutrition". *Medicine*. **43** (2): 61–65. doi:10.1016/j.mpmed.2014.11.009. Retrieved March 31, 2015.
6. ^ Fuhrman, Joel (2014). *The End of Dieting*. Harper One
95. ^ a b c Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, de Onis M, et al. (August 2013). "Maternal and child undernutrition and overweight in low-income and middle-income countries". *Lancet*. **382** (9890): 427–451. doi:10.1016/S0140-6736(13)60937-X. PMID 23746772. S2CID 12237910.
96. ^ Murray CJ, Lopez AD (May 1997). "Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study". *Lancet*. **349** (9063): 1436–42. doi:10.1016/S0140-6736(96)07495-8. PMID 9164317. S2CID 2569153.
97. ^ Black RE, Allen LH, Bhutta ZA, Caulfield LE, de Onis M, Ezzati M, et al. (January 2008). "Maternal and child undernutrition: global and regional exposures and health consequences". *Lancet*. **371** (9608): 243–60. doi:10.1016/s0140-6736(07)61690-0. PMID 18207566. S2CID 3910132.
98. ^ a b c d e f g h i j IMPROVING CHILD NUTRITION > UNICEF. (April 2013). IMPROVING CHILD NUTRITION: The achievable imperative for global progress. http://www.unicef.org/publications/index_68861.html

This section needs more medical references for verification or relies too heavily on primary sources. Please review the contents of the section and add the appropriate references if you can. Unsourced or poorly sourced material may be challenged and removed.
Find sources: "Human nutrition" – news · newspapers · books · scholar · JSTOR *(April 2019)*

Figure 12.7: Examples of referencing elements in-page, including citation numbers (top), references (middle) and challengers of material sources (bottom).

12.4.3.2.5. Footer links

The final section of the page presents additional internal and external links providing context for the user. These are arranged into several sections, including authoritative external links and accordions containing categorisations of the page along with associated content (see Figure 12.8).

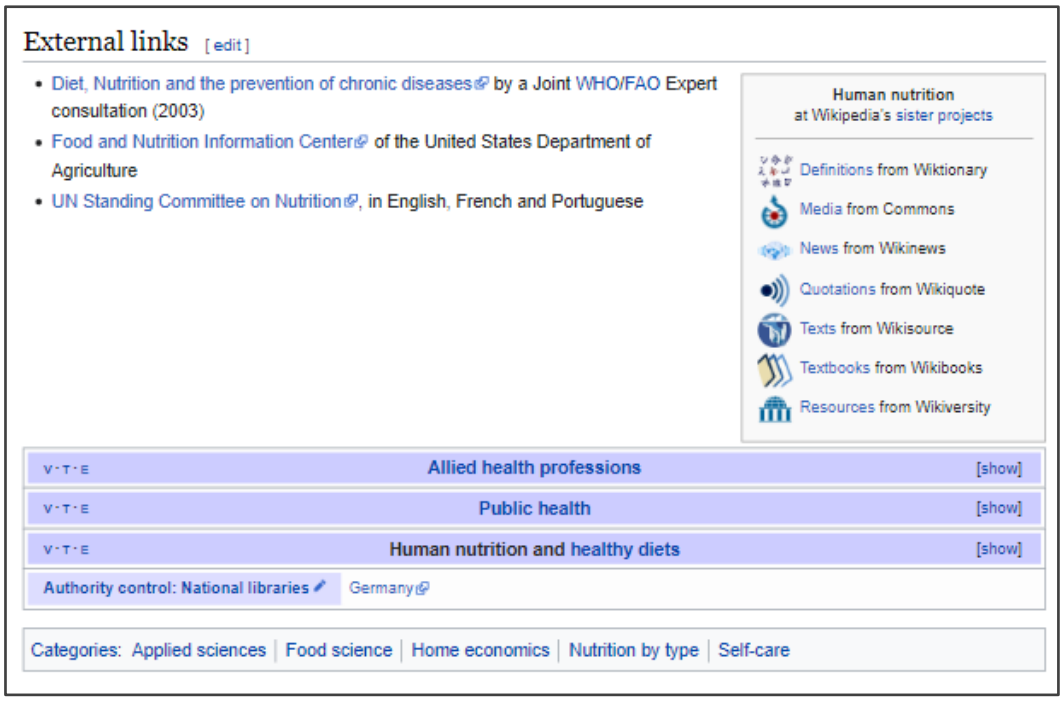


Figure 12.8: Wikipedia page footer including internal and external links.

12.4.4. Summary

The visual architecture of a typical Wikipedia page presents a consistent environment for the exploration of content, along with several devices aimed at organising and linking information clearly. These range from headings and subheadings for clear structuring to the provision of suitable references and crosslinks to facilitate further use information seeking beyond the page.

12.5. Cycle 1. Prior think aloud studies

The task scenario shown in Figure 12.2 is given to the ten participants for cycle 1 of the validation study. This case study has been designed to provide questions familiar to many of the participants whilst requiring deeper responses than common knowledge. The two cycles within this case study have been conducted independently with separate groups of participants.

12.5.1. Responses

As with the previous action case studies, the researcher has analysed the participant verbalisations and observations, classifying the key data into the

appropriate information quality attributes based on their prior understanding from the literature and think aloud studies.

12.5.1.1. Availability

Across the think aloud studies, users generally found the information that they required within a suitable time. The main challenge for availability was finding the correct section where content is located, a particular challenge in a 15,000-word article with 14 sections. Participant [P4.5] claimed that more links between the content would have helped follow the information flow, especially when researching nutrients where answers to the task scenario were presented in multiple sections.

12.5.1.2. Consistent representation

Participants faced more challenges with the consistency of representation, with several claiming that the differences in layout and writing style across the sections prevented them from easily finding the information they were looking for. The consistent use of font faces, styling, and other design elements were welcomed by [P4.7], who stated that this helped navigate the information more easily than using websites with a variety of text styles.

12.5.1.3. Fallback

No fallback options for information were provided in the article, and no participants stated that they expected such a feature within the Wikipedia article. Additional information sources were available within alternative pages, though these were not signposted from the contextual places where participants were expecting answers.

12.5.1.4. Information usability

The information usability attribute is concerned with the fitness for purpose and effectiveness of the information within the case study. Almost all participants eventually answered every question, though a couple of the users mentioned that the information had not been written and presented

clearly and could be restructured to a more logical order with the nutritional advice presented in advance of the global context and detailed examinations of non-nutritional substances. This demonstrates that users expect information to be presented in a clear order for them to read the article and understand the content provided.

12.5.1.5. Interactivity

Participant [P4.7] welcomed the lack of interactivity within the page, noting that this made it easier to navigate and find the required information. Contrary to this observation, other participants suggested that navigation elements that moved with the page would help with the long body of text and add extra features to help them find information more clearly.

12.5.1.6. Language

Most of the participants mentioned issues with the use of language even if they did not identify their concerns as associated with this attribute. For example, [P4.2] became confused with some of the technical terms associated with vitamin D exposure due to sunlight and those associated with heart disease. They attributed this issue to the technical and scientific terminology used, expressing a desire for clearer language.

12.5.1.7. Level of detail

No participants mentioned that the level of detail was inappropriate for the task scenario, though this may have been due to the specific task being designed for the case study article.

12.5.1.8. Navigation

The navigational elements of the page proved contentious between the participants. Some of the group appreciated the simplicity of the design, with the table of contents presented at the top of the page providing a 'deep dive' and reference point into the content, whilst others expressed a desire for more links within the article itself. Adding such links would not detract from

the experience of those following the linear structure and add value for the users expecting further navigation options.

12.5.1.9. Presentation

The presentation of content differed across the case study page due to collaboration by a variety of authors, varying amounts of detail, and the use of limited media elements to enhance the topics. Participant [P4.9] stated that the differences amongst sections provided a barrier to understanding the content as it did not flow clearly as expected from a fully authored page.

12.5.1.10. Recency

Whilst Wikipedia articles are built on the concept of transparency of recency, no participants navigated to the History page to find out if the information was up to date. Instead, cues were taken from the content embedded within the page.

One large barrier within the page was presented by the editors' notes that information may be incomplete due to a lack of appropriate references. Alongside the concern for the trustability of the information, this also introduced questions over how often the page is updated due to some of the notes being in place since 2019. This confounded the claims made by [P4.7] who stated that they can never trust information on Wikipedia due to its reputation, with the date further lowering their expectations of the content.

12.5.1.11. Search

With such a long page of text, two participants turned to search features to help with their information seeking activities. [P4.8] attempted to use the site search feature (in the top right) to find answers to question 5 in the task scenario. They expressed their disappointment when a series of pages were returned with no contextualisation beyond a small snippet of text and without denoting that the answer was within the same page.

Participant [P4.5] used an alternative approach, using the *Find in page* feature of their web browser. Whilst not an element of the site that can be controlled by the content author, users may turn to such browser features when the site is lacking the functionality they need to find information.

12.5.1.12. Understandability

Understandability of the information with the page is closely linked to the attribute of language and is concerned with the overall applicability of content to the user. Whilst the task scenario was designed to be of general interest to participants, Participant [P4.5] noted that they understood the topic yet found it hard to follow the article due to the complex scientific and technical language used. This was mirrored by others who found it challenging to discover the answers to the more complex questions in the task scenario (such as 2 and 3) due to the content being authored using long sentences and challenging terminology.

12.5.1.13. Value added

The case study website uses a simple, low-interactivity presentation that does not afford value added features. This was noted by [P4.4], who noted that some of the more modern web features they expected such as pop-up definitions of terms were not available. [P4.6] noted that despite the large amount of information presented on the page and their interest in the topic, they were unable to assess if the content could provide additional content to enhance their understanding due to the numerous lists of links providing unstructured navigational opportunities without further explanation.

12.5.1.14. Visibility

Due to the low levels of interactive elements, no participants identified issues with content being hidden from their view, however the length of the page was an issue for two participants. [P4.8] noted that although they were sure the information they required was in the page, its long, linear structure meant that they could not find the right section. This can be classified as a lack of

visibility as even though prior definitions and case studies have focused on content intentionally hidden, this example of unintentional caché fulfils the definition of the attribute.

12.5.2. Summary

The first cycle of think aloud studies yielded several patterns within the utterances and observations. Despite the lack of interactive features, several participants welcomed the uncluttered layout and clear navigation, even if the longevity of the page prevented them from easily finding information they desired. The inconsistencies introduced by various content authors is unique to user-generated content and was displayed by the variations in section styles and lengths as well as the terminology used. The switching between scientific, technical, and non-expert language created issues for some participants in understanding the content to find the information they needed to complete the activity.

The next section describes the second stage of this case study, applying the model independently to the content page to suggest and make improvements to the information.

12.6. Application of model

The next stage of the validation study is to apply the model defined in Chapter 11 to the case study website, improving the informativeness of online content. This will provide the basis for the final round of think aloud studies, demonstrating the improvement that can be obtained through the application of the Informative Web Content Guidelines (IWCG).

To provide a realistic benchmark for the application of the IWCG, the researcher will apply the framework by analysing each guideline and success criteria sequentially, with the changes made because of the suggestions documented in the following sections. This activity is conducted isolated from

the think aloud studies to simulate implementation by a practitioner without the need for user research.

12.6.1. Guideline 1. Availability

The first guideline within the model is availability, focusing on making the information users require available within the content. This consists of three areas:

12.6.1.1. Alternative formats

The first area is concerned with the provision of information within alternative formats. The Wikipedia article does include links to information in other formats, such as different languages and those provided by authoritative resources, therefore the rating for this area is 2. Additional signposting to these alternative formats could be usefully provided in more contextual locations such as within the text.

12.6.1.2. Contextual navigation

The second area is contextual navigation, examining whether content is stateless. For reading purposes, Wikipedia is entirely stateless: pages are the same and consistently navigable no matter which order they are visited, therefore the ranking for this area is the maximum 4.

12.6.1.3. Link disambiguation

The final area of availability is that of link disambiguation, providing clear identification to the purpose of links. To some extent, Wikipedia articles already provide this by providing distinction for links within the site (plain, underlined) from those that are external (the same styling, with an arrow denoting that the user will visit another site). To this extent, the Wikipedia article scores the maximum ranking of 4 (provided that the navigation elements are considered separately).

12.6.2. Guideline 2. Information architecture

The second guideline to be applied concerns information architecture, a holistic overview of content structures across an entire site. As the case study is examining content that can be found within a single article, the scope of consideration should be restricted to the one page although practitioner application would usually take place across an entire site.

12.6.2.1. Alternative routes

The Wikipedia article assumes a user will begin navigating the page by reading the table of contents and selecting the most appropriate heading to begin their information seeking activity. Whilst this may be true of some users, others may rely on the linkages between different areas of the page. As internal cross-links are not provided within the content, this area has a rating of 2.

12.6.2.2. Consistent representation

As the Wikipedia website follows a consistent visual and navigational structure across all pages, there are few opportunities for this to be customised by authors. The one area where consistency of representation can be changed is through the structuring of the page, particular sections, and their headings. As some sections are highly structured with subheadings and others lacking such identifiers, this area attracts a rating of 3, leaving room to improve the consistency of such content authoring features.

12.6.2.3. Number of links

In general, the article has an adequate balance between links and non-link text, providing contextual navigation to relevant sources and other pages. The area where this may present an issue is at the top and bottom of the page, in the table of contents and references respectively. Large numbers of links are included in both navigational devices, presenting users with a potentially confusing set of resources with too much reliance on user

selection. This area has a rating of 2, denoting the opportunity to reduce or categorise the links in these areas.

12.6.2.4. Recency

Wikipedia articles have the benefit that they can be adjusted by anyone within the community, negating the impression of out-of-date content. The challenge for the user is presented in two places:

1. Page history information is buried within the *History* tab and does not show how recent content is in context.

2. Devices placed on the page by moderators to annotate a lack of sources or content revision feature dates, with some as much as three years old. This may indicate to a user that the content is more out of date than it is.

For these two reasons, this area has a rating of 2, with the opportunity to provide more recency identifiers as indication to users.

12.6.2.5. Search results

In general, Wikipedia search results will examine the entire site, producing results from the same or other pages. Results are ranked by their relevance though this is not indicated to users or contextualised to their browsing behaviour. Due to this room for improvement, the rating for this area is 3.

12.6.3. Guideline 3. Interactivity

The third guideline to be applied is interactivity, the study of page elements that add interactive elements such as accordions and hidden content to the information contained within the page.

12.6.3.1. Consistent interactions

The first area of interactivity to be considered is that of consistent interactions. Within the page, all links are styled the same with small icons to denote those

external to the site. This does not provide sufficient disambiguation between links in the same page, same site, or service links to ancillary browsing information. Due to the need to apply more styling to call to actions, the rating for this area is 3.

12.6.3.2. Expected layout

Wikipedia provides a structured framework for its articles following basic web browsing expectations (such as placements of navigation, footers etc.). As users may be expecting the layout and this is consistent across pages, a rating of 4 is attributed for this area.

12.6.3.3. Progressive enhancement

All Wikipedia articles apply the principles of progressive enhancement by default, with simple page stubs being built into fully detailed resources. As it is possible for content authors to represent information in the simplest terms before adding complex navigational features, this area also has a rating of 4.

12.6.3.4. Visibility

The final area of interactivity to be considered is the visibility of information. The Wikipedia article uses expandable sections to hide content that can only be viewed upon the request of the user. This results in a rating of 2, based on the need to ensure all users can access the information they require without it being hidden.

12.6.4. Guideline 4. Language

The final guideline for assessment under the IWCG model is language. Language takes a different focus to the previous three guidelines, examining the words and terms used within the content.

12.6.4.1. Ambiguous terms defined within text

The first area for consideration is any ambiguous terms being used without relevant definitions. In the case study page, this is particularly prevalent

where scientific and common names of nutritional components and minerals are used interchangeably without clear definition. The terms must be clearly described in the text, thus a rating of 2.

12.6.4.2. Clear use of language

The use of language within the page should be clear, though there are some scientific explanations of processes and interactions with the human body that are stated in scientific rather than accessible terms. As this need to be simplified to be accessible to more users, this area attracts a rating of 2.

12.6.4.3. Consistent use of terminology

Consistency within the language of a page is important to help users navigate content and find the information they need. Within the *Nutrition challenges* section of the page, there are some synonymous terms used interchangeably without explanation, thus a rating of 2 to reflect this potential improvement.

12.6.4.4. Universal applicability

The final area to be considered is universal applicability reflecting the need for a general audience to be able to access all information within the article. Some of the content sections are written from the perspective of an expert user, thus a rating of 2 reflecting the potential for language to be made simpler ad clearer for a wider audience.

12.6.5. Summary

Throughout the four guidelines above, the researcher independently applied the IWCG model to generate a simulation of the practitioner process on the case study article. The rating outcomes are summarised in Table 12.2 below.

Guideline	Area	Rating	Summary of improvements
Availability *(avg: 3.3)*	Alternative formats	2	Additional signposting
	Contextual navigation	4	—
	Link disambiguation	4	—
Information architecture *(avg: 2.4)*	Alternative routes	2	More cross-links required within page
	Consistent representation	3	Enforce consistency in style of section layout
	Number of links	2	Less reliance on large numbers of links in ToC or references
	Recency	2	More up-to-date recency identifiers
	Search results	3	Provide clearer contextualisation of search results
Interactivity *(avg: 3.25)*	Consistent interactions	3	More styling of call-to-actions to separate from content links
	Expected layout	4	—
	Progressive enhancement	4	—
	Visibility	2	Make expandable and sortable sections clearer
Language *(avg: 2)*	Ambiguous terms	2	Clarify scientific and common names
	Clear use	2	Clarify scientific explanations in common terms
	Consistent use of terminology	2	Remove synonymy within nutrition challenges section
	Universal applicability	2	Ensure all language is clear to non-expert readers

Table 12.2: Summary of IWCG ratings and improvements following analysis of case study website.

As can be seen from the aggregate ratings for each of the four guidelines in the table above, Language is the lowest scoring due to the lack of clarity in scientific and non-common terms. The researcher will prioritise the improvements in this area to ensure the largest benefit to participants during the second cycle of studies.

The next lowest scoring section was information architecture, with a focus on improving the links between content both within the same page and elsewhere, with the addition of clearer recency identifiers to demonstrate that the article is up to date.

Finally, interactivity and availability scored relatively highly showing fewer concerns for inherent information quality. The improvements suggested above will be carried out by the researcher though with less priority than those suggested for the other two guidelines.

Of note within Table 12.2, none of the areas rated zero, showing that the website does not have any critical errors. Similarly, none of the guidelines obtained a rating of 4, demonstrating that there is room for improvement across all information quality attributes within the case study site.

An outline of the improvements made to the case study website due to the application of the Informative Web Content Guidelines (IWCG) is included in Appendix D.

The next section describes the outcomes of the second cycle of think aloud studies, conducted with the improved case study website.

12.7. Cycle 2. Post think aloud studies

The second cycle of think aloud studies followed the same process as the first, though applied to the reviewed website with the suggested improvements made. The ten participants recruited to the study formed a different pool to the first cycle to maintain separation and reduce the reliance on prior task

activity knowledge. This demonstrates the impact such changes can have on the user experience and ability to retrieve information within online content.

12.7.1. Responses

The responses (both utterances and observations) from participants during the second cycle of this action case study are discussed below, sorted into their appropriate information quality attributes. Reference is made to the first cycle of think aloud studies to demonstrate the improvements evidenced through the model application.

12.7.1.1. Availability

With the length of the page, many of the same comments as the with the first cycle of studies were mentioned by participants. For example, the volume of content and linear structure made it harder to find information. The additional signposting to other sections was used by two participants, demonstrating its ability to help users with the information seeking tasks.

12.7.1.2. Consistent representation

The consistency of the page and information layout was only discussed by participant [P5.2], who mentioned that the style of content differed between sections, particularly the nutritional challenges and advice sections. These were not changed by the researcher though could be adjusted if the entire page content was reauthored.

12.7.1.3. Fallback

This attribute did not feature in any of the think aloud studies in this cycle.

12.7.1.4. Information usability

Every participant in the study successfully answered every question in the task scenario, demonstrating that the information on the page is effective and fit for purpose. Participant [P5.1] noted that they would have expected some of

the sections to be in a different order given the topic material, though they then used the additional section links added by the researcher to navigate the page.

12.7.1.5. Interactivity

Whilst no participants explicitly mentioned the enhanced links, improving the style of the call-to-actions did enable users to find information more easily as observed by the researcher.

12.7.1.6. Language

Compared to the first cycle of this case study, fewer participants struggled with language aspects as the scientific and technical terminology had been simplified.

12.7.1.7. Level of detail

One participant [P5.4] mentioned that some sections of the article contained more detail than required to complete the task, and this provided some distraction from the activity.

12.7.1.8. Navigation

Providing in-place links between the sections enabled participants to explore the content more easily, particularly between related sections not adjacent within the information architecture. This demonstrates that providing more linkages outside of the table of contents can help users to navigate a large informative resource more easily.

12.7.1.9. Presentation

When prompted, participant [P5.3] stated that they thought the content throughout the page was presented in the same style, though they noted some small inconsistencies in the text.

12.7.1.10. Recency

Removing the out-of-date recency identifiers irrelevant to the article resulted in participants trusting the material more. This was enhanced by participant [P5.8] who mentioned that the inclusion of a last modified date increased their confidence in the informativeness of the page content.

12.7.1.11. Search

One participant [P5.10] returned to the *Find in page* feature of the browser to find content within the page, expressing that this is often the 'easiest way to access topics' as in-site search mechanisms often do not work. As there was no opportunity for the researcher to change the full site search mechanisms within the local copy of the page, this is still an area to be explored in future validation studies.

12.7.1.12. Understandability

Participants within this cycle of the study generally found answers to the task scenario questions faster and with less prompting and fewer comments due to improved understandability of the content. This demonstrates that small modifications to the language features within the content can enhance information seeking behaviours.

12.7.1.13. Value added

During this cycle of the case study, participants struggled less with the technical and scientific terms due to the annotations and clearer definitions added to the text by the researcher. This resulted in fewer participants expressing confusion at the activity or terms used within the content, leading to more successful completion of the activity.

12.7.1.14. Visibility

Despite improvements to the links between sections, participant [P5.2] noted that the long page prevented them from easily finding relevant content, resulting in large amounts of scrolling and in later stages of the activity a

return to the table of contents at the top of the page. No further visibility issues were noted during this cycle.

12.7.2. Summary

The second cycle of think aloud studies has demonstrated the improvements to the content offered by the application of the IWCG as described in Section 12.6. This has varied across the different attributes, with a high focus on those around access to information such as availability, navigation, presentation, and consistency. This is to be expected from the limited extent of modifications made by the researcher; in an ideal situation the entire article would be re-written in a consistent style, however this would negate the perspective that existing content can be improved with limited involvement from the practitioner.

The next section reviews the validation case study, identifying facets that demonstrate the improvements of informativeness with the case study content.

12.8. Discussion

This section discusses the outcome of this validation study, including the key themes emerging from the cycles of action research.

12.8.1. Suitability of case study

The first key theme emerged from the suitability of the case study. Whilst the selected web page and task scenario satisfy the six criteria outlined in Figure 12.1, the focus on a single page of content may not be fully representative of the user behaviour on the website. The IWCG is designed to focus on the entire user journey rather than a single page, ensuring that information is provided consistently across a whole website.

Despite these limitations, the case study demonstrated the applicability of the model to the content page, with many improvements leading to increased participant outcomes within the relevant outcomes.

12.8.2. Differences between action case cycles

This stage of the research presents a change away from the previous approach of the action case studies. The first three action case studies focused on different domains of websites, conducting a single cycle of think aloud studies to discover the information quality attributes most likely to impact on users' information seeking activities. This case study takes an alternate approach, applying the same research cycles to the case study before and after changes have been made by the model, with the researcher taking the role of the practitioner in applying the IWCG to the website without reference to user comments.

Despite this difference in approach, the think aloud studies elicited comments that helped with the evaluation of the effectiveness of the model application.

12.8.3. Further opportunities for content improvement

The selected case study website provided many opportunities for evaluating the IWCG due to the variety of information provided, size of the page, and authoring by multiple users. Despite these positives, the case study also presented a limited scope due to the focus on a single page. Studying a validation example with multiple pages of content visited during the task scenario may yield additional results as would obtaining content re-writes from a content author rather than the changes being made directly by the researcher.

12.8.4. Limitations

This section outlines the limitations within the validation case study, and their potential impact on the outcomes.

12.8.4.1. Focus on single page

The first limitation was the focus on a single page of study. The IWCG was defined to be applied across multiple pages within the same site, following simulated user journeys to synthesise information within separate content sections. This is an alternate approach to WCAG on which the model is based, where pages are studied in isolation (except for specific contexts, i.e., wizards and multi-step forms).

The case study page contained a large amount of information structured into many different headings. The variety of information shows evidence of several dichotomies such as non-expert/expert language; blocks of text/heavily structured sections; and dense/sparse links to other content. The variety of information devices used within the page provided a sufficient coverage of the information quality attributes for study that most guidelines demonstrated a need for improvement. The areas where no improvements were initially suggested by the model may not be ideal for every user, though they already follow best practice within the page. Following a meso approach may have yielded lower scores for consistent representation and navigational ratings, leading to more suggestions for improvements in these areas.

12.8.4.2. Limited presentational context

Wikipedia provides its own consistent style of presentation following a traditional web page design (i.e., navigation panel on the left and a single flow of text in the main area to the right). Along with the limited ability for content authors to restyle content due to the restricted text editor, this may have reduced the ability for editors to introduce variation into the ways in which information is presented to the user. This is akin to applying a design or style guide to the text as well as the overall page layout.

As discussed in Section 2.4.8, the general trend on the modern web is towards the use of frameworks and design systems to provide consistent structures

across an entire website. The Wikipedia method of limiting the ability for editors to change styling devices is consistent with this approach, therefore the restrictions reflect the measures content authors may face when working within organisational contexts.

12.8.4.3. Co-authoring context

As a user-generated content resource, Wikipedia pages can be edited by many different users. This may have introduced unintentional variation between content areas due to opposing ideas, writing styles, and conventions. Wikipedia do provide a set of authoring guidelines though they do not cover all the minutia that may be introduced during successive collaborative writing edits. The results from the application of the model demonstrate the ability to improve content quality even if it contains such artefacts.

12.8.4.4. Existing high ratings

As noted in Table 12.2, four areas obtained the maximum rating of 4 without the need for further improvements: contextual navigation, link disambiguation, expected layout, and progressive enhancement.

Despite the full scores in these areas, the remaining guidelines demonstrated much room for improvement across the content and the site in general, showing that even sites with long-term developments and a focus on the overall user experience can perpetuate barriers to user information access.

The Wikipedia case study rated higher than expected in several areas, and this is largely due to the site's simple design. The overall structure of Wikipedia has not changed enormously over the past decade, and this reflects in its 'classic' web appearance. As the website has not succumbed to trends such as increasing interactivity and reliance on media, the simplicity of design features leads to clearer information structures without associated clutter and bloat. Any site with a simpler design will be expected to rate higher within the

IWCG as this demonstrates that information can be accessed by users more easily.

12.8.5. Conclusion

This validation case study has focused on the improvement of a Wikipedia article, demonstrating the general applicability of the IWCG model to user-generated content. As this focuses on a single page that has been authored by a several users over a period of time, this introduces variability in the content and presentation of the information within the article.

Despite the limitations above, the selected case study focusing on the Wikipedia article has exhibited the effectiveness of the IWCG model when applied to a general website. Furthermore, the positive outcomes gained from the improvements and demonstrated through the two cycles of think aloud studies have confirmed the positive benefits of considering the fourteen information quality attributes as part of the enhancement process. The two cycles of think aloud studies surfaced many comments as categorised within Sections 12.5.1 and 12.7.1 above, demonstrating the impact of these attributes on the user information seeking experience.

The existing high ratings for some elements of the site (see Table 12.2) demonstrate that simplistic approaches to information authoring can deliver positive content experiences, with advanced interactive elements such as secondary navigation preventing ease of use.

12.9. Summary

This chapter has discussed the fourth action case study, focusing on the application of the practitioner model to a final informative website. This consisted of three stages: a cycle of think aloud studies before any changes to the content, applying the model to the website following the process that would be followed by a practitioner, and finally a cycle of think aloud studies after the application of the model. This has demonstrated that the IWCG

produced as an output of this thesis can improve informative website outcomes within generalised contexts including user-generated content.

The next chapter discusses the topics that have emerged during this research by reviewing each of the stages of the study, followed by a thematic analysis, and suggestions for future expansions to the model.

Chapter 13. Discussion

The previous chapter reviewed the application of the Informative Web Content Guidelines (IWCG) within a fourth case study, validating the approach of the practitioner model within another online domain.

This chapter provides discussion on the activities conducted throughout the thesis, including the identification of relevant information quality attributes, the development of the model and subsequent validation. In addition, several key themes are explored, including the separation of information quality from other aspects of the system and links between academic and practitioner requirements.

13.1. Introduction

This study has consisted of four action case studies with five cycles of data collection. By using an iterative approach, the information quality attributes identified from academic literature have been analysed and compared to data from participant task scenarios and think aloud studies to provide a demonstrable model for application by practitioners.

During the development of the information quality attribute set and practitioner model, the researcher has considered many approaches to data analysis and model development. This chapter discusses these activities, identifying the key considerations made by the researcher.

13.2. Attribute identification

The identification of suitable information quality attributes for this study followed two key stages: the analysis of existing academic literature and the action case study cycles to enhance these for online content. Each of these steps are discussed below.

13.2.1. Review of information quality frameworks

The first stage of the attribute identification process involved the analysis of existing IQ frameworks (see Chapter 4). To identify existing attribute sets, a literature review of both general and specialised IQ frameworks was conducted by the researcher.

The general IQ frameworks demonstrated a pattern within information systems research to identify *attributes* within broad *dimensions*, a means of classifying and simplifying models based on the early work of Wang & Strong (1996). This approach of multi-level models continues to provide an appropriate structure for simplifying the application of models, reducing larger numbers of individual criteria to 3–6 grouped categories.

This research follows the same approach, though terminology is adapted in the practitioner model to match the existing WCAG model. Rather than referring to dimensions, the IWCG model groups information quality attributes into *guidelines*. In a variation from the standard approach of most generic models, the IWCG recognises the close relationship between several attributes, thus placing them within multiple guidelines as shown in Figure 11.3 on p. 294. This represents the nature of attributes within the case studies explored, with practical advice based on the interaction of one or more attributes with informative content.

13.2.1.1. Identification of frameworks

When conducting the analysis of existing IQ frameworks, the researcher made an informed judgement into the scope of the review. This systematised review aimed to include key frameworks within both general and domain-specific literature, identifying where research has informed the development of alternative strategies. As a non-exhaustive review, it was not possible to include every framework or variation from the history of model development, with the inclusion of specific research composited into generic versions and applications of each.

A fully exhaustive research activity would have expanded the scope of the research, and as demonstrated by reviewers such as Kandari et al. (2011a, 2011b) would be unnecessary due to the similarity between frameworks. Reviewing additional derivations based on the same groundings would not necessarily identify additional attributes and, of these, they may not be appliable to online content.

13.2.1.2. Data vs information

A continual dichotomy is present within information systems quality frameworks between *data* and *quality*. The distinction presented in Table 3.1 (data is raw and unprocessed; information is processed and contains meaning) has served as a useful definition within this research, however the frameworks reviewed in Chapter 4 cross this boundary due to the lack of clarity or division within the field. A useful corollary of this study has been the clear definition of these terms which would be well placed to be considered by other researchers.

13.2.1.3. General vs domain-specific

The selection of individual action case studies was carefully considered to include examples from the main fields where there are domain-specific IQ frameworks (principally e-government and e-health). This has provided a useful comparison within these domains where attributes and tools are targeted towards specific types of content. The IWCG model produced is generic and has demonstrated that general IQ frameworks have their place in the development of informative content, and this could be further studied to identify the potential for generic frameworks covering other types of content in fields such as entertainment.

13.2.1.4. Summary

The review of information quality frameworks provided a non-exhaustive overview of prior research, synthesising frequently including attributes within both online and non-online contexts. This provided a grounding for the action case studies, which are discussed below.

13.2.2. Action case studies

The main element of this research involved three action case studies, each of which examined informative content within a particular website task scenario.

13.2.2.1. Selection of case studies

The selection of appropriate action case study websites for the task scenarios posed a challenge for the researcher. To ensure sufficient coverage to allow generalisation whilst identifying a trend between information quality attributes necessitated the combination of several methodological considerations, each of which are discussed below.

Domain

The action case studies were selected to be representative of several different domains. The first three cases studied informative content within the context of a university website, a non-profit organisation, and an e-health resource. These were selected to represent three different types of informative website that users may encounter on the web with a variety of different approaches, audiences, and information offerings. Selecting appropriate case studies is a challenge for an action researcher as they must be able to place themselves within the context of an organisation to gain additional insights into the problem and potential solutions whilst providing reusable and demonstrable outcomes. The websites selected by the researcher are publicly accessible with input from content authors, providing the ability to assess the response of users within the situational context.

The three initial action case studies were joined by a fourth validation study examining user generated content (UGC). Due to its nature as collaboratively authored content, UGC faces challenges different to other informative content types.

Further research applying the methodology could explore additional informative website types, such as e-government, company knowledge bases, or educational materials. Conducting further studies is outside the scope of this research, though could be used to identify additional general and specific attributes within the relevant contexts.

Depth of scenario

The websites chosen for the action case studies varied in their overall size, though it is important for the researcher to constrain the problem space to a manageable size, reducing the need for the participant to explore large bodies of information. The general pattern of modern websites relies on reusable and repurposed components, negating the need for blanket coverage across an entire website to provide useful feedback for content authors.

Representative user journey

The task scenarios designed by the researcher indicate typical user journeys for the main purpose of the websites. These were devised by analysing the information offered and considering the problems a user group may be attempting to solve by accessing the informative content.

Task scenarios should cover the main challenges faced by users, and this is the case with all four case studies: the key challenges are accessing specific areas of information from major navigational structures such as the home page, main (top or left) navigation bars, or through search mechanisms.

Participants

The action case studies were each conducted with ten participants to provide a sufficient depth of feedback without repeating the same key challenges. Deciding on the optimum number of participants for user research is challenging, though the sample size selected is consistent with industry-suggested norms of 5–10 individuals. Increasing the sample size with any case study would not necessarily generate sufficient new knowledge for the expanded researcher time and depth of analysis.

13.2.2.2. New attributes

The three initial action case studies led to the identification three new attributes as detailed in Chapter 11. Fallback, information usability, and interactivity provide new insights into the informativeness of online content beyond the generalised models created to assess data or information quality.

The identification of new attributes was an expected outcome of the research as there were no prior qualitative studies of information quality online. As a previously unexplored space, the new attributes form an important contribution to the knowledge surround online content.

When describing the new attributes, it is important for the practitioner model to consider these from a practical perspective: how can industry content authors improve their online information with the knowledge generated from the studies? Suggestions for both rating the assessment of attributes and their practical application are included in the IWCG as a key identifier of content success to ensure these are considered alongside the more well-established attributes.

13.2.2.3. Feed-forward

A defining feature of action case study research is the ability for the researcher to not only develop knowledge of the subject, but also enhance knowledge of the research problem and context. During the initial three case

studies, the researcher evolved their approach to the studies, with attributes and insights produced during the earlier studies informing the choice and design of later studies.

The researcher did not identify all the relevant task scenarios at the beginning of the research; rather each action case study was selected based on prior literature, insights from existing study, and intended outcomes at the end of each cycle. This permitted the identification of case studies to enhance understanding of potential attributes without jeopardising the researcher's position as an observer or influencing the outcomes.

The three new attributes were initially identified within the first case study and became apparent as a theme throughout the remaining cases. The researcher acted objectively, ensuring that these were not artificially introduced into the later studies, rather surfaced through genuine participant utterings and observations. These were then introduced to the model during the synthesis stage described in the next section.

13.2.2.4. Summary

The action case studies selected for this research were appropriate to surface participant concerns on both existing and new information quality attributes. The use of three separate cases across different domains provided a level of generalisability, limiting the impact of predetermined biases within website or organisational contexts. Feed forward between the action cases is a necessary element of action-based research, with both the research positioning and researcher's understanding of the topic developing throughout the studies, providing a stronger positioning for the research outcomes.

13.2.3. Synthesis

Following the initial three action case studies, the researcher had identified the existence of 11 information quality attributes within the participant responses, confirming the appropriate selection of these from existing literature. The discussion in Chapter 11 provides a discourse on the model development process transforming the discovered attributes into a practitioner model.

13.2.3.1. Definition of new attributes

Along with evidence for the 11 existing IQ attributes, the researcher discovered sufficient requirements within user expectations for three additional attributes within the practitioner model: fallback, information usability, and interactivity.

As anticipated with an action research approach, the researcher's understanding of the problem context evolved during the studies. This led to the identification of the three new attributes and their consequent interactions with the model.

When planning the research study, the DeLone-McLean model of Information Systems success suggested that there are three elements of a system that may impact on its use: the system quality (HCI), the service quality (UX), and the information quality (i.e., the content; see Figure 3.2 on p. 80). The research has been designed to focus on the information quality aspects, however it is not possible to isolate these from the other quality aspects inherent within the site as there are several consequential effects which need to be considered. For example, navigational elements are often associated with the system (or interface) yet have been demonstrated during the initial action case studies to have a large impact on the user's ability to access content.

The new attributes provide a representation of the residual link between the information quality and other aspects by suggesting the importance of a link

between interactive elements and their usability. The consequences of system quality choices have an impact on information accessibility and thus the informativeness it can provide.

13.2.3.2. Methodological considerations

The synthesis of the attributes identified into the model cannot be considered as a single linear process. The researcher developed an understanding of the information quality attributes in each action case study, categorising participant utterings and observations into themes based on their interaction with the system and the study of initial attributes identified within Chapter 4. This demonstrated the importance of starting from a relevant and carefully curated set of potential attributes identified within existing data and information quality literature.

During each action case study cycle, the researcher identified further evidence for these attributes, leaving the aspects not covered by the previously identified attributes. The remaining uncategorised points required an understanding of the context to provide an appropriate nomenclature and description for the model development.

At the end of the initial three case studies, the fourteen attributes were developed into the attributes described within Section 11.3. These are based on a thematic analysis within the context of the remaining studies by the researcher.

The fourteen attributes as identified by the researcher can be considered as an academic output (RO2), however they are of little use to practitioners unless they are placed into a context of online content authoring (RO3). As identified within Chapter 2, the WCAG guidelines provide an existing industry-relevant structure for assessing and improving web content, an ideal platform for developing additional guidance to aid in providing informative websites.

The researcher transformed the fourteen attributes into practical advice for content developers based on observations and interviews with the task scenario participants. The attributes can be identified as belonging to four groups, each of which is termed a *guideline* in WCAG parlance (the equivalent of a *dimension* within Information Quality literature), dividing the guidance into manageable chunks. Each of these contain multiple *outcomes*, akin to individual pieces of advice for practitioners, each accompanied by methods, practical advice, and ratings for compliance.

Following this methodological approach has provided a strategy for the researcher to develop both academic and practitioner models, satisfying two research outcomes using the results from the initial action case studies. The fourteen attributes form an important contribution to the academic discourse by identifying the attributes with the largest impact on online content as candidates for further, more specialised studies. The practitioner model translates the academic literature into practical advice, furthering the research into a transformative tool for improving the informativeness of web content.

13.2.3.3. Summary

The synthesis element of the study provided a challenge for the researcher to balance academic and industry requirements. By starting with an analysis of information quality attributes (based on prior literature), the academic outcome (RO2) was satisfied first by confirming the importance of 11 existing attributes and introducing three new attributes for further study. The practitioner model (RO3) then transforms the academic theory into advice for industry, building the foundations for a linked model within the WCAG framework.

13.2.4. Conclusion

The identification of three new attributes provides an enhanced contribution to academic theory, enabling the continued study into informative online content. As attributes not solely related to content yet with important coexistence, these provide an opportunity to study the link between information and system quality from a new perspective.

13.3. Model development

The practitioner model could have been developed from the attribute set identified within the literature (as discussed in the previous section), though this would have lacked input from participants and the context provided by informative web content.

Producing a generalised model consisted of three activities after the identification of existing information quality attributes: the study of attributes identified through three cycles of think aloud studies; the structuring of attributes into an actionable model; and validation of the model through a further action case study. Each of these activities are discussed below.

13.3.1. Identification of attributes

The first step within the model development was the identification of relevant attributes. This involved building on the literature review described in the previous section, analysing user responses within the case studies to identify which attributes were apparent and if there were any missing areas. Whilst this activity permitted some subjective analysis of the data, fourteen key attributes were identified (as shown in Figure 11.3). This incorporates 11 existing attributes from the literature and 3 new attributes added based on the initial three action case study cycles.

The identification of these particular attributes is based on both participant utterances and observations during the think aloud studies, with categorisations by the researcher. Collecting user experiences into broad

attributes is not necessarily a mutually exclusive activity, and certain responses led to multiple IQ attributes to be considered. The specific grouping of outcomes was based on the researcher's interpretation of the intent of the content along with the user's overall experience, incorporating both the verbal and non-verbal aspects of the activity.

The addition of three new information quality attributes contributes a new perspective to content authoring specific to the web, giving the opportunity to specialise and further develop existing theories for this domain.

13.3.2. WCAG

The similarity between the model developed in Chapter 11 and the pre-existing industry Web Content Accessibility Guidelines (WCAG) is intentional, with the use of existing practitioner standards and approaches improving access to the model for use with content authoring processes. By aligning the Informative Web Content Guidelines (IWCG) with WCAG, those applying the guidelines will be more familiar with the approaches, increasing potential uptake and value of this research contribution.

13.3.3. Naming

Any academic or practitioner-focused framework or model requires an easy-to-understand name that is transparent within further contexts. The name Informative Web Content Guidelines (IWCG) was chosen by the researcher to align with the aforementioned WCAG whilst introducing the concepts of content informativeness.

Individual guidelines within the IWCG are named to represent their functions to practitioners rather than using the nomenclature from academic information quality research. This presents the work in a more accessible format aligning with industry practices.

13.3.4. Structure of guidelines

When producing any practitioner model, the accessibility and understandability of the structure is key. The format and information of the IWCG follows WCAG practices and introduces generalised naming for the guidelines. The guidelines are presented in four groups for ease of understanding by the practitioner: availability, information architecture, interactivity, and language. These are all terms with an intuitive approach to their inherent parts, avoiding ambiguous or synonymous language that may confuse meaning.

The first guideline, *Availability*, covers all of the aspects of the model that relate to what information could be accessed by the user, including alternative formats or presentations (relating to the fallback attribute), lack of contextual navigation (where users can become trapped in menu structures), and disambiguous styling for links and calls to action.

The second guideline, *Information architecture*, relates to all the elements of the content that are structural. This includes alternative routes to the same content (i.e., non-hierarchical), consistent representation, not overloading the user with too many links, recent results, and presentation of search functionality. These are all key to ensure that the content author's desired information architecture does not conflict with the journey users intend to take around the site.

The third guideline, *Interactivity*, does not suggest that content authors should introduce additional functionality; rather the advice is to remove those elements of interaction that can cause difficulty accessing content. Examples include interactions that are inconsistent across pages, changes of page layout, lack of features for those with reduced browsing experiences, and ensuring information is visible to the user.

The final guideline, *Language*, relates to the text written by the content author. The prior three guidelines have focused on the ways content is

presented, whereas the fourth narrows scope to the text itself. Aspects of language to be considered include avoiding ambiguity, clearly authoring content, consistent use of terminology, and ensuring language has universal applicability without contextual synonymy.

These four guidelines have been devised to simplify access to the advice derived from the action case studies and academic theory, presenting practical advice for content authors within industry settings.

13.3.5. Beyond textual content

This research has focused on *textual* content as an information-rich resource within informative websites. Further studies could generate advice for non-textual content such as images, videos, and other multimedia though these are outside of the scope of this research. The analysis of only textual content provides a concentrated view of the types of advice that can be generated for practitioners, highlighting the possibilities for future research to extend into informative content presented via other contemporaneous means. Despite the evolution of web technologies, the guidelines generated in the IWCG will continue to be applicable even as specific devices and technologies evolve within a future online world.

13.3.6. Summary

The development of the model has transformed prior academic studies into information quality and the action case think aloud studies conducted by the researcher into a model for practitioner use. This model includes 16 separate snippets of advice for producing more informative online content, building on 11 prior information quality attributes, along with the addition of 3 new attributes.

The model is presented using the same approach as the industry-standard Web Content Accessibility Guidelines (WCAG) 3.0. This facilitates a common

approach, building on existing practitioner knowledge in the application of frameworks and guidelines to improve websites.

13.4. Validation study

The fourth action case study provides an opportunity to validate the practitioner model and application methods using another informative website. This is structured as per the previous case studies, with two cycles of evaluation both before and after application of the model.

13.4.1. Purpose of validation

Oates (2005, p. 132) defines two types of validity: internal and external. Of these, the latter is most relevant to the generalisability of research to provide predictions of future contexts. One method to provide a high level of external validity is repetition of studies in as many situations as possible, as demonstrated in this study by the use of multiple action case studies. In this research, the external validity is extended by the development of the fourth action case study within another context, providing an additional domain where the model is applied.

13.4.2. Selection of case study

Selecting the fourth case study posed a challenge for the researcher, as the purpose of validation is to demonstrate the research outcomes in another context without relying upon similar constructs that may bias further results. After conducting the first three action case studies, the researcher selected a Wikipedia page (user generated content) to provide additional insight into another domain of informative website.

As discussed in Chapter 12, the specific page and task scenario were selected to provide insight into an information-rich page with representative actions of what a typical user may be seeking during their interactions with the content. This increases the number of representative test cases, diversifying the data to be studied.

13.4.3. Application of model

A key element of the fourth case study is the application of the model to the test web page, demonstrating the increased successful outcomes of the participant think aloud studies after the changes compared to the native content. The researcher simulated the activities that would be conducted by a practitioner on their own content, following the guidance included within the Informative Web Content Guidelines (IWCG) and discussed in Section 11.5.

13.4.4. Challenges of user-generated content

User-generated content poses its own challenges for the action case study, as it can contain characteristics of the author's own writing, is often unedited and not officially provided by any authoritative organisation. Despite these challenges, the English Wikipedia resource provides information for an average of 837 million unique users per month (Wikimedia, 2022).

The definition of the task scenario facilitated the researcher in the selection of a mature, fully referenced, and curated page to generate questions for the participant activity. The questions and tasks were designed to simulate the types of questions a user may visit the page to answer, including finding specific information about nutrition.

As a validation case study, the user-generated content is suitable as it further demonstrated the applicability of the IWCG in another context. The case study website may not have been suitable for the development of the model due to its focus on one page that may not have been representative of content or typical user tasks across the entire site.

13.4.5. Impact of online study

Whilst the fourth action case study was conducted using the same methodology and think aloud approach as the earlier cycles, it is the only element of the research that was conducted online rather than in-person.

Comparative studies have demonstrated that both physical and virtual think aloud sessions can provide effective results (Brush et al., 2004; Tullis et al., 2002), though there are challenges in creating a comfortable environment for the participant (Fan et al., 2020, p. 95). Attempts to mitigate the effects were made by providing unstructured time at the beginning of the study and encouraging verbalisations, though the nature of interactions can vary when using remote tools. The researcher recognises that the nature and depth of research activities may have varied dependent on the means of the study, however the change to the use of online technologies is isolated to the fourth validation case study with a different purpose (providing external validity) to the first three case studies (providing insights and exploratory data).

13.4.6. Summary

The application of the IWCG to a fourth action case study has helped the researcher to demonstrate the validity of the model within both another site and another domain (user-generated content). This provides further evidence of the practitioner value of considering the quality of online content contained within their website.

The next section explores the key themes that emerged during the research, identifying both successes and challenges for future studies.

13.5. Themes

Interpretive research explores complex organisational contexts in-situ, identifying problems leading to theory construction. This is often apparent as action research gravitates towards key themes during the activities. Throughout this study, the researcher has observed several key themes across the case studies and model development, each of which is discussed below.

13.5.1. Definitions and nomenclature

A key challenge during this research has been the lack of clear definitions of terminology. Much of this has been the result of information systems being a multi-disciplinary field, drawing together research from multiple separate domains.

The first challenging term encountered was content (see Section 2.2). Many fields and authors within the Information Systems space have conflicting definitions which provided an obstacle for this study. The researcher defined content as 'any information that is expressly provided to the user to transmit meaning within the message' (p. 41), a loose definition that covers both textual and non-textual content. This encompasses the various academic and practitioner definitions, including an intuitive sense when applied to the online context, though may be too broad for the ongoing study of the model. Future research may be required to refine this term for more specific applications.

The second challenging term was information. Although the study follows Checkland's Soft Systems Methodology that provides one definition of this term, *information* can have different meanings based on the researcher or practitioner perspective. For those in industry, information is generally understood as a broad term for any content, whilst academics may have a narrower scope as *content with meaning* or *processed data*. This distinction became evident in the review of information quality frameworks (see Chapter 4), where attributes and dimensions of data and information quality have been considered interchangeably by prior works. With the Information Systems field, future research needs to carefully consider these terms and their application, identifying a strict interpretation of the research space to provide clarity for further works. The model developed during this study is specifically focused on information and could not be applied to raw data encountered on the web due to the lack of meaning within the message.

The final term without a distinct definition was informativeness (see p. 67). Building on the concept of information (with its own differences in interpretation), informativeness is not clearly defined within prior works. Recognising the importance of user perceptions, the relative understanding of informativeness may differ based on the interpretation of whether volume or quality of information is more important. As has been established through the think aloud studies, neither of these concepts can be considered within an isolated space as the construct of informativeness relies upon up to 14 different attributes.

13.5.2. Content separate to the system

The separation of content and the system is not often considered within HCI research; however, this is an essential distinction for the study of information quality, which only pertains to the content and not the system in which it is stored.

The DeLone-McLean model (see p. 78) provides the clearest distinction between the content and the system, with elements categorised into three areas: the system (HCI aspects of the interface), the information (content and meaning provided by the website), and the service (user experience considerations). Even with this model, it can be challenging to provide distinct classifications for the three areas, with each having an impact on user information seeking patterns and outcomes.

Existing practitioner frameworks such as WCAG consider system and service aspects without a focus on the information provided to the user. This can result in the improvement of the user's experience yet no increase in overall information systems success. The provision of the information quality framework produced by this study aims to address this gap by identifying and suggesting areas where content can be improved to have a positive effect on user outcomes.

Continuing to examine web pages as a whole, combining text and multimedia content with layout, navigation, and system artefacts does not help to disambiguate the specific attributes responsible for informativeness, leading to unintended consequences. For example, navigational issues within a page could be considered as either a content or system issue depending on the researcher's perspectives. Whilst this study has delineated between the two based on an intuitive sense of which navigational elements are directly a result of content authoring and which are a consequence of system/interface design, future research could expand on this area to identify the potential difficulties of not fully identifying the nature of website features which cross such boundaries.

Keeping content separate from the system features has been a continual challenge since the early days of the web: websites of the 1990s typically mixed data with presentation elements, using features such as inline styles. More contemporaneous web development practices suggest keeping data and presentation aspects separated though they cannot be considered wholly mutually exclusive as certain elements of experience design are required to present information within a user accessible manner.

13.5.3. Generic or domain-specific

Early information quality frameworks examined data and information from a generalised perspective, applying the same attribute sets and dimensions to determine an overall quality metric. As specific online fields such as e-government and e-health have grown, individual communities of researchers have further specialised these attributes to be suitable for their own purposes.

The overview of key frameworks in Chapter 4 identified several such examples where individual questions or criteria have been introduced to handle the existence of domain-specific content such as medical or contact information. Whilst the efficacy of such specialised frameworks has been

examined elsewhere, their common redefinition within disparate fields suggests both the utility and interest within quality research.

The research objectives of this study (particularly RO3) did not restrict this research to one such domain. This was a deliberate choice by the researcher as a focus on only one type of website would both constrain the possible outcomes and create an instrument that could only be used in a confined set of circumstances. The model produced and defined during this research should be generalisable across all online informative content, as demonstrated using four action case studies across different domains.

The production of a generalised information quality model leads to the question of whether this can or should be refined for an individual domain. In a similar manner to the quantitative frameworks discussed earlier in this thesis, it would be possible to distil the questions to more focused questions depending on the category of websites to be studied, however this may preclude advancing research into the model across further examples. This is often the case with domain-specific IQ frameworks that become too specialised, leading to their study and revision by a small group of researchers without reference to evolving trends of the web. A generalised model such as the IWCG would be most useful to practitioners in the same virtue as the WCAG model is universally applicable and not spawned a number of specialisations to various types of website.

13.5.4. Non-informative content

At the commencement of this study, the research scope focused on informative content. This was a deliberate decision to study content with a single common purpose: to enable the user to know more about a particular topic after interacting with the site. This is dependent on user perceptions of the content, with a close link to the quality of information provided by an organisation.

In addition to information, there are many other forms of content provided online, including for entertainment, marketing, or social purposes. The incorporation of these into the study would have necessitated further understanding of user intentions and perceptions, including further development of task scenarios with other outcomes. Conducting task scenarios for non-informative content could be challenging, as websites developed for e.g., entertainment may not enable every user to achieve the same purpose. Future studies could expand the model for non-informative websites; however, this could require the introduction of several new attributes and dimensions to cover the other aspects of user experience, detracting from the aim of providing clear advice to content authors.

13.5.5. Alternative content types

As discussed in the action case studies, the main object of research was textual content. Content consisting of solely text is easiest to separate from other systems aspects, as the text is the same across all platforms and however the site is accessed, stripping several of the interactive elements that may distract a user. Despite this clarity, there is an intrinsic connection between the content and how it is displayed to the user, hence some navigational and presentational aspects need to be considered due to their impact on the overall informativeness of content.

This study and its task scenarios have avoided the use of interactive or multimedia content, search engines, social media, and other navigational aids in the information seeking journey. Whilst this has provided a useful abstraction and simplification of scope for the study, further research could consider the interaction of these various elements as useful points in a typical user information session. In addition, the interaction of the information within a page with other content types such as appropriate imagery and device-based navigational aids could introduce additional considerations into the model for content authors.

13.5.6. Tendency for metrics-based analysis

A key theme throughout the overview of information quality frameworks in Chapter 4 was the use of quantitative metrics for analysing informativeness of content. Metrics introduce a common standard of measurement, facilitating comparison between different cases or examples. This can be useful for those performing comparative studies, however this research focused on a practitioner model for improving content within their context, which may not involve comparison to any other content.

The challenge of applying metrics to any web page is that they can only produce results in terms of conformance to criteria rather than an understand of how users perceive the content. In an idealised model, all users would behave the same and the two would have an identical outcome, however perceptions can differ between sessions and information seeking goals. The consideration of interpretivistic data such as the utterings and observations from the think aloud study encouraged further insight beyond what could be obtained by a mechanised processing of web page features.

Following the initial three action case studies, the IWCG model was developed to facilitate the analysis of website content informativeness. The model follows the WCAG 3.0 pattern, which introduces the idea of ratings for each guideline. Whilst this appears to contravene the above paragraph, the ratings serve a useful purpose for practitioners: by analysing their own site according to the examples and criteria, any areas which score less than the maximum may indicate criteria where content improvement works could be conducted to improve the rating. This reduces metrics from a comparison between cases approach to comparing evolutions of the same content between iterations of development.

13.5.7. Link between academic and practitioner models

This research draws together two disparate fields of study: academic models such as information success and quality frameworks, and models such as WCAG for use by practitioners. Despite the longevity of both fields, the lack of common research and literature highlights the challenges of academic study within industry as well as the lack of application of academic research within practitioner contexts.

RO3 of this study defined a key outcome as the creation of a practitioner model: this was considered as a useful output and contribution of the research as studies into web topics rarely make the transition between academic and industry spaces. By developing an instrument that can be applied by practitioners, the long tradition of information quality studies can be utilised within a new context providing benefits for end users as well as content authors.

13.5.8. Lack of academic research into industry best-practice

Despite the renewed focus on industry best practice within this thesis, academic research often lacks a connection to the actions and advances of practitioners. An example of this disconnect is the development of the upcoming WCAG 3.0 standard for accessibility. This has been worked on by an international committee of practitioners for several years, yet few academics have acknowledged the existence of the upcoming changes to standard WCAG assessments. This is consistent with other areas of information systems research where academic outputs can often lag behind technological and digital developments.

Future researchers within the web space could continue with the perspectives of this research in applying academic models, theories, and methodologies to industry practices, helping to improve access to information for all.

13.5.9. Technology changes

Whilst the action case studies discussed earlier in this thesis examined user perceptions of content with a short timescale, the wider changes and longevity of web technologies also need to be considered. The concept of a website has evolved a long way since the early text-only pages accessed over slow speed connections.

The ubiquity of web access and high levels of digital literacy have changed the way an ordinary user interacts with the web, leading to a higher demand for online content. Along with this increased demand, expectations of organisations' web presence have also increased, with users expecting to be able to conduct many activities online through self-service, easy-to-use platforms. This highlights the importance of accessibility and informativeness within content authoring, as this may be users only interaction with a specific organisation.

All of the think aloud activities were conducted on standard PCs, though this may not represent the typical web interaction: for instance, 26% of UK adults only use a device other than a traditional computer for their online browsing (Ofcom, 2021, p. 17). This highlights the need for future research to consider content displayed on other devices such as smartphones, tablets, and media systems. Even though the same content may be accessed on any device, the nuances of the user journey and restrictions of navigation should be considered for any potential impact on the overall informativeness of the site.

13.5.10. Summary

Whilst this research study has met the stated research aims and outcomes, it has highlighted several areas that could be considered either for future studies or further development in the information systems field.

Ensuring nomenclature is consistent and concisely defined is crucial for the accurate development of models, especially those that bridge academic

and industry practices. This could lead to the study of other types of content outside the scope of established definitions, broadening understanding of user expectations and interactions, especially beyond the classic PC interface.

The following section discusses possible avenues for further study as an outcome of this research.

13.6. Towards a future research agenda

As with all long-term research projects, several possibilities for future avenues of studies were identified during the activities. These fall into three main categories as discussed below.

13.6.1. Extensions to model

The first category of future research is the possibility of extending the practitioner model. The Informative Web Content Guidelines (IWCG) generated through this research give content authors 16 pieces of advice structured into four guidelines, a shorter model than the comparable WCAG. Future research studies could conduct the same process, applying an iterative action case study cycle to generate additional advice based on participant utterances and observations. This may produce additional attributes and further specification of advice to further guide practitioners in the production of informative content.

This study has focused on textual content within informative websites, a deliberate scoping decision of the researcher. On the modern web, users are faced with many different media such as images, audio, video, and interactive elements that also convey meaning. Information can often be presented in alternative formats, structured into diagrams to produce clarity of processes and procedures (reminiscent of the adage *a picture is worth a thousand words*), enhancing the density of meaning. Extending the model to other types of media would require further study of information retrieval by

users, introduction of new guidelines and criteria for specific considerations. As media can often be viewed in alternative formats, the content author would need to be reminded to review all possible interpretations of the data.

An additional extension to the model could be provided by additional analysis of the interaction between the content and the system. Despite the distinctions provided by DeLone-McLean's Model (see Section 3.2.4), the content and system elements cannot always be mutually delineated without hidden dependencies. This is evidenced by the existence of guidelines targeting navigation, styling, and layout concerns that are embedded within not only information but also the system itself. A future study could enhance the body of knowledge of the understanding between these two elements by examining the relationship and questions of what is content and what is the system?

13.6.2. Extensions to domains

An alternative approach for future research would be to continue the action case studies within alternative domains. The initial three case studies focused on informative content within a university, a non-profit organisation, and a health website, three examples of traditionally informative websites. This was augmented with the final validation case study based on user generated (albeit highly curated) content.

Whilst the case studies were selected by the researcher to be representative of the textual information seeking exercises a typical web user may encounter, there is the scope to increase the research into multiple other domains such as corporate knowledge bases, e-learning resources, portals, and current affairs. This would achieve two research outcomes:

1. Further demonstrate the applicability of the research model and IWCG within additional domains.

2. Provide the opportunity to supplement the model with additional information quality attributes and practical advice.

Both outcomes provide additional benefits to practitioners, enabling the development of the model into a single resource for improving informative websites.

An additional possible avenue would be the specialisation of the model for non-informative websites. Whilst the focus of this research has been on those spaces where a user is performing an information seeking activity with answers provided by the textual content, other types of site such as applications, entertainment resources and e-commerce could be considered as extensions to the existing model. Further derivation of applicable attributes would be required to ensure compatibility with the existing model defined by this research and the general aim to improve informativeness rather than conversions or enjoyment.

13.6.3. Alternative approaches

This research study has been deliberately undertaken within the interpretive paradigm based on the philosophy of the researcher and the nature of the problem. Data collection using qualitative methods has focused the model and outcomes on participant utterances and observations, identifying their cognitive patterns as representative of typical users.

A key element of the literature review was the identification of most prior works as positivistic in nature, providing an opportunity for this research to develop an alternative approach to analysing content quality in line with industry user research activities.

Whilst the researcher recognises the importance of qualitative data within the analysis of content quality, the IWCG provides many opportunities for the practitioner to translate their site's current performance into a scaled, quantitative value. The model's counterpart, WCAG 3.0, aims to automate some of the more mechanical elements of website analysis, and the same approach could be applied to this model. For example, it would be possible to create tools to enhance the analysis of availability and information architecture (Guidelines 1 and 2) providing an augmented measure along with the practitioner's interpretation of their own site. Activities to develop such measures could help with the application of the model, increasing its potential use within industry.

13.6.4. Conclusion

Overall, there are several different avenues for future research following the IWCG model produced by this study. These include increasing the depth of the model with further information quality attributes, increasing the breadth of applicability with further domains of study, and aiding practitioners in its use by providing automations where possible. Future research themes are further discussed in Section 14.9.

13.7. Summary

This chapter has reviewed the key themes emerging during this study, reviewing the potential areas for future development of the model based on the use of prior information quality attributes and those derived during the action case/think aloud studies. Any future developments should be produced with the same intent as this original research, focusing on the ways in which practitioners can access advice to improve their own websites.

The next and final chapter of this thesis synthesises the discussion above with the original research questions, aims and objectives to demonstrate the successful outcome of the study and identify potential future directions of the research.

Chapter 14. Conclusions

This chapter synthesises the discussion documented in Chapter 12 by reviewing each of the research questions, aims, objectives and contributions to analyse how these have been met by the research study. Subsequently the methodological approach selected by the researcher is reviewed along with the study limitations and considerations for a future research agenda.

14.1. Introduction

This research study explores how web content quality can be analysed and improved via information quality attributes. To understand how the original research problem has been studied, this chapter reviews each of the key research questions, aims, objectives and contributions defined within Chapter 1, and discusses how these have been addressed.

An important consideration for any research study is the methodological approach, reviewed and discussed within Section 14.5. Following this, the limitations of the study are explored, leading to a reflection by the researcher and avenues for future study based on the discussion in Section 13.6.

14.2. Research questions

At the beginning of this thesis on p. 13, the researcher defined three key questions that identified the problem to be solved by this study and refined the scope for subsequent research activities. Each of these questions are discussed below.

14.2.1. How can the information quality of online content be evaluated? (RQ1)

The first research question framed the challenge of evaluating the information quality of online content. To answer this question, several elements needed to be synthesised by the researcher.

A key challenge of this research study has been the definition of appropriate nomenclature for online content and information quality. Chapters 2 and 3 explored the various definitions of *content*, *informativeness*, and *information quality* used throughout academic works and industry approaches.

During the discourse in Section 2.2, the researcher defines online content to be any information provided to the user with an intended meaning within the message. This draws together the practitioner stance that content is providing messages in various formats, whilst academic literature (evolved from Shannon's *theory of communication*, 1948) promotes the meaning provided by information. Information itself is composed of multiple pieces of raw data, each of which do not convey meaning on their own yet can be combined into a collection of information (c.f., Checkland, 1981).

The evaluation of information quality (IQ) can be provided by one of a myriad of IQ frameworks. Many of these frameworks are based on the works of Wang & Strong (1996), however contextualisation to informativeness of online content has not yet been provided by the academic community. This provides the research gap where this thesis intended to explore possible evaluation techniques.

Combining academic and practitioner evaluation along with the interpretive philosophical paradigm outlined in this thesis has resulted in the definition of a new approach: the Informative Web Content Guidelines (IWCG) in the same format as the practitioner-led Web Content Accessibility Guidelines (WCAG). This provides an easy-to-use framework in a format already known within industry to analyse and improve informativeness of web content.

14.2.2. How are content quality and user achievement of online information-based objectives related? (RQ2)

To demonstrate the effectiveness of the Informative Web Content Guidelines (IWCG) as a model to improve web content, it has been necessary to first understand the link between content quality and achievement of online objectives.

The application of task scenarios within the action case cycles (see Section 6.6) has simulated user achievement of information-based objectives by providing participants with a problem context and individual goals. Through the observation of these activities using the think aloud protocol (see Section 6.7), the researcher identified key aspects of informativeness of the case study websites which could be improved, developed a model to frame these as a set of guidelines, and demonstrated the positive link between the process and task outcomes in the validation case study.

14.2.3. How can practitioners apply the key attributes to improve informativeness of online content? (RQ3)

Academic and practitioner approaches to information systems analysis can often differ, resulting in a mismatch in nomenclature and benefits to be provided by existing contributions of knowledge. The final research question directs the study towards practitioner outcomes, with a focus on the provision of a model to disseminate the findings beyond the information systems field.

To provide a practitioner-accessible model, the Informative Web Content Guidelines (IWCG) parallel developments within the Web Content Accessibility Guidelines (WCAG), a widely known and followed framework for accessibility. This has resulted in the development of an evidence-based approach for evaluating informativeness of content alongside existing practice aimed to make the web more inclusive for all.

Further guidance on the application of the model is discussed in Chapter 11.

14.2.4. Summary

The three questions defined at the beginning of this study have been answered by this research. By providing a synthesis between existing academic literature and industry practices, the outputs of the study are applicable within both the field of information systems (for further development towards informative content attributes) and web development activities (for improving the informativeness of existing web content).

The next section discusses how the study meets the original research aims.

14.3. Aims

At the beginning of this study, the researcher defined two aims. Each of these are analysed below to demonstrate how they have been achieved by the research outcomes.

14.3.1. To identify which information quality attributes influence the informativeness of online content. (RA1)

The first aim targeted the research activities towards the identification of both existing and new attributes of information quality which may have an impact on the informativeness of online content. This was an important activity, as the attributes that have an impact on other types of content and in other domains may not be appropriate on the web or may fail to influence the user's perception of the informativeness of the system.

In the preliminary review of existing information quality attributes from prior academic studies, 11 attributes were identified by the researcher as potential candidates for exploration in the action case studies. Furthermore, the initial three action case studies provided evidence for 11 of these within informative online content, with a further three attributes (fallback, information usability, and interactivity – see Figure 11.3) identified as being important to web-based information yet not included within the academic models.

The validation case study documented in Chapter 12 further demonstrated the relevance of the fourteen identified attributes to informative online content, providing confirmation that application of the IWCG can help practitioners to improve their web content.

14.3.2. To provide a practitioner model for online information quality based on academic literature and user studies. (RA2)

Following the literature study and first three action case studies, the researcher defined the Informative Web Content Guidelines (IWCG) as a practitioner-focused model for the study of online content with respect to the fourteen attributes identified within RA1. When formulating the practitioner model, the researcher identified that the industry led WCAG provided an accessible format for advice to content developers in a familiar format, enhancing the potential for communicating the model and encouraging its use. The WCAG has been developed by a community of web professionals over the past decade to be both universally applicable and easy-to-access for those within the industry. Furthermore, academic studies of the WCAG have demonstrated the model to be mostly effective at providing generalised advice for web-based systems.

The IWCG as intended for practitioners is included in Appendix A of this thesis.

14.3.3. Summary

The two research aims of this study provided a framework for the study of information quality attributes of online content along with the intention for the model to be accessible to practitioners as well as academic researchers. Both aims have been met with the synthesised study of attributes within both academic literature and the first three action case studies, followed by the development of a model following industry best practice and the established WCAG. Additional opportunities for future development of the model and dissemination to practitioners are discussed in Section 14.9 below.

14.4. Objectives

The intention of research objectives are to provide a 'clear sense of purpose and direction' (Saunders et al., 2003, p. 25) for the study with more detail than research questions and aims, with a specific focus on how the outcomes can be *measured*.

At the beginning of this thesis, the researcher defined three objectives to be studied over the course of the doctoral research, each of which is discussed below.

14.4.1. To identify information quality attributes from academic literature inherent in online content. (RO1)

The first research objective relates to the study of information quality attributes within existing academic literature. Much of the literature review in Chapter 4 focused on the identification of relevant attributes within frameworks in academic literature. Through analysis of existing frameworks, eleven key information quality attributes were identified for further study (see Section 7.2).

Selecting the existing attributes relevant to information quality provided the first step towards RO2, the identification of further attributes inherent in online content yet not identified within prior academic frameworks.

14.4.2. To determine the key attributes that impact on the success of online content. (RO2)

Following the study discussed in RO1 above, the researcher identified the attributes that impact on the informativeness of online content. This element of the study was conducted using the three action case studies discussed in Chapters 8, 9 and 10. Each action case study focused on the identification of information quality attributes within a different domain following the cognitive patterns of ten representative users conducting a task scenario, with thought processes identified through the think aloud protocol. This determined the

final set of 14 attributes (11 from the literature and 3 additional) to be included within the model.

Following the identification of the 14 information quality attributes, the fourth action case study provides validation through application of the model to another domain. Whilst there are limitations to the completeness of this approach, it enabled the reanalysis of the model within an additional context, demonstrating the value that can be provided by the practitioner model.

14.4.3. To create and validate a practitioner model to describe the relationship between key information quality attributes and online content. (RO3)

The final research objective centred on the development of a practitioner model to aid in the identification of the relationship between information quality and online content. This is provided as the Informative Web Content Guidelines (IWCG; see Appendix A) in a format matching the existing industry standard WCAG model. The model's approach has been validated using the fourth action case study documented in Chapter 12.

14.4.4. Summary

The research conducted to meet the objectives identified at the beginning of the study has enabled the researcher to identify both the key information quality attributes that can improve online content and a suitable model for communicating this relationship to practitioners for application to their own website content.

14.5. Methodology

A key element of any research study is the methodology chosen as this can guide the outcomes of the activity. This section will review the methodological selections and their appropriateness to the findings and further directions of the research.

14.5.1. Interpretive approach

The philosophical discourse in Chapter 5 led to the selection of an *interpretive* paradigm for this research study. This enabled the researcher to explore underlying meanings embedded within the relationships between different themes and contexts in a specific problem space. Interpretive approaches to research within the Information Systems field are increasing as a response to the multi-disciplinary nature of the domain and the lack of concrete hypotheses to be tested with quantitative measures (as would be the case in positivist studies).

One of the key challenges of an interpretive approach is the interpretation of data. Rather than the application of well-defined methods, authors such as Klein & Myers (1999) suggest guiding principles for the IS researcher analysing qualitative data. The overall direction is towards balance rather than rigid rules to provide a context-aware perspective of the challenges.

The researcher has addressed this challenge by applying iterative methods such as action case studies and cycles of data analysis. This has aided in the validation of key data, leading to the practitioner model. The use of a validation case study in Chapter 12 has further demonstrated the applicability of the approach and identification of key attributes.

The following sections evaluate the research within the context of Klein & Myers' (1999) seven principles for the conduct of interpretive Information Systems studies.

14.5.1.1. Principle 1. The Fundamental Principle of the Hermeneutic Circle

The first principle develops the notion that to understand an area of study the researcher needs to consider both the individual concepts and the whole picture once these are combined. This is encapsulated in the Soft Systems Methodology by the consideration of the problem domain and its constituent components, identifying the role of each actor plus the system in its entirety.

This research has been conducted using these principles: each of the attribute selection study and action case studies identify the applicability of information quality attributes within a particular context, with the model development outlined in Chapter 11 combining the individual findings into a set of guidelines for generalised applicability across informative websites.

14.5.1.2. Principle 2. The Principle of Contextualization

The context of research data is an important consideration as the situation in which it is collected may have an impact on the phenomena observed. For the action case studies, the researcher analysed the individual scenarios and developed task scenarios to account for the typical audience providing specialisations of the research questions for the different domains. This contextualisation then fed through to the generalised model which aimed to reduce the influence of isolated contexts from the case studies and identify the key repeatable attributes and guidance that can be applied to other informative websites.

14.5.1.3. Principle 3. The Principle of Interaction Between the Researchers and the Subjects

One of the key challenges of interpretive IS studies is the interaction between researcher and participants. The major element of this is the social construct under which the context is studied. For this research study, the researcher minimised the potential conflict between research objectives and participant objectivity by conducting a series of concurrent think aloud observations incorporating open discussions of task scenario completion. To mitigate the potential Hawthorne Effect (where individuals modify their behaviour in response to knowledge of the observation, each cycle of the study was conducted with ten participants, far exceeding Nielsen's (1994) conclusion that five participants would be sufficienct to observe 80% of usability problems (see p. 185).

14.5.1.4. Principle 4. The Principle of Abstraction and Generalization

The philosophical and methodological considerations of this research study have been discussed in Chapters 5 and 6, where the researcher has stated and justified their selections within interpretive IS research. Principle 4 extends beyond these developments, suggesting that observations and results should be generalisable within the context of the research. The IWCG model produced during this research is applicable to a wide variety of informative contexts whilst being grounded in the core Information Systems theories of the DeLone-McLean model and prior Information Quality research.

14.5.1.5. Principle 5. The Principle of Dialogical Reasoning

Interpretive studies can often uncover unexpected research outcomes. Within this research study, the identification of three new information quality attributes with an impact on online content success (fallback, information usability, and interactivity) is an example of such outcomes. The use of an iterative methodology with several cycles of research mitigates the risk that these attributes are only apparent in one particular context, demonstrating their continued applicability across the case studies (see Figure 11.3 on p. 294).

14.5.1.6. Principle 6. The Principle of Multiple Interpretations

As with all research studies involving qualitative participant responses, multiple interpretations of the same narratives could lead to different results. To mitigate this impact, the researcher conducted multiple cycles of research with multiple participants representing a wide range of audiences without relying on a single demographic of user. This more accurately represents the typical user base of an informative website, bringing a variety of prior experiences to the study. In addition, the data has been evaluated using cyclical processes to objectively introduce new observations and utterances rather than relying on the input from prior cycles of study.

14.5.1.7. Principle 7. The Principle of Suspicion

Klein & Myers' (1999) final principle concerns the introduction of false preconceptions through data interpretation distortions. To mitigate such misunderstandings, the researcher applied a critical perspective to the observations and utterances from participants to identify those data items related to the information quality properties of the content rather than the system. By performing three initial action case cycles, the researcher reduced the reliance on input from any particular aspect of the case studies that may have caused social distortions within participant data.

14.5.1.8. Summary

The research documented in this thesis applied and followed the seven principles for interpretive IS research as described above. The researcher minimised the effects of unintended biases, contextualisations, and interactions through the considered selection of appropriate case studies and evaluation of data.

14.5.2. Soft Systems Methodology

Within the interpretive paradigm, the Soft Systems Methodology (SSM) provides a model to link three key elements: a framework of ideas, a methodology and an area of concern and is particularly appropriate for IS studies focusing on *softer* challenges.

SSM provided a suitable methodology for this study as it enabled the exploration of loose themes connected by research questions generated through the researcher's own experiences. The challenges of providing informative content to a varied audience across several domains required the study of a complex problem, with the questions and objectives evolving with the input of each action case study cycle's outcomes.

14.5.3. Action case research

Action case research is an emerging methodology within Information Systems, identified to fill the gap between hard case studies and iterative action research. The main difference between action case and action research is that in the former, the researcher does not play a role in the activity being studied whilst they would participate in the latter.

An action case approach was a suitable selection for this study as the researcher played the role of an observer throughout the early three case studies (see Chapters 8–10), whilst their role was limited to changing content between rather than within action cycles during the validation case (see Chapter 12). This limited the impact the researcher may have caused by their own involvement in simulated activities, moving towards observation rather than active participation.

14.5.4. Task scenarios

Within the fields of usability, Human-Computer Research (HCI), and information retrieval studies, task scenarios are commonly used as they are well suited to raising questions about an activity, its problems, and its inherent relationships (Carroll, 1999, pp. 5–6; J. Kim, 2012, p. 300; Rosson & Carroll, 2002, p. 2). Task scenarios provided a suitable approach to this study as they provided the opportunity to observe users working through simulated information seeking activities, capturing their visual and verbal responses in a realistic and repeatable manner.

One key aspect of task scenarios is the repeatability of the evaluation, with some researcher suggesting anywhere from 5% to 65% agreement between participants (Hertzum & Jacobsen, 2003). This is a double-edged sword as agreement can provide both consistency of outcomes with fixed scenarios facilitating comparisons and transparency (Hoppmann, 2009, p. 213); and obscurity for additional considerations within the research. The latter is often less of a concern in usability studies, where it has been observed that

between five and eight participants can identify a majority of issues (Ericsson & Simon, 1984; Lindgaard & Chattratichart, 2007, p. 1422; Nielsen, 1994). Thus, conducting each action case cycle with ten participants provided sufficient coverage to identify facets of the information quality with a high level of certainty. It is possible that the appearance of some facets may have been missed due to the size of the cohort, however this has been further mitigated by conducting four separate cycles of action case research (three for the initial study, and a final two cycles for the validation case study).

14.5.5. Think aloud protocol

The think aloud protocol is not a single technique; rather a group of usability evaluation techniques that can be applied to explore participant cognitive patterns during an information seeking activity.

A key determinant in the successful application of the think aloud protocol is the relationship between the participant and researcher (c.f., the user and the evaluator in Clemmensen et al., 2009, p. 216). This was considered by the researcher in the construct of the studies, with the development of familiarity and rapport in the early stages (through welcome discussion and initial unstructured task) combined with using an environment familiar to the participant (whether lab-based or online in later cycles).

One element that helped develop the relationship between participants and researchers was the use of the concurrent think aloud technique. As the researcher continually engaged with participants through questions and other verbalisations, the users were encouraged to discuss their cognitive thoughts and perceptions of the case studies rather than continuing alone. This moved the activity away from an observer–observee relationship towards a friendlier, more co-operative approach that helped uncover deeper insights.

14.5.6. Data analysis

As with all interpretive studies, data analysis required subjective and thematic evaluation to synthesise attribute findings. To help with the data analysis, a cyclical process was used to identify key attributes and then refine the individual attribute sets. Using such an approach is common within industry contexts, with the Plan-Do-Check-Act cycle defined within ISO 9001 (2015), consistent with the cyclical nature of action research.

The use of a cyclical approach enabled the flow of outcomes from the first action case study into later activities, and this helped to establish and further develop the attributes identified within multiple domains. Whilst this has led to consistent and successful outcomes, the continuation of any data across multiple case studies introduces the opportunity for researcher biases or observations to persist in later data. This was avoided by reviewing all attributes during the model development stage and selecting those most likely to be impacted once outcomes from three action case studies had been produced.

Further action cycles could have introduced additional data into this study, however this may not have been sufficient to change the overall model. For each of the 14 attributes, there is evidence for both their occurrence and impact within online content that can be independently justified. One avenue for future studies would be to continue the action case study cycles within additional domains and with alternative websites to further demonstrate the attributes with the largest impact on information quality of online content.

14.5.7. Summary

The methodological approaches selected at the beginning of this research study were chosen for their suitability to the research questions, aims and objectives in addition to the perspectives of the researcher.

The action case cycle approach enabled the researcher to combine think aloud studies with existing academic literature on information quality attributes, producing the practitioner model and subsequent validation outcomes. This has resulted in a model accessible to those in industry to improve their own web content based on examples within four different domains, providing a generalised approach to online informative content.

Considerations of additional methodological approaches in further studies are discussed in Section 14.9.2.

14.6. Contributions

In Section 1.5 at the beginning of this thesis, the researcher outlined the intended contributions of this study in three areas: knowledge, theory, and practice. The contributions produced by the research in each area are summarised below.

14.6.1. Contribution to theory

As identified within Chapter 1, there is one contribution to theory generated by this thesis, relating primarily to Research Objectives RO1 and RO2. This contribution is evaluated below.

C1. Identification of existing information quality attribute set relevant to online content

The first planned contribution was the analysis of existing information quality attributes, particularly those relevant for online content. Since the seminal work of Wang & Strong (1996), the academic study of data and information quality attributes has increased. This has produced numerous sets of non-mutually exclusive attributes for consideration by researchers. A key challenge for this research was the distillation of IQ frameworks and attributes into a key set for consideration as likely candidates for the study of online content. This contribution is documented within Chapter 7, where the

researcher mapped information quality attributes and justified a selected subset for consideration.

The study progressed by considering the identified attribute set to identify which are relevant to the study of online content. This reduced the set analysed earlier in the research to the final set of attributes for inclusion in the action case study elements as documented in Section 7.3, with those excluded from study justified in Section 7.3.1.

An additional element within this contribution was the application of the application of Information Systems (IS) success theory to online content. Theoretical models such as the one defined by DeLone-McLean (see p. 78) had not been previously applied to website content quality. Considering the problem context within the domains and relationships identified by this model produced a new contribution to theory that could be further developed for the other domains of HCI (System quality) and UX (Service quality).

14.6.2. Contribution to knowledge

This study has produced two key contributions to knowledge as documented below.

C2. Identification of new information quality attributes for online content

A key contribution to knowledge produced by this research is the identification of three new attributes pertaining to information quality of online content: fallback, information usability, and interactivity. These introduce new concepts to the existing attribute set, identifying additional considerations for the evaluation of website content for informative purposes.

C3. Development of enhanced attribute set relevant to online content

The first contribution to knowledge is the development of an enhanced information quality (IQ) attribute set relevant to the study of online content. Various IQ frameworks have been previously developed and studied for

websites, however none of these have focused on the quality of the content within the system. This is a unique perspective that has isolated and justified the attributes within online content across four separate domains. A key addition to this contribution is the new attributes identified within the action case study elements of this research documented as C5 below.

14.6.3. Contribution to practice

As a practitioner-focused research project, this study has emphasised the three contributions to practice documented below.

C4. Methods for analysing online information quality

This project has required the definition of new methods for analysing online content using IQ frameworks. Despite the large body of academic theory generated over the past 25 years, applying information quality attributes to content with an interpretive approach has not been explored often. Most IQ research is positivistic in nature, ignoring the benefits of qualitative studies into user perceptions of online content. Introducing industry-led methods such as the think aloud protocol and action case studies is novel within the study of online content and is a methodology that could be applied to future studies to continue delivering the benefits of understanding user interaction with websites first explored in this thesis.

C5. Practitioner model for application to website content

A key research output of this thesis is the practitioner model to introduce the academic information quality attributes to content authors as a mechanism for improving their content. The attributes identified during the action case studies have been distilled into the Informative Web Content Guidelines (IWCG), a model following the popular WCAG format. This provides an accessible and easy-to-use approach to self-evaluation of content, producing recommendations for how online content can be improved.

C6. Guidance for practitioners to improve online content quality

An important element of the Informative Web Content Guidelines (IWCG) is the practical guidance provided to practitioners. As part of the model development process (see Chapter 11), the 11 core and 3 additional attributes are transformed into 16 outcomes within 4 broad guidelines, providing an accessible approach to exploring the outcomes of the study. Stepping away from the individual attributes and transforming their intent within academic definitions into web-based considerations provides practical methods and categorisations for the improvement of online content. This is an important element of the WCAG model, which provides not only a measure of webpage accessibility but also suggestions for improvements that can be made to improve the score and overall experience of the page for users.

14.6.4. Summary

The six contributions by this research defined within Chapter 1 and evaluated above have been met by the study. The key contributions to theory have included the reinterpretation of existing academic theories to apply them to an interpretivistic approach to studying online content quality. The key contributions to knowledge have focused on the development and identification of the new attribute set for online content quality, distilling the myriad of information quality frameworks into a single device for application to website content. As a practitioner-focused research project, the contributions to practice have centred around the application of research outcomes by content authors, furthering the aims of this study to produce advice for those outside the Information Systems field to apply improvements to their own websites.

14.7. Limitations

As with any study, the research documented in this thesis was limited by several factors, each of which could have been mitigated by compromising the design of other elements of the study.

This section discusses each of the limitations and their impact on the research outcomes.

14.7.1. Limited domains

The first limitation within this study was the focus on four domains: a university website, a non-profit website, an e-Health website, and a user-generated encyclopædia. These four case studies were selected by the researcher for the reasons documented earlier in this thesis, which included prior bodies of academic literature in specific fields, accessibility of organisations and participants, and sufficient depth of content for task scenarios. Whilst they are intended to be representative of the types of information seeking challenges users may face online, it is not possible to select a single 'one-size-fits-all' site or domain that covers all informative possibilities. The model developed and documented in Chapter 11 is generalised to work across any informative website, though validation within every domain and every combination of features is beyond the scope of this study.

Future research could provide validation with other domains as discussed in Section 14.9.1 below.

14.7.2. Coverage of attributes

When designing action case study cycles within a simulated environment, a key challenge is the ability to create task scenarios that cover all the required areas of study without limiting or guiding the participant.

The task scenarios for the first three action case studies were authored from the perspective that they should represent a typical user journey through the

site, based on the researcher's notion of the average use case. Whilst this forms a useful comparative base for the study and areas of content likely to be interacted with by the user, there is the possibility that participants were guided to certain attributes or issues by the task scenario design.

The researcher mitigated against this possibility by conducting three separate case studies across different domains with different participants, reducing the likelihood of consistent repetition across the think aloud outcomes. This could have been further developed by designing multiple task scenarios and including additional cycles of research, though would have increased the scope of the overall study.

The final validation action case study (see Chapter 12) included additional potential for a reduced attribute set: focusing on a single page of user generated content diminished the potential for a content author to make adjustments in the areas of information usability, interactivity, navigation, presentation, and search as these attributes are all controlled by the constraints of the system. The Wikipedia article was still a suitable choice for the study as it demonstrates the overall effectiveness of the IWCG model for another type of informative website.

14.7.3. Qualitative study

Conducting a qualitative study of information quality is an unusual approach within the information quality space. Most of the prior literature follows quantitative methods, developing metrics for assessment and comparison of content.

This research has followed an interpretive, action case study approach introducing new concepts beyond existing models, including in-situ studies and analysis of subjective participant utterings and observations through the think aloud protocol. Although this is a limitation in terms of the scale of websites and cases that could be evaluated, the unique approach providing

insight into the meaning behind online information has led to a unique contribution to the field.

14.7.4. Types of content

A key limitation of this research is the focus on text-based content. In the early days of the web, this would have been the normal format for interaction with online information. Modern developments on the web such as interactive, media-rich pages, search-as-default behaviours and enhanced navigational structures have led users to become accustomed to accessing information through richer interfaces.

Limiting the content studied to purely the text (and inherent navigational and presentational elements) was necessary to provide an appropriate scope for the information quality model, masking the impact of the non-content attributes covered by conventional HCI and UX research. The narrow scope to produce a practitioner model for solely the textual aspects follows the refined nature of existing models such as WCAG that aim to improve a single aspect of a website.

14.7.5. Summary

Any properly scoped research study includes limitations because of the selection of philosophical approach and methodology. This study was designed to mitigate several of these limitations by demonstrating consistent outcomes across several case studies and four different website domains. To produce specific guidance for practitioners that could be validated with an additional action case study, the research focused on text-based content and a qualitative approach; both factors could be further studied with the future research agenda outlined in Section 14.9.

14.8. Role of the researcher

The role of the researcher is important to any doctoral study; however, this is even more crucial when applying an interpretive approach. Walsham

(1995, p. 376) highlights the importance of the researcher in interpretive studies as they are required to interact with the subjects to explore the problem space outside of controlled laboratory environments and applying their own interpretation of the data.

Throughout this study, the researcher has followed an interpretive approach incorporating their own experiences of online user experience to develop guidance and advice for application to a range of informative websites. Based on their prior industry experience, the researcher has focused on the application of academic models to practitioner resources to improve industry accessibility to key information systems theories.

With qualitative studies and evaluation methods, there is an ever-present danger that the researcher introduces their own biases and opinions, directing research outcomes to their preconceived notions of the topics. This has been mitigated by using the Plan-Do-Check-Act evaluative approach and reducing participant guidance throughout the think aloud studies to observe users solving their own problems. By taking a hands-off approach during these elements of the study and asking how and why questions, the unintended consequences of predisposition have been minimised.

However, the use of action methodologies introduces the researcher as a key actor in the research journey: with true action research as an observer and participant. By following an action case study approach, the researcher removed themselves from the context being studied, maintaining their role as solely observer.

14.9. Future research

As with any research study, there are several avenues through which the Informative Web Content Guidelines (IWCG) could be further developed and validated. This section summarises some of these approaches, highlighting

areas where a future research agenda could strengthen the research outcomes and expand the model into contiguous opportunities.

14.9.1. Validation of further domains

The first area which could be explored with further research activity is the validation of additional website domains in alternate fields. As discussed as a limitation in Section 14.7.1 above, this study focused on four types of informative website. This is a non-exhaustive selection of possible domains and could be expanded with additional case studies in other fields such as financial, educational, and current affairs websites. Such studies could be conducted using the same task scenario and think aloud approach to confirm the validity of the existing model, and potentially identify additional attributes for inclusion with specific use cases.

14.9.2. Alternative methodologies

This research study focused on an interpretive approach applying task scenarios with the think aloud protocol. Whilst this activity matched standard user research practice, varying the methodology may further validate the model and suggest alterations to improve its effectiveness. For example, moving towards observation of real user activities rather than a simulated task scenario may help uncover additional insights within the organisational context.

14.9.3. Practitioner validation

Following the establishment of the Informative Web Content Guidelines (IWCG), a further round of validation by practitioners would demonstrate the potential for the model to improve the informativeness of web content. This could take the form of a focus group or series of interviews aimed at eliciting the views on the model of a sufficient number of content author practitioners. Using their own experiences, practitioners could inform the approaches to meeting guidelines and suggested advice to incorporate further empirical suggestions as frequently occurs with the parallel WCAG guidelines.

14.9.4. Additional content types

This research has been conducted with a focus on text-based online content as this is consistent across user devices and abstracts from the system and Human-Computer Interface (HCI) elements of a website. Whilst content has been defined by the researcher to include information, that is data with a meaning (see p. 41), the focus has been solely on text rather than the images and media that could also be considered as informative.

The Web Content Accessibility Guidelines (WCAG) recognise the important of non-text content and include specific provisions for ensuring that these are accessible to all users. In a similar manner, the Informative Web Content Guidelines (IWCG) could be expanded to include guidance to produce informative content within images, video, and other media. This would require further studies to identify typical user interactions with these elements, the appropriate information quality attributes, and the addition of guidelines to the model to aid practitioners. This would benefit content authors on the modern web where there are increasing expectations for interactive, multimedia-rich websites as an alternative to plain text.

14.9.5. Interaction between content and system

In Section 3.2.4 of the literature review, the author introduced the DeLone-McLean model for information systems success. This provides an approach to analysing the impact of three aspects of quality within a technological artefact: the system, the information, and the service. A distinction is made between the three elements, with the system and service being accounted for in academic studies of HCI and practitioner applications of UX, with a lack of focus on the information quality.

Whilst the model treats these three elements as interconnected concepts of quality, HCI research often faces the challenge of decoupling the system from the information contained within. This is reflected in the WCAG

practitioner model, which includes advice about both presentational (system/HCI) and content aspects.

This research study has been designed to minimise the impact of system elements such as navigational, layout and presentation structures. This has necessitated the use of websites consistently within each case study and isolation of the attributes related solely to these aspects during both the literature review ana analysis stages. Despite the researcher's efforts, there may be elements of the study that still rely on specific presentational aspects, which could be mitigated by further study of additional task scenarios within both the same and additional case studies.

14.9.6. Alignment with WCAG

The practitioner model has been inspired by the Web Content Accessibility Guidelines (WCAG), following the same approach as the forthcoming WCAG 3.0 standard. This provides a consistent documentation and approach for content authors to understand and apply the recommendations, translating the academic advice into practical guidelines.

Even with this approach, there is the opportunity for future research to align the model even closer with the WCAG. For example, incorporating content aspects as an additional guideline would provide a wider reach and enhance the existing guidance for web professionals without disrupting existing practice. Additional study into the widespread applicability of the model would be required to ensure consistency of guidance, testing, scoring, and conformity across both sets of guidelines.

14.9.7. Automated analysis

The transition from WCAG 2.1 to WCAG 3.0 introduces the possibility of automated analysis of several accessibility guidelines, with outcomes denoted by numerical ratings (akin to a Likert scale) rather than the former

levels A, AA, and AAA. By using the WCAG 3.0 format for the IWCG model, future research could identify how individual information quality attributes could be automatically analysed and reported on. This would enable practitioners to discover which elements of their content could be improved upon without performing the analysis themselves, increasing the utility of the model whilst reducing the time required for its implementation.

14.9.8. Further extensions to WCAG

The development of the IWCG followed the established WCAG model for accessibility. The W3C have several working groups working on related guidelines for authoring tools, user agents, mobile devices, and other technological developments. The model proposed through this research provides one possible extension of the WCAG standard to incorporate informativeness as an additional consideration for content authors, though there is the possibility of further studies into other areas such as information architecture and immersive/multimedia environments. This will become more essential as the web develops away from traditional browsing experiences in the future.

14.9.9. Summary

There are several possible avenues for future research emerging from this study that would further enhance the outcomes. Conducting studies in additional domains (whether through existing or new methodologies) and supplementary validation with both practitioners and representative users would provide further evidence that the Informative Web Content Guidelines (IWCG) provide a suitable academically grounded approach to improving the informativeness of website content.

14.10. Conclusion

This research study has developed and validated a new model for analysing and improving the informativeness of online content for stakeholders, customers, and service users alike. Through this model, practitioners can

explore the information quality attributes with the largest impact on their content, with practical advice for applying improvements to text-based information. This has been demonstrated by the validation case study and could be further explored through the potential research agenda outlined above.

References

Abascal, J., & Nicolle, C. (2005). Moving towards inclusive design guidelines for socially and ethically aware HCI. *Interacting with Computers*, *17*(5), 484–505.

Ackoff, R. (1999). *Ackoff's best: His classic writings on management*. John Wiley & Sons.

Adam, A., & Kreps, D. (2009). Disability and discourses of web accessibility. *Information, Communication & Society*, *12*(7), 1041–1058.

Adams, A., & Cox, A. L. (2008). Questionnaires, in-depth interviews and focus groups. In P. Cairns & A. L. Cox (Eds.), *Research Methods for Human-Computer Interaction* (pp. 17–34). Cambridge University Press.

Adams, S. A. (2010). Revisiting the online health information reliability debate in the wake of "web 2.0": An inter-disciplinary literature and website review. *International Journal of Medical Informatics*, *79*(6), 391–400.

Ahmed, I., & Blustein, J. (2006). Influence of spatial ability in navigation: Using look-ahead breadcrumbs on The Web. *International Journal of Web Based Communities*, *2*(2), 183–196.

Aladwani, A. M., & Palvia, P. C. (2002). Developing and validating an instrument for measuring user-perceived web quality. *Information & Management*, *39*(6), 467–476.

Al-Debei, M. M., Jalal, D., & Al-Lozi, E. (2013). Measuring web portals success: A respecification and validation of the DeLone and McLean information systems success model. *International Journal of Business Information Systems*, *14*(1), 96–133.

Alexander, R., Murray, D., & Thompson, N. (2017). Cross-cultural web usability model. *International Conference on Web Information Systems Engineering*, 75–89.

Alfano, M., Lenzitti, B., Taibi, D., & Helfert, M. (2021). On-Line Retrieval of Health Information Based on Language Complexity, Information Customization and Information Quality. *Information and Communication Technologies for Ageing Well and E-Health: 6th International Conference, ICT4AWE 2020, Prague, Czech Republic, May 3–5, 2020, Revised Selected Papers*, 1.

Al-Jefri, M. (2019). *Automatic Evaluation of Online Health Information Quality* [PhD Thesis]. University of Brighton.

Al-Jefri, M., Evans, R., Uchyigit, G., & Ghezzi, P. (2018). What Is Health Information Quality? Ethical Dimension and Perception by Users. *Frontiers in Medicine*, *5*(260).

Altuntaç, P. (2015). *The Comparison of Concurrent and Retrospective Think Aloud Methods in Unmoderated Remote Usability Testing* [Universiteit Leiden]. https://openaccess.leidenuniv.nl/bitstream/handle/1887/34669/Altuntac%2C%20Pelin-s1414852-MA%20Thesis-ACP-2015.pdf

Andreasen, M. S., Nielsen, H. V., Schrøder, S. O., & Stage, J. (2007). What Happened to Remote Usability Testing?: An Empirical Study of Three Methods. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1405–1414.

AoIR. (2012). *Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0)*. https://aoir.org/reports/ethics2.pdf

Arazy, O., Kopak, R., & Hadar, I. (2017). Heuristic principles and differential judgments in the assessment of information quality. *Journal of the Association for Information Systems*, *18*(5), 1.

Argyris, C., & Schon, D. (1978). Organizational learning: A theory of action approach. *Reading, MA: Addision Wesley*.

Astley, W. G., & Van de Ven, A. H. (1983). Central perspectives and debates in organization theory. *Administrative Science Quarterly*, 245–273.

Ataloglou, M. P., & Economides, A. A. (2009). Evaluating European ministries' websites. *International Journal of Public Information Systems*, *5*(3).

Avison, D. E., Davison, R., & Malaurent, J. (2018). Information systems action research: Debunking myths and overcoming barriers. *Information & Management*, *55*, 177–187.

Avison, D. E., Lau, F., Myers, M. D., & Nielsen, P. A. (1999). Action research. *Communications of the ACM*, *42*(1), 94–97.

Babbie, R. (2010). *The Practice of Social Research*. Wadsworth Cangage Learning.

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern Information Retrieval: The Concepts and Technology behind Search* (2nd ed.). ACM Press.

Ball, M. J., & Lillis, J. (2001). E-health: Transforming the physician/patient relationship. *International Journal of Medical Informatics*, *61*(1), 1–10.

Ballou, D. P., & Pazer, H. L. (1995). Designing information systems to optimize the accuracy-timeliness tradeoff. *Information Systems Research*, *6*(1), 51–72.

Barker, D. (2016). *Web content management: Systems, features, and best practices*. O'Reilly Media.

Barnes, S. J., & Vidgen, R. T. (2002). An integrative approach to the assessment of e-commerce quality. *J. Electron. Commer. Res.*, *3*(3), 114–127.

Barnes, S., & Vidgen, R. (2000). WebQual: An exploration of website quality. *ECIS 2000 Proceedings*, *74*.

Baskerville, R. L. (1999). Investigating information systems with action research. *Communications of the AIS*, *2*(19).

Baskerville, R. L., & Myers, M. D. (2004). Special issue on action research in information systems: Making IS research relevant to practice: Foreword. *MIS Quarterly*, 329–335.

Baskerville, R. L., & Myers, M. D. (2009). Fashion waves in information systems research and practice. *Mis Quarterly*, 647–662.

Baskerville, R. L., & Pries-Heje, J. (1999). Grounded action research: A method for understanding IT in practice. *Accounting, Management and Information Technologies*, *9*(1), 1–23.

Baskerville, R. L., & Wood-Harper, A. T. (1996). A critical perspective on action research as a method for information systems research. *Journal of Information Technology*, *11*(3), 235–246.

Bates, M. J. (2002). Toward an integrated model of information seeking and searching. *The New Review of Information Behaviour Research*, *3*(1), 1–15.

Bauer, H. H., Falk, T., & Hammerschmidt, M. (2006). eTransQual: A transaction process-based approach for capturing service quality in online shopping. *Journal of Business Research*, *59*(7), 866–875.

Baughan, A., August, T., Yamashita, N., & Reinecke, K. (2020). Keep it Simple: How Visual Complexity and Preferences Impact Search Efficiency on Websites. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–10.

BCS. (2015). *Code of Conduct*. http://www.bcs.org/upload/pdf/conduct.pdf

Benbasat, I., & Barki, H. (2007). Quo vadis TAM? *Journal of the Association for Information Systems*, *8*(4), 211–218.

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The case research strategy in studies of information systems. *MIS Quarterly*, 369–386.

Benbasat, I., & Zmud, R. W. (1999). Empirical research in information systems: The practice of relevance. *MIS Quarterly*, 3–16.

Benbasat, I., & Zmud, R. W. (2003). The identity crisis within the IS discipline: Defining and communicating the discipline's core properties. *MIS Quarterly*, 183–194.

Benbunan-Fich, R. (2001). Using protocol analysis to evaluate the usability of a commercial web site. *Information & Management*, *39*(2), 151–163.

Berners-Lee, T. J. (1989). *Information management: A proposal*.

Bickford, P. (1997, October). Worth the Wait? *Netscape/View Source Mag*.

Blandford, A. (2019). Semi-structured qualitative studies. In *The Encyclopedia of Human-Computer Interaction* (2nd ed.). https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/semi-structured-qualitative-studies

Bødker, S. (2006). When second wave HCI meets third wave challenges. *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles*, 1–8.

Boland, R. J., & Hirschheim, R. A. (1985). *Series Introduction to the first volume of the Wiley Series on Information Systems*.

Boren, T., & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication*, *43*(3), 261–278.

Borlund, P., & Ingwersen, P. (1997). The Development of a Method for the Evaluation of Interactive Information Retrieval Systems. *Journal of Documentation*, *53*(3), 225–250.

Bovee, M., Srivastava, R. P., & Mak, B. (2003). A conceptual framework and belief-function approach to assessing overall information quality. *International Journal of Intelligent Systems*, *18*(1), 51–74.

Boyer, C., Selby, M., Scherrer, J.-R., & Appel, R. (1998). The health on the net code of conduct for medical and health websites. *Computers in Biology and Medicine*, *28*(5), 603–610.

Braa, K., & Vidgen, R. (1999). Interpretation, intervention, and reduction in the organizational laboratory: A framework for in-context information

system research. *Accounting, Management and Information Technologies*, *9*(1), 25–47.

Brajnik, G. (2009). Validity and reliability of web accessibility guidelines. *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility*, 131–138.

Brajnik, G., Yesilada, Y., & Harper, S. (2012). Is accessibility conformance an elusive property? A study of validity and reliability of WCAG 2.0. *ACM Transactions on Accessible Computing (TACCESS)*, *4*(2), 8.

Briggs, R. O., De Vreede, G.-J., Nunamaker, J. F., & Sprague, R. H. (2003). Information systems success. *Journal of Management Information Systems*, *19*(4), 5–8.

Brush, A. J. B., Ames, M., & Davis, J. (2004). A Comparison of Synchronous Remote and Local Usability Studies for an Expert Interface. *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, 1179–1182. https://doi.org/10.1145/985921.986018

Bryant, A., Black, A., Land, F., & Porra, J. (2013). Information Systems history: What is history? What is IS history? What IS history?… and why even bother with history? *Journal of Information Technology*, *28*, 1–17.

Buckland, M. K. (1991). Information as Thing. *Journal of the American Society for Information Science*, *42*(5), 351–360.

Budiu, R. (2017, October 1). *Quantitative vs. Qualitative Usability Testing*. Nielsen Norman Group. https://www.nngroup.com/articles/quant-vs-qual/

Buenadicha Mateos, M., Chamorro Mera, A., Miranda González, F. J., & González López, Ó. R. (2001). A new Web assessment index: Spanish universities analysis. *Internet Research*, *11*(3), 226–234.

Burrell, G., & Morgan, G. (1979). *Sociological paradigms and organisational analysis: Elements of the sociology of corporate life*. Heinemann.

Büttner, O. B., & Göritz, A. S. (2008). Perceived trustworthiness of online shops. *Journal of Consumer Behaviour: An International Research Review*, *7*(1), 35–50.

Caballero, I., Verbo, E., Calero, C., & Piattini, M. (2007). A Data Quality Measurement Information Model Based On ISO/IEC 15939. *ICIQ*, 393–408.

Capra, R., Marchionini, G., Oh, J. S., Stutzman, F., & Zhang, Y. (2007). Effects of structure and interaction style on distinct search tasks. *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*, 442–451.

Carroll, J. M. (1999). Five reasons for scenario-based design. *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32.*, 1–11.

Castañeda, J. A., Muñoz-Leiva, F., & Luque, T. (2007). Web Acceptance Model (WAM): Moderating effects of user experience. *Information & Management*, *44*(4), 384–396.

Cecez-Kecmanovic, D., Davison, R. M., Fernandez, W., Finnegan, P., Pan, S. L., & Sarker, S. (2020). Advancing qualitative IS research methodologies: Expanding horizons and seeking new paths. *Journal of the Association for Information Systems*, *21*(1), 246–263.

Chakraborty, G., Srivastava, P., & Warren, D. L. (2005). Understanding corporate B2B web sites' effectiveness from North American and European perspective. *Industrial Marketing Management*, *34*(5), 420–429.

Charnock, D., Shepperd, S., Needham, G., & Gann, R. (1999). DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *Journal of Epidemiology & Community Health*, *53*(2), 105–111.

Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal*, *12*(2).

Checkland, P. (1981). *Systems Thinking, Systems Practice*. Wiley.

Checkland, P. (1985). From optimizing to learning: A development of systems thinking for the 1990s. *Journal of the Operational Research Society*, *757–767*.

Checkland, P. (1991). From framework through experience to learning: The essential nature of action research. *Information Systems Research: Contemporary Approaches and Emergent Traditions*, 1–7.

Checkland, P., & Holwell, S. (1997). *Information, systems and information systems: Making sense of the field*. Wiley.

Checkland, P., & Poulter, J. (2006). *Learning for action: A short definitive account of soft systems methodology and its use, for practitioners, teachers and students*. John Wiley & Sons Ltd.

Checkland, P., & Scholes, J. (1990). *Soft systems methodology in action*. Wiley.

Cheung, M. Y., Luo, C., Sia, C. L., & Chen, H. (2009). Credibility of electronic word-of-mouth: Informational and normative determinants of on-line

consumer recommendations. *International Journal of Electronic Commerce*, *13*(4), 9–38.

Chopra, K., & Wallace, W. A. (2003). Trust in electronic environments. *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference On*, 10-pp.

Chu, J., & Yang, G. (2010). A culture-based study on information density of e-commerce websites. *2010 International Conference On Computer Design and Applications*, *4*, V4-60.

Chuttur, M. Y. (2009). Overview of the technology acceptance model: Origins, developments and future directions. *Working Papers on Information Systems*, *9*(37), 9–37.

Clark, D. (2007). Content management and the separation of presentation and content. *Technical Communication Quarterly*, *17*(1), 35–60.

Clarkson, P. J., & Coleman, R. (2015). History of Inclusive Design in the UK. *Applied Ergonomics*, *46*, 235–247.

Clemmensen, T., Hertzum, M., Hornbœk, K., Shi, Q., & Yammiyavar, P. (2009). Cultural cognition in usability evaluation. *Interacting with Computers*, *21*(3), 212–220.

Cockton, G., & Woolrych, A. (2001). Understanding Inspection Methods: Lessons from an Assessment of Heuristic Evaluation. In A. Blandford, J. Vanderdonckt, & P. Gray (Eds.), *People and Computers XV—Interaction without Frontiers* (pp. 171–191). Springer London.

Coombs, C. (2017). Coherence and transparency: Some advice for qualitative researchers. *Production*, *27*.

Cooper, M., Sloan, D., Kelly, B., & Lewthwaite, S. (2012). A challenge to web accessibility metrics and guidelines: Putting people and processes first. *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*.

Corbin, J., & Strauss, A. (2014). *Basics of qualitative research*. Sage Publications.

Coyier, C. (2020, October 21). *A Complete Guide to Links and Buttons*. CSS Tricks. https://css-tricks.com/a-complete-guide-to-links-and-buttons/

Creswell, J. W. (2008). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (3rd ed.). Sage Publications Ltd.

Creswell, J. W., & Clark, V. L. P. (2007). *Designing and conducting mixed methods research*. Sage publications.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. Sage.

Dardailler, D. (2009, June). *WAI early days*. W3C Web Accessibility Initiative. https://www.w3.org/WAI/history

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319–340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*(8), 982–1003.

Davis, G. B. (1999). A research perspective for information systems and example of emerging area of research. *Information Systems Frontiers*, *1*(3), 195–203.

Davis, G. B., & Olson, M. H. (1985). *Management information systems: Conceptual foundations, structure, and development*. McGraw-Hill, Inc.

Davis, N. (2011). Information overload, reloaded. *Bulletin of the American Society for Information Science and Technology*, *37*(5), 45–49.

Davison, R., Martinsons, M. G., & Kock, N. (2004). Principles of canonical action research. *Information Systems Journal*, *14*(1), 65–86.

De Villiers, M. (2005). Three approaches as pillars for interpretive information systems research: Development research, action research and grounded theory. *Proceedings of the 2005 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries*, 142–151.

Debevc, M., Kosec, P., & Holzinger, A. (2011). Improving multimodal web accessibility for deaf people: Sign language interpreter module. *Multimedia Tools and Applications*, *54*(1), 181–199.

DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, *3*(1), 60–95.

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, *19*(4), 9–30.

DeLuca, D., Gallivan, M. J., & Kock, N. (2008). Furthering information systems action research: A post-positivist synthesis of four dialectics. *Journal of the Association for Information Systems*, *9*(2), 48.

Deming, W. E. (1986). *Out of the Crisis*. MIT press.

Dervin, B., & Nilan, M. (1986). Information needs and uses. *Annual Review of Information Science and Technology*, *21*, 3–33.

Dickens, L., & Watkins, K. (1999). Action research: Rethinking Lewin. *Management Learning*, *30*(2), 127–140.

Dillon, A., & Turnbull, D. (2005). *Information architecture*. New York: Marcel Dekker.

Dinet, J., Chevalier, A., & Tricot, A. (2012). Information search activity: An overview. *European Review of Applied Psychology*, *62*(2), 49–62.

Donahue, G. M. (2001). Usability and the bottom line. *IEEE Software*, *18*(1), 31.

Donnelly, A., & Magennis, M. (2002). Making accessibility guidelines usable. *ERCIM Workshop on User Interfaces for All*, 56–67.

Dunlap, J. C., & Lowenthal, P. R. (2016). Getting graphic about infographics: Design lessons learned from popular infographics. *Journal of Visual Literacy*, *35*(1), 42–59.

Dutta-Bergman, M. J. (2004). The impact of completeness and web use motivation on the credibility of e-health information. *Journal of Communication*, *54*(2), 253–269.

Dwivedi, Y. K., Wastell, D., Laumer, S., Henriksen, H. Z., Myers, M. D., Bunker, D., Elbanna, A., Ravishankar, M., & Srivastava, S. C. (2015). Research on information systems failures and successes: Status update and future directions. *Information Systems Frontiers*, *17*(1), 143–157.

Ellcessor, E. (2010). Bridging disability divides: A critical history of web content accessibility through 2001. *Information, Communication & Society*, *13*(3), 289–308.

Ellcessor, E. (2014). <ALT="Textbooks">: Web accessibility myths as negotiated industrial lore. *Critical Studies in Media Communication*, *31*(5), 448–463.

Ellis, W. D. (2013). *A source book of Gestalt psychology*. Routledge.

EN 301 549. (2014). *Accessibility requirements suitable for public procurement of ICT products and services in Europe*. European Commission. http://mandate376.standards.eu/standard

Eppler, M. J. (2015). 11. Information quality and information overload: The promises and perils of the information age. *Communication and Technology*, *5*.

Eppler, M. J., & Muenzenmayer, P. (2002). Measuring Information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an

Application Methodology. *Proceedings of the Seventh International Conference on Information Quality (ICIQ-02)*, 187–196.

Eppler, M. J., & Wittig, D. (2000). Conceptualizing Information Quality: A Review of Information Quality Frameworks from the Last Ten Years. *Proceedings of the 2000 Conference on Information Quality*, 83–96.

Ericsson, K. A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. *The Cambridge Handbook of Expertise and Expert Performance*, 223–241.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. the MIT Press.

European Commission. (2021, July 19). *Web accessibility*. European Digital Strategy. https://digital-strategy.ec.europa.eu/en/policies/web-accessibility

European Parliament. (2016). *Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies*. http://data.europa.eu/eli/dir/2016/2102/oj/eng

Eveland, W. P., & Dunwoody, S. (2000). Examining information processing on the World Wide Web using think aloud protocols. *Media Psychology*, *2*(3), 219–244.

Ewusi-Mensah, K. (2003). *Software development failures*. MIT Press.

Ewusi-Mensah, K., & Przasnyski, Z. H. (1994). Factors contributing to the abandonment of information systems development projects. *Journal of Information Technology*, *9*(3), 185–201.

Eysenbach, G. (2001). What is e-health? *Journal of Medical Internet Research*, *3*(2), e20.

Fadahunsi, K. P., O'Connor, S., Akinlua, J. T., Wark, P. A., Gallagher, J., Carroll, C., Car, J., Majeed, A., & O'Donoghue, J. (2021). Information quality frameworks for digital health technologies: Systematic review. *Journal of Medical Internet Research*, *23*(5), e23479.

Fan, M., Lin, J., Chung, C., & Truong, K. N. (2019). Concurrent think-aloud verbalizations and usability problems. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *26*(5), 1–35.

Fan, M., Shi, S., & Truong, K. N. (2020). Practices and Challenges of Using Think-Aloud Protocols in Industry: An International Survey. *Journal of Usability Studies*, *15*(2).

Fan, M., Zhao, Q., & Tibdewal, V. (2021). Older Adults' Think-Aloud Verbalizations and Speech Features for Identifying User Experience Problems. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–13.

Feldman, J. (2011). Computers and thought—The back story. *IEEE Annals of the History of Computing*, *33*(4), 56–61.

Fenn, T., & Hobbs, J. (2014). The information architecture of meaning making. In *Reframing information architecture* (pp. 11–30). Springer.

Ferre, X., Juristo, N., & Moreno, A. (2005). Framework for integrating usability practices into the software process. *Product Focused Software Process Improvement*, 202–215.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Addison-Wesley.

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. *Journalism & Mass Communication Quarterly*, *77*(3), 515–540.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221–233.

Floridi, L. (2005). Is semantic information meaningful data? *Philosophy and Phenomenological Research*, *70*(2), 351–370.

Flyvbjerg, B., & Budzier, A. (2011). Why Your IT Project Might Be Riskier Than You Think. *Harvard Business Review*, *89*(9), 601–603.

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do Procedures for Verbal Reporting of Thinking Have to Be Reactive? A Meta-Analysis and Recommendations for Best Reporting Methods. *Psychological Bulletin*, *137*(2), 316–344.

Francis, J. E., & White, L. (2002). PIRQUAL: a scale for measuring customer expectations and perceptions of quality in internet retailing. *Marketing Throry and Applications*, 263–270.

Freire, A. P., Fortes, R. P., Turine, M. A., & Paiva, D. M. (2008). An evaluation of web accessibility metrics based on their attributes. *Proceedings of the 26th Annual ACM International Conference on Design of Communication*, 73–80.

Funk, J., Lee, Y., & Wang, R. (1998). Institutionalizing information quality practice. *Proceedings of the Conference on Information Quality*, *3*, 1–17.

Fusch, P. I., & Ness, L. R. (2015). Are we there yet? Data saturation in qualitative research. *The Qualitative Report*, *20*(9), 1408–1416.

Gable, G. G. (1996). A multidimensional model of client success when engaging external consultants. *Management Science*, *42*(8), 1175–1198.

Gable, G. G., Sedera, D., & Chan, T. (2008). Re-conceptualizing information system success: The IS-impact measurement model. *Journal of the Association for Information Systems*, *9*(7), 377–408.

Galliers, R. D., & Land, F. F. (1987). Choosing appropriate information systems research methodologies. *Communications of the ACM*, *30*(11), 901–902.

Garrett, J. J. (2010). *The elements of user experience: User-centered design for the web and beyond*. Pearson Education.

Gates, B. (1996, January 3). *Content is King* [Blog]. http://www.microsoft.com/billgates/columns/1996essay/essay960103.asp

Ge, M., & Helfert, M. (2007). A review of information quality research—Develop a research agenda. *Paper Presented at the International Conference on Information Quality 2007*.

Ge, M., Helfert, M., & Jannach, D. (2011). Information quality assessment: Validating measurement dimensions and processes. *Proceedings of the European Conference of Information Systems (ECIS)*.

Gelfert, A. (2018). Fake news: A definition. *Informal Logic*, *38*(1), 84–117.

George, C. E., & Scerri, J. (2007). Web 2.0 and User-Generated Content: Legal challenges in the new frontier. *Journal of Information, Law and Technology*, *2*.

Gernert, D. (1996). Pragmatic information as a unifying concept. *Information: New Questions to a Multidisciplinary Concept*.

Ghasemaghaei, M., & Hassanein, K. (2015). Online information quality and consumer satisfaction: The moderating roles of contextual factors–A meta-analysis. *Information & Management*, *52*(8), 965–981.

Glass, R. L. (2005). IT Failure Rates-70% or 10-15%? *IEEE Software*, *22*(3), 109–111.

Go, K., & Carroll, J. M. (2004). The Blind Men and the Elephant: Views of Scenario-based System Design. *Interactions*, *11*(6), 44–53.

Goldkuhl, G. (2012). Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems*, *21*(2), 135–146.

Goles, T., & Hirschheim, R. (2000). The paradigm is dead, the paradigm is dead… long live the paradigm: The legacy of Burrell and Morgan. *Omega*, *28*(3), 249–268.

Goodhue, D. L. (2007). Comment on Benbasat and Barki's 'Quo Vadis TAM' article. *Journal of the Association for Information Systems*, *8*(4), 219–222.

Google. (2021). *Google Books Ngram Viewer, content*. https://books.google.com/ngrams/graph?content=content&year_start=1750&year_end=2019&corpus=26&smoothing=10

Gorla, N., Somers, T. M., & Wong, B. (2010). Organizational impact of system quality, information quality, and service quality. *The Journal of Strategic Information Systems*, *19*(3), 207–228.

Government Digital Service. (2012, April 3). *Government design principles*. https://www.gov.uk/guidance/government-design-principles

Government Digital Service. (2016, February 25). *Content design: Planning, writing and managing content*. Gov.Uk Guidance. https://www.gov.uk/guidance/content-design/writing-for-gov-uk

Government Digital Service. (2018, May 9). *New accessibility regulations for public sector websites and apps*. https://www.gov.uk/guidance/accessibility-requirements-for-public-sector-websites-and-apps

Government Digital Service. (2020, April 28). *Making your service accessible: An introduction*. GOV.UK Service Manual. https://www.gov.uk/service-manual/helping-people-to-use-your-service/making-your-service-accessible-an-introduction

Greene, J. C., Caracelli, V. J., & Graham, W. F. (2008). Toward a conceptual framework for mixed-method evaluation designs. In C. L. Plano Clark & J. W. Creswell (Eds.), *The Mixed Methods Reader*. Sage.

Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly*, 611–642.

Grover, V., Jeong, S. R., & Segars, A. H. (1996). Information systems effectiveness: The construct space and patters of application. *Information & Management*, *31*(4), 177–191.

Gunawardena, C., & Brown, D. H. (2007). IS initiatives in the vocational and technical education sector of developing Asian countries: A systems

approach to the management of project intervention processes. *The Electronic Journal of Information Systems in Developing Countries*, *30*, 1–19.

Halawi, L. A., McCarthy, R. V., & Aronson, J. E. (2008). An empirical investigation of knowledge management systems' success. *Journal of Computer Information Systems*, *48*(2), 121–135.

Hargittai, E., Fullerton, L., Menchen-Trevino, E., & Thomas, K. Y. (2010). Trust online: Young adults' evaluation of web content. *International Journal of Communication*, *4*, 468–494.

Hassenzahl, M., & Monk, A. (2010). The inference of perceived usability from beauty. *Human–Computer Interaction*, *25*(3), 235–260.

Haug, A., Zachariassen, F., & Van Liempd, D. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management*, *4*(2), 168–193.

Hayashi Jr, P., Abib, G., & Hoppen, N. (2019). Validity in Qualitative Research: A Processual Approach. *The Qualitative Report*, *24*(1), 98–112.

Heeks, R. (2006). Health information systems: Failure, success and improvisation. *International Journal of Medical Informatics*, *75*(2), 125–137.

Helfert, M. (2001). Managing and measuring data quality in data warehousing. *Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics*, *38*.

Hellstén, S.-M., & Markova, M. (2006). The DeLone and McLean model of information systems success-original and updated models. *Proceedings of the SIGCHI Conference*, 1–5.

Heo, J., & Han, I. (2003). Performance measure of information systems (IS) in evolving computing environments: An empirical investigation. *Information & Management*, *40*(4), 243–256.

Hernández, B., Jiménez, J., & Martín, M. J. (2009). Key website factors in e-business strategy. *International Journal of Information Management*, *29*(5), 362–371.

Heron, J., & Reason, P. (1997). A participatory inquiry paradigm. *Qualitative Inquiry*, *3*(3), 274–294.

Hertzum, M., Hansen, K. D., & Andersen, H. H. K. (2009). Scrutinising usability evaluation: Does thinking aloud affect behaviour and mental workload? *Behaviour & Information Technology*, *28*(2), 165–181.

Hertzum, M., & Jacobsen, N. E. (2003). The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. *International Journal of Human–Computer Interaction*, *15*(1), 183–204. https://doi.org/10.1207/S15327590IJHC1501_14

Hertzum, M., & Pejtersen, A. M. (2000). The information-seeking practices of engineers: Searching for documents as well as for people. *Information Processing & Management*, *36*(5), 761–778.

Hirschheim, R., & Klein, H. K. (2012). A glorious and not-so-short history of the information systems field. *Journal of the Association for Information Systems*, *13*(4), 188.

Hirschheim, R., & Newman, M. (1988). Information systems and user resistance: Theory and practice. *The Computer Journal*, *31*(5), 398–408.

Ho, K., & Sculli, D. (1994). Organizational theory and soft systems methodologies. *Journal of Management Development*, *13*(7), 47–58.

Holgeid, K., & Thompson, M. (2013). A reflection on why large public projects fail. *The Governance of Large-Scale Projects*, 219–244.

Hoppmann, T. K. (2009). Examining the 'point of frustration'. The think-aloud method applied to online search tasks. *Quality & Quantity*, *43*(2), 211–224.

Hornbæk, K., & Hertzum, M. (2017). Technology acceptance and user experience: A review of the experiential component in HCI. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *24*(5), 33:1-33:30.

Hughes, J., & Wood-Harper, T. (1999). Systems development as a research act. *Journal of Information Technology*, *14*(1), 83–94.

Huizingh, E. K. (2000). The content and design of web sites: An empirical study. *Information & Management*, *37*(3), 123–134.

Iivari, N. (2005). Usability Specialists – 'A Mommy Mob', 'Realistic Humanists' or 'Staid Researchers'? An Analysis of Usability Work in the Software Product Development. In M. F. Costabile & F. Paternò (Eds.), *Human-Computer Interaction—INTERACT 2005* (pp. 418–430). Springer Berlin Heidelberg.

Ingwersen, P., & Järvelin, K. (2006). *The turn: Integration of information seeking and retrieval in context*. Springer Science & Business Media.

ISO 9001. (2015). *Quality management systems. Requirements*. International Organization for Standardization.

ISO 9241-151. (2008). *Ergonomics of human-system interaction—Part 151: Guidance on World Wide Web user interfaces*. International Organization for Standardization.

ISO 13407. (1999). *Human-Centred Design Processes for Interactive Systems*. International Organization for Standardization.

ISO TR 16982. (2002). *Usability methods supporting human-centred design*. International Organization for Standardization.

Janan, D., & Wray, D. (2013). Research into readability: Paradigms and possibilities. *New Literacies: Reconstructing Language and Education, A. Pandian, C. Liew, C. Ling and C. Tan Ai Lin (Eds)*, 367–377.

Jaspers, M. W. M. (2009). A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence. *International Journal of Medical Informatics*, *78*(5), 340–353.

Juristo, N., Windl, H., & Constantine, L. (2001). Guest Editors' Introduction: Introducing Usability. *IEEE Software*, *18*(1), 20–21.

Kahn, B. K., Strong, D. M., & Wang, R. Y. (2002). Information quality benchmarks: Product and service performance. *Communications of the ACM*, *45*(4), 184–192.

Kąkol, M., Jankowski-Lorek, M., Abramczuk, K., Wierzbicki, A., & Catasta, M. (2013). On the subjectivity and bias of web content credibility evaluations. *Proceedings of the 22nd International Conference on World Wide Web*, 1131–1136.

Kandari, J. S., Jones, E. C., Nah, F. F.-H., & Bishu, R. R. (2011a). Information quality on the world wide web: A framework to measure and its validation. *Proceedings of the 17th International Conference on Information Systems Analysis and Synthesis*.

Kandari, J. S., Jones, E. C., Nah, F. F.-H., & Bishu, R. R. (2011b). Information quality on the world wide web: Development of a framework. *International Journal of Information Quality*, *2*(4), 324–343.

Kang, Y.-S., & Kim, Y. J. (2006). Do visitors' interest level and perceived quantity of web page content matter in shaping the attitude toward a web site? *Decision Support Systems*, *42*(2), 1187–1202.

Karanam, S., Van Oostendorp, H., Sanchiz, M., Chevalier, A., Chin, J., & Fu, W. T. (2015). Modeling and predicting information search behavior. *Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics*, 1–12.

Kelton, K., Fleischmann, K. R., & Wallace, W. A. (2008). Trust in digital information. *Journal of the American Society for Information Science and Technology*, *59*(3), 363–374.

Keutel, M., Michalik, B., & Richter, J. (2014). Towards mindful case study research in IS: a critical analysis of the past ten years. *European Journal of Information Systems*, *23*(3), 256–272.

Kim, J. (2012). Scenarios in information seeking and information retrieval research: A methodological application and discussion. *Library & Information Science Research*, *34*(4), 300–307.

Kim, S., & Stoel, L. (2004). Apparel retailers: Website quality dimensions and satisfaction. *Journal of Retailing and Consumer Services*, *11*(2), 109–117.

Klaus, T., & Blanton, J. E. (2010). User resistance determinants and the psychological contract in enterprise system implementations. *European Journal of Information Systems*, *19*(6), 625–636.

Klein, B. (2002). When do users detect information quality problems on the world wide web? *AMCIS 2002 Proceedings*, 1101–1103.

Klein, H. K., & Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly*, 67–93.

Knight, S., & Burn, J. (2005). Developing a framework for assessing information quality on the World Wide Web. *Informing Science*, *8*.

Koffka, K. (1935). *Principles of Gestalt Psychology*. Harcourt.

Kostur, P. (2006). Incorporating Usability into Content Management. *IEEE International Professional Communication Conference*, 193–196.

Kreps, D. (2018). *Against Nature: The Metaphysics of Information Systems*. Routledge.

Kreps, D., & Goff, M. (2015). Code in action: Closing the black box of WCAG 2.0, A Latourian reading of Web accessibility. *First Monday*, *20*(9).

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage publications.

Lawson, B. (2006). Introduction. In *Web accessibility: Web standards and regulatory compliance* (pp. xxv–xlvii). Springer Verlag.

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction*. Morgan Kaufmann.

Lazonder, A. W., Biemans, H. J. A., & Wopereis, I. G. J. H. (2000). Differences Between Novice and Experienced Users in Searching Information on

the World Wide Web. *Journal of the American Society for Information Science*, *51*(6), 576–581.

Lederer, A. L., Maupin, D. J., Sena, M. P., & Zhuang, Y. (2000). The technology acceptance model and the World Wide Web. *Decision Support Systems*, *29*(3), 269–282.

Lee, A. S. (2004). Thinking about social theory and philosophy for information systems. *Social Theory and Philosophy for Information Systems*, *1*, 26.

Lee, J. C., & Myers, M. D. (2004). Dominant actors, political agendas, and strategic shifts over time: A critical ethnography of an enterprise systems implementation. *The Journal of Strategic Information Systems*, *13*(4), 355–374.

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment. *Information & Management*, *40*(2), 133–146.

Lew, P., Olsina, L., & Zhang, L. (2010). Quality, quality in use, actual usability and user experience as key drivers for web application evaluation. *Web Engineering*, 218–232.

Lewandowski, D. (2005). Web searching, search engines and Information Retrieval. *Information Services & Use*, *25*(3–4), 137–147.

Lewis, C. (1982). Using the'thinking-aloud'method in cognitive interface design. *Research Report RC9265, IBM TJ Watson Research Center*.

Lewthwaite, S. (2014). Web accessibility standards and disability: Developing critical perspectives on accessibility. *Disability and Rehabilitation*, *36*(16), 1375–1383.

Lindgaard, G., & Chattratichart, J. (2007). Usability Testing: What Have We Overlooked? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1415–1424. https://doi.org/10.1145/1240624.1240839

Liu, C., & Arnett, K. P. (2000). Exploring the factors associated with Web site success in the context of electronic commerce. *Information & Management*, *38*(1), 23–33.

Loiacono, E. T., Watson, R. T., & Goodhue, D. L. (2007). WebQual: An instrument for consumer evaluation of web sites. *International Journal of Electronic Commerce*, *11*(3), 51–87.

Lone, S., Harboul, N., & Weltevreden, J. (2021). *2021 European E-commerce Report*. Hogeschool van Amsterdam.

https://pure.hva.nl/ws/files/23594824/European_Ecommerce_Report_2021.pdf

Loranger, H. (2014). Accordions Are Not Always the Answer for Complex Content on Desktops. *Nielsen Norman Group*. https://www.nngroup.com/articles/accordions-complex-content/

Lowry, P., Karuga, G., & Richardson, V. (2007). Assessing leading institutions, faculty, and articles in premier information systems research journals. *Communications of the Association for Information Systems*, *20*, 142–203.

Lynch, P. J., & Horton, S. (2016). *Web Style Guide* (4th ed.). Yale University Press.

Madnick, S. E., Wang, R. Y., Lee, Y. W., & Zhu, H. (2009). Overview and framework for data and information quality research. *Journal of Data and Information Quality (JDIQ)*, *1*(1), 2:1-2:22.

Mardiana, S., Tjakraatmadja, J. H., & Aprianingsih, A. (2015). DeLone–McLean information system success model revisited: The separation of intention to use-use and the integration of technology acceptance models. *International Journal of Economics and Financial Issues*, *5*(1S), 172–182.

Markus, M. (1999). Thinking the Unthinkable: What happens if the IS field as we know it goes away? *Rethinking Management Information Systems*, 175–203.

Mårtensson, P., & Lee, A. S. (2004). Dialogical action research at omega corporation. *MIS Quarterly*, 507–536.

Mason, R. O. (1978). Measuring information output: A communication systems approach. *Information and Management*, 219–234.

Matera, M., Rizzo, F., & Carughi, G. T. (2006). Web usability: Principles and evaluation methods. In *Web engineering* (pp. 143–180). Springer.

Mathiassen, L., & Sandberg, A. (2013). How a professionally qualified doctoral student bridged the practice-research gap: A confessional account of Collaborative Practice Research. *European Journal of Information Systems*, *22*(4), 475–492.

Mawhinney, C. H., & Lederer, A. L. (1990). A study of personal computer utilization by managers. *Information & Management*, *18*(5), 243–253.

McCandless, M. (1996). Content distribution on the Web. *IEEE Expert*, *11*(5), 8–9.

McDonald, S., & Petrie, H. (2013). The effect of global instructions on think-aloud testing. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2941–2944.

McDonald, S., & Stevenson, R. J. (1996). Disorientation in hypertext: The effects of three text structures on navigation performance. *Applied Ergonomics*, *27*(1), 61–68.

McKay, J., & Marshall, P. (2001). The dual imperatives of action research. *Information Technology & People*, *14*(1), 46–59.

McLuhan, M. (1994). *Understanding media: The extensions of man*. MIT Press.

Mertens, D. M. (2010). Transformative mixed methods research. *Qualitative Inquiry*, *16*(6), 469–474.

Miller, H. (1996). The multiple dimensions of information quality. *Information Systems Management*, *13*(2), 79–82.

Miller, R. B. (1968). Response time in man-computer conversational transactions. *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, 267–277.

Mingers, J. (2001). Combining IS research methods: Towards a pluralist methodology. *Information Systems Research*, *12*(3), 240–259.

Mingers, J., & Brocklesby, J. (1997). Multimethodology: Towards a framework for mixing methodologies. *Omega*, *25*(5), 489–509.

Mingers, J. C. (1995). Information and meaning: Foundations for an intersubjective account. *Information Systems Journal*, *5*(4), 285–306.

Mingers, J., & Gill, A. (1997). *Multimethodology: Theory and practice of combining management science methodologies*. Wiley.

Miranda González, F., & Bañegil Palacios, T. (2004). Quantitative evaluation of commercial web sites: An empirical study of Spanish firms. *International Journal of Information Management*, *24*(4), 313–328.

Moody, D. L. (2000). Building links between IS research and professional practice: Improving the relevance and impact of IS research. *Proceedings of the Twenty First International Conference on Information Systems*, 351–360.

Moran, K. (2015a, June 28). *The Roots of Minimalism in Web Design*. Nielsen Norman Group. https://www.nngroup.com/articles/roots-minimalism-web-design/

Moran, K. (2015b, July 12). *The Characteristics of Minimalism in Web Design*. Nielsen Norman Group. https://www. nngroup. com/articles/characteristics-minimalism

Morris, M. G., & Dillon, A. (1997). How user perceptions influence software use. *IEEE Software*, *14*(4), 58–65.

Morris, Z. S., Wooding, S., & Grant, J. (2011). The answer is 17 years, what is the question: Understanding time lags in translational research. *Journal of the Royal Society of Medicine*, *104*(12), 510–520.

Morville, P., & Rosenfeld, L. (2006). *Information Architecture for the World Wide Web: Designing Large-Scale Web Sites* (3rd ed.). O'Reilly.

Müller, J., & Christandl, F. (2019). Content is king–But who is the king of kings? The effect of content marketing, sponsored content & user-generated content on brand responses. *Computers in Human Behavior*, *96*, 46–55.

Myers, M. D. (1997). Qualitative research in information systems. *Management Information Systems Quarterly*, *21*(2), 241–242.

Naumann, F., & Rolker, C. (2000). Assessment methods for information quality criteria. *Proceedings of the Fifth International Conference on Information Quality*.

Nelson, R. R. (2007). IT project management: Infamous failures, classic mistakes, and best practices. *MIS Quarterly Executive*, *6*(2).

Nelson, R. R., Todd, P. A., & Wixom, B. H. (2005). Antecedents of information and system quality: An empirical examination within the context of data warehousing. *Journal of Management Information Systems*, *21*(4), 199–235.

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A., & Nielsen, R. K. (2017). *Reuters Institute digital news report 2017*. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf

NHS Digital. (2021, December 31). *NHS App turns three with 22 million users*. https://digital.nhs.uk/news/2021/nhs-app-turns-three-with-22-million-users

Nielsen, J. (1993). *Usability engineering*. Morgan Kaufmann.

Nielsen, J. (1994). Estimating the number of subjects needed for a thinking aloud test. *International Journal of Human-Computer Studies*, *41*(3), 385–397.

Nielsen, J. (2001, May 12). *Search: Visible and Simple*. Nielsen Norman Group. https://www.nngroup.com/articles/search-visible-and-simple/

Nielsen, J. (2012). *Thinking aloud: The #1 usability tool*. Nielsen Norman Group. https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/

Nielsen Norman Group. (2014, January 12). *Turn User Goals into Task Scenarios for Usability Testing*. https://www.nngroup.com/articles/task-scenarios-usability-testing/

Nkanata, M. G. (2019). Applying DeLone and McLean information systems success model in the evaluation of e-government initiatives: A literature review. *Proceedings of 20th Annual IS Conference*, *18*, 287.

Nørgaard, M., & Hornbæk, K. (2006). What do usability evaluators do in practice?: An explorative study of think-aloud testing. *Proceedings of the 6th Conference on Designing Interactive Systems*, 209–218.

Norman, D. A. (1998). *The Invisible Computer*. MIT Press.

Nunnally, B., & Farkas, D. (2016). *UX research: Practical techniques for designing better products*. O'Reilly Media.

Nurse, J. R., Rahman, S. S., Creese, S., Goldsmith, M., & Lamberts, K. (2011). Information quality and trustworthiness: A topical state- of- the- art review. *The International Conference on Computer Applications and Network Security (ICCANS) 2011*, 492–500.

Oates, B. J. (2005). *Researching information systems and computing*. Sage.

OED. (2021). *Oxford English Dictionary*.

Ofcom. (2021). *Online Nation report*. https://www.ofcom.org.uk/__data/assets/pdf_file/0013/220414/online-nation-2021-report.pdf

Office of National Statistics. (2020, August 7). *Internet access – households and individuals, Great Britain: 2020*. https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2020

Olmsted-Hawala, E. L., Murphy, E. D., Hawala, S., & Ashenfelter, K. T. (2010). Think-aloud protocols: A comparison of three think-aloud protocols for use in testing data-dissemination web sites for usability. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2381–2390.

Olsina, L., Sassano, R., & Mich, L. (2009). Towards the quality for Web 2.0 applications. *8th International Workshop on Web Oriented Software Technology (IWWOST2009)*.

Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization Science*, *3*(3), 398–427.

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information Systems Research*, *2*(1), 1–28.

Ou, C. X., & Sia, C. L. (2010). Consumer trust and distrust: An issue of website design. *International Journal of Human-Computer Studies*, *68*(12), 913–934.

Palvia, S. C. J., & Sharma, S. S. (2007). E-government and e-governance: Definitions/domain framework and status around the world. *International Conference on E-Governance*, *5*, 1–12.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*, *12*(3), 801–823.

Pappas, I. O., Sharma, K., Mikalef, P., & Giannakos, M. N. (2018). A Comparison of Gaze Behavior of Experts and Novices to Explain Website Visual Appeal. *PACIS*.

Parker, M., Moleshe, V., De la Harpe, R., & Wills, G. (2006). *An evaluation of Information quality frameworks for the World Wide Web*.

Paul, T., & Vipond, D. (2019, June 4). *Adapting the GOV.UK Design System for the NHS*. Government Digital Service Blog. https://gds.blog.gov.uk/2019/06/04/guest-post-adapting-the-gov-uk-design-system-for-the-nhs/

Persson, H., Åhman, H., Yngling, A. A., & Gulliksen, J. (2015). Universal design, inclusive design, accessible design, design for all: Different concepts—One goal? On the concept of accessibility—Historical, methodological and philosophical aspects. *Universal Access in the Information Society*, *14*(4), 505–526.

Petrie, H., & Kheir, O. (2007). The relationship between accessibility and usability of websites. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 397–406.

Petrie, H., Savva, A., & Power, C. (2015). Towards a unified definition of web accessibility. *Proceedings of the 12th Web for All Conference*, 35.

Petter, S., DeLone, W. H., & McLean, E. R. (2008). Measuring information systems success: Models, dimensions, measures, and interrelationships. *European Journal of Information Systems*, *17*(3), 236–263.

Petter, S., DeLone, W., & McLean, E. R. (2012). The past, present, and future of" IS Success". *Journal of the Association for Information Systems*, *13*(5), 341–362.

Petter, S., DeLone, W., & McLean, E. R. (2013). Information systems success: The quest for the independent variables. *Journal of Management Information Systems*, *29*(4), 7–62.

Petter, S., & McLean, E. R. (2009). A meta-analytic assessment of the DeLone and McLean IS success model: An examination of IS success at the individual level. *Information & Management*, *46*(3), 159–166.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211–218.

Pitt, L. F., Watson, R. T., & Kavan, C. B. (1995). Service quality: A measure of information systems effectiveness. *MIS Quarterly*, 173–187.

Pokorny, J. (2004). Web searching and information retrieval. *Computing in Science & Engineering*, *6*(4), 43–48.

Power, C., Freire, A., Petrie, H., & Swallow, D. (2012). Guidelines are only half of the story: Accessibility problems encountered by blind users on the web. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 433–442.

Pradhan, S. (2005). Believability as an Information Quality Dimension. *ICIQ*.

Price, R., Neiger, D., & Shanks, G. (2008). Developing a measurement instrument for subjective aspects of information quality. *Communications of the Association for Information Systems*, *22*(1), 49–74.

Rafique, I., Lew, P., Abbasi, M. Q., & Li, Z. (2012). Information quality evaluation framework: Extending ISO 25012 data quality model. *International Journal of Computer and Information Engineering*, *6*(5), 568–573.

Rai, A., Lang, S. S., & Welker, R. B. (2002). Assessing the validity of IS success models: An empirical test and theoretical analysis. *Information Systems Research*, *13*(1), 50–69.

Rains, S. A., & Karmikel, C. D. (2009). Health information-seeking and perceptions of website credibility: Examining Web-use orientation,

message characteristics, and structural features of websites. *Computers in Human Behavior*, *25*(2), 544–553.

Ranpara, R., & Kumbharana, C. (2021). Contextual Information Retrieval Search Engine Challenges. In *Rising Threats in Expert Applications and Solutions* (pp. 269–273). Springer.

Resmini, A., & Rosati, L. (2012). A brief history of information architecture. *Journal of Information Architecture*, *3*(2), 1–12.

Reynolds, R. (2013, June 11). *How we've been improvig site search on GOV.UK*. GOV.UK Blog. https://gds.blog.gov.uk/2013/06/11/improving-site-search-on-gov-uk/

Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology*, *53*(2), 145–161.

Rieh, S. Y., & Belkin, N. J. (1998). Understanding judgment of information quality and cognitive authority in the WWW. *Proceedings of the 61st Annual Meeting of the American Society for Information Science*, *35*, 279–289.

Rippen, H., & Risk, A. (2000). E-Health code of ethics (May 24). *Journal of Medical Internet Research*, *2*(2), e9.

Robey, D. (1996). Research commentary: Diversity in information systems research: Threat, promise, and responsibility. *Information Systems Research*, *7*(4), 400–408.

Rogerson, S., Miller, K. W., Winter, J. S., & Larson, D. (2019). Information systems ethics–challenges and opportunities. *Journal of Information, Communication and Ethics in Society*, 87–97.

Rosemann, M., & Vessey, I. (2005). Linking theory and practice: Performing a reality check on a model of IS success. *ECIS 2005 Proceedings*, 1–12.

Rosemann, M., & Vessey, I. (2008). Toward improving the relevance of information systems research to practice: The role of applicability checks. *MIS Quarterly*, 1–22.

Rosenfeld, L., Morville, P., & Arango, J. (2015). *Information architecture* (4th ed.). O'Reilly Media.

Rosson, M. B., & Carroll, J. M. (2002). *Usability Engineering: Scenario-Based Development of Human-Computer Interaction*. Elsevier Science.

Ruzza, M., Tiozzo, B., Mantovani, C., D'Este, F., & Ravarotto, L. (2017). Designing the information architecture of a complex website: A

strategy based on news content and faceted classification. *International Journal of Information Management*, *37*(3), 166–176.

Sabherwal, R., Jeyaraj, A., & Chowa, C. (2006). Information system success: Individual and organizational determinants. *Management Science*, *52*(12), 1849–1864.

Sadiq, S., Yeganeh, N. K., & Indulska, M. (2011). 20 years of data quality research: Themes, trends and synergies. *Proceedings of the Twenty-Second Australasian Database Conference-Volume 115*, 153–162.

Saltzer, J. H., & Schroeder, M. D. (1975). The protection of information in computer systems. *Proceedings of the IEEE*, *63*(9), 1278–1308.

Sauer, C. (1993). *Why information systems fail: A case study approach*. Alfred Waller Ltd., Publishers.

Saunders, M. N., Lewis, P., & Thornhill, A. (2003). *Research methods for business students* (3rd ed.). Prentice Hall.

Sauro, J. (2016). *The origins and evolution of thinking aloud*. MeasuringU. https://measuringu.com/thinking-aloud/

Schaupp, L. C., Fan, W., & Belanger, F. (2006). Determining success for different website goals. *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, 6, 107b–107b.

Schmutz, P., Heinz, S., Métrailler, Y., & Opwis, K. (2009). Cognitive load in ecommerce applications: Measurement and effects on user satisfaction. *Advances in Human-Computer Interaction*, 1–9.

Seffah, A., & Metzker, E. (2004). The obstacles and myths of usability and software engineering. *Communications of the ACM*, *47*(12), 71–76.

Seymour, T., Frantsvog, D., Kumar, S., & others. (2011). History of search engines. *International Journal of Management & Information Systems (IJMIS)*, *15*(4), 47–58.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423.

Shannon, C. E., & Weaver, W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press.

Sherer, S. A., Kohli, R., & Baron, A. (2003). Complementary investment in change management and IT investment payoff. *Information Systems Frontiers*, *5*(3), 321–333.

Shneiderman, B. (2000). Universal usability. *Communications of the ACM*, *43*(5), 84–91.

Silberg, W. M., Lundberg, G. D., & Musacchio, R. A. (1997). Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor—Let the reader and viewer beware. *Jama*, *277*(15), 1244–1245.

Sillence, E., Briggs, P., Fishwick, L., & Harris, P. (2004). Trust and mistrust of online health sites. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 663–670.

Simons, H. D. (1971). Reading comprehension: The need for a new perspective. *Reading Research Quarterly*, 338–363.

Simonsen, J. (2009). A Concern for Engaged Scholarship: The challenges for action research projects. *Scandinavian Journal of Information Systems*, *21*(2), 1.

Siricharoen, W. V. (2013). Infographics: The new communication tools in digital age. *The International Conference on E-Technologies and Business on the Web (Ebw2013)*, 169–174.

Six, J. M., & Macefield, R. (2016, January 4). *How to Determine the Right Number of Participants for Usability Studies*. UX Matters.

Slatin, J. M., & Rush, S. (2002). *Maximum accessibility: Making your web site more usable for everyone*. Addison-Wesley Longman Publishing Co., Inc.

Sloan, D., & Kelly, B. (2011). Web accessibility metrics for a post digital world. *RDWG Symposium on Website Accessibility Metrics*.

Stallman, R. M. (1996). *Words to Avoid (or Use with Care) Because They Are Loaded or Confusing*.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). 10 potholes in the road to information quality. *Computer*, *30*(8), 38–46.

Stvilia, B., Twidale, M. B., Gasser, L., & Smith, L. C. (2005). Information quality discussions in Wikipedia. *Proceedings of the 2005 International Conference on Knowledge Management*, 101–113.

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital Media, Youth, and Credibility* (pp. 73–100). The MIT Press.

Susman, G. I., & Evered, R. D. (1978). An assessment of the scientific merits of action research. *Administrative Science Quarterly*, 582–603.

Swanson, E. B. (1982). Measuring user attitudes in MIS research: A review. *Omega*, *10*(2), 157–165.

Swanson, E. B. (1997). Maintaining IS quality. *Information and Software Technology*, *39*(12), 845–850.

Taddeo, M. (2010). An Information-based Solution for the Puzzle of Testimony and Trust. *Social Epistemology*, *24*(4), 285–299.

Taylor, R. S. (1986). *Value-added processes in information systems*. Greenwood Publishing Group.

TeBlunthuis, N., Bayer, T., & Vasileva, O. (2019). Dwelling on Wikipedia. *Reading Time*, *2019*. https://upload.wikimedia.org/wikipedia/commons/f/f5/Dwelling_on_Wikipedia_slides_from_Opensym_2019.pdf

Teo, T. S., Lim, V. K., & Lai, R. Y. (1999). Intrinsic and extrinsic motivation in Internet usage. *Omega*, *27*(1), 25–37.

Termens, M., Ribera, M., Porras, M., Boldú, M., Sulé, A., & Paris, P. (2009). Web content accessibility guidelines: From 1.0 to 2.0. *Proceedings of the 18th International Conference on World Wide Web*, 1171–1172.

Thatcher, J., Waddell, C. D., Henry, S. L., Swierenga, S., Urban, M. D., Burks, M., Regan, B., & Bohman, P. (2003). *Constructing accessible web sites*. Apress.

Thielsch, M. T., & Hirschfeld, G. (2019). Facets of website content. *Human-Computer Interaction*, *34*(4), 279–327.

Thielsch, M. T., & Moshagen, M. (2015, April 7). *VisAWI Manual*. http://www.thielsch.org/download/VisAWI/VisAWI_Manual_EN.pdf

Thielsch, M. T., Thielsch, C., & Hirschfeld, G. (2019). How informative is informative? Benchmarks and optimal cut points for E-Health Websites. *Mensch Und Computer 2019-Workshopband*.

Thorlacius, L. (2007). The Role of Aesthetics in Web Design. *Nordicom Review*, *28*(1).

Tibau, M., Siqueira, S. W., & Nunes, B. P. (2021). Think-Aloud Exploratory Search: Understanding Search Behaviors and Knowledge Flows. *Research and Innovation Forum 2020: Disruptive Technologies in Times of Change*, 303–315.

Tullis, T., Fleischman, S., McNulty, M., Cianchette, C., & Bergel, M. (2002). An empirical comparison of lab and remote usability testing of web sites. *Usability Professionals Association Conference*.

UK Parliament. (2018). *The Public Sector Bodies (Websites and Mobile Applications) (No. 2) Accessibility Regulations 2018*. http://www.legislation.gov.uk/uksi/2018/952/contents/made

Urbach, N., Smolnik, S., & Riempp, G. (2009). The state of research on information systems success. *Business & Information Systems Engineering*, *1*(4), 315–325.

Usability.gov. (2019). *Running a Usability Test*. https://www.usability.gov/how-to-and-tools/methods/running-usability-tests.html

Van de Ven, A. H. (2007). *Engaged scholarship: A guide for organizational and social research*. Oxford University Press on Demand.

van den Haak, M. J., de Jong, M. D. T., & Schellens, P. J. (2003). Retrospective vs. Concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour & Information Technology*, *22*(5), 339–351.

van Velsen, L., van der Geest, T., & Klaassen, R. (2007). Testing the usability of a personalized system: Comparing the use of interviews, questionnaires and thinking-aloud. *2007 IEEE International Professional Communication Conference*, 1–8.

van Velsen, L., van der Geest, T., & Klaassen, R. (2011). Identifying Usability Issues for Personalization During Formative Evaluations: A Comparison of Three Methods. *International Journal of Human-Computer Interaction*, *27*(7), 670–698.

Vanderheiden, G. C. (1995, January 31). *Design of HTML (Mosaic) Pages to Increase their Accessibility to Users with Disabilities Strategies for Today and Tomorrow*. http://trace.wisc.edu/archive/html_guidelines/version1.html

Varela, M., Mäki, T., Skorin-Kapov, L., & Hoßfeld, T. (2013). Towards an understanding of visual appeal in website design. *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, *70–75*.

Vaziri, R., & Mohsenzadeh, M. (2012). A questionnaire-based data quality methodology. *International Journal of Database Management Systems*, *4*(2), 55.

Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences*, *39*(2), 273–315.

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly*, *37*(1).

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46*(2), 186–204.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 425–478.

Vidgen, R., & Braa, K. (1997). Balancing interpretation and intervention in information system research: The action case approach. In *Information systems and qualitative research* (pp. 524–541). Springer.

von Neumann, J. (1945). *First Draft of a Report on the EDVAC*. University of Pennsylvania.

Von Neumann, J. (1993). First Draft of a Report on the EDVAC. *IEEE Annals of the History of Computing*, *15*(4), 27–75.

W3C. (1999). *Web content accessibility guidelines (WCAG) 1.0*. https://www.w3.org/TR/WAI-WEBCONTENT/

W3C. (2005). *Web Accessibility Initiative: Introduction to Web Accessibility*. https://www.w3.org/WAI/fundamentals/accessibility-intro/

W3C. (2008). *Web content accessibility guidelines (WCAG) 2.0*. https://www.w3.org/TR/WCAG20/

W3C. (2016a). *HTML & CSS*. W3C. https://www.w3.org/standards/webdesign/htmlcss

W3C. (2016b). *Understanding Conformance*. Understanding WCAG 2.0. https://www.w3.org/TR/UNDERSTANDING-WCAG20/conformance.html

W3C. (2018a, June 5). *Web Content Accessibility Guidelines (WCAG) 2.1*. https://www.w3.org/TR/WCAG21/

W3C. (2018b). The Business Case for Digital Accessibility. *Web Accessibility Initiative*. https://www.w3.org/WAI/business-case/

W3C. (2021, January 14). *W3C Accessibility Guidelines (WCAG) 3.0*. W3C. https://w3c.github.io/silver/guidelines/

W3Techs. (2021). *Historical yearly trends in the usage statistics of content management systems*. Web Technology Surveys. https://w3techs.com/technologies/history_overview/content_management/all/y

Walker, S., Black, A., & Carey, C. (2013). *GOV.UK content principles: Conventions and research background*.

https://www.gov.uk/government/publications/govuk-content-principles-conventions-and-research-background

Walsham, G. (1995). The emergence of interpretivism in IS research. *Information Systems Research*, *6*(4), 376–394.

Walsham, G. (2006). Doing interpretive research. *European Journal of Information Systems*, *15*(3), 320–330.

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, *39*(11), 86–95.

Wang, R. Y. (1998). A product perspective on total data quality management. *Communications of the ACM*, *41*(2), 58–65.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5–33.

Watson, R. (2001). Research in Information Systems: What We Haven't Learned. *MIS Quarterly*, *25*(4), v–xv.

Weber, R. (1997). *Ontological foundations of information systems*. Coopers & Lybrand and the Accounting Association of Australia and New Zealand Melbourne.

Weber, R. (2003). Editor's Comments. *MIS Quarterly*, *27*(2), iii–xi.

Weinberger, D. (2007). *Everything is miscellaneous: The power of the new digital disorder*. Macmillan.

Wertheimer, M. (1924). *Ueber Gestalttheorie, Lecture Before the Kant Gesellschaft*.

Westerman, D., Spence, P. R., & van der Heide, B. (2014). Social Media as Information Source: Recency of Updates and Credibility of Information. *Journal of Computer-Mediated Communication*, *19*, 171–183.

Whitenton, K. (2013, November 10). *Flat vs. Deep Website Hierarchies*. Nielsen Norman Group. https://www.nngroup.com/articles/flat-vs-deep-hierarchy/

Wikimedia. (2022, January 22). *Wikimedia Statistics*. https://stats.wikimedia.org/#/en.wikipedia.org/reading/unique-devices/normal|line|2-year|(access-site)~mobile-site*desktop-site|monthly

Williams, M. D., Rana, N. P., Dwivedi, Y. K., & Lal, B. (2011). Is UTAUT really used or just cited for the sake of it? A systematic review of citations of UTAUT's originating article. *ECIS*, 231.

Wilson, P. (1983). *Second-hand knowledge: An inquiry into cognitive authority*. Præger.

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation*, *55*(3), 249–270.

Winter, R., Griffiths, M., & Green, K. (2000). The'academic'qualities of practice: What are the criteria for a practice-based PhD? *Studies in Higher Education*, *25*(1), 25–37.

Wolfinbarger, M., & Gilly, M. C. (2003). ETailQ: dimensionalizing, measuring and predicting etail quality. *Journal of Retailing*, *79*(3), 183–198.

World Wide Web Foundation. (2020). *History of the Web*. https://webfoundation.org/about/vision/history-of-the-web/

Wu, J.-H., & Wang, Y.-M. (2006). Measuring KMS success: A respecification of the DeLone and McLean's model. *Information & Management*, *43*(6), 728–739.

Wyrwoll, C. (2014). User-Generated Content. In *Social Media* (pp. 11–45). Springer.

Yablonski, J. (2020). *Laws of UX: Using Psychology to Design Better Products & Services*. O'Reilly Media.

Yesilada, Y., Brajnik, G., Vigo, M., & Harper, S. (2012). Understanding web accessibility and its drivers. *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*, 19.

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Sage publications.

Yoo, B., & Donthu, N. (2001). Developing a scale to measure the perceived quality of an Internet shopping site (SITEQUAL). *Quarterly Journal of Electronic Commerce*, *2*(1), 31–45.

Zmud, R. W. (1978). An empirical investigation of the dimensionality of the concept of information. *Decision Sciences*, *9*(2), 187–195.

Zúñiga, G. L. (2001). Ontology: Its transformation from philosophy to information systems. *Proceedings of the International Conference on Formal Ontology in Information Systems*, 187–197.

Appendix A. Informative Web Content Guidelines

The Informative Web Content Guidelines (IWCG) documented below form advice for the analysis and improvement of online web content. The guidance follows the same structure as the upcoming WCAG 3.0 guidelines, providing a way to analyse your web content for informativeness. By ranking each of sixteen key criteria, areas for suggested improvements are discovered, providing evidence-based means for content enhancements without the need to conduct user studies.

Context and domains of use

The guidance contained within the IWCG has been produced to target informative, textual website content. Typical examples of such websites include corporate knowledge bases, medical guides, and informational articles across any domain. Advice produced as a result of applying the guidelines will help you to improve your website's content by providing suggestions for improvements.

How to apply these guidelines

These guidelines provide a structured means to assess your web content and generate areas for improvement. They can be applied by a practitioner without knowledge of the underlying theory, with suggestions generated via the rating mechanism. Note that although ratings are generated for each of the guidelines and criteria, the overall sum of these is not intended to be used as an aggregate score to compare websites; rather they feature as indicators for which areas to focus content improvement efforts.

To apply the guidelines, work through each in turn, analysing web page content for the listed ratings. For each area, rank your own content based on the properties described, remembering that these should be applied to a cross-section of the entire site rather than individual pages as with WCAG.

Not all ratings are used for every area; this is standard practice for the WCAG 3.0 model.

Once there is a score for each area, average the scores in each guideline (containing three, four or five areas each). These will give an indication of which areas to focus on to deliver the most improvement to content. Starting with the lowest scoring guideline, work through the areas applying the guidance to improve the criteria scores.

Each area within the guidelines also contains a critical error path. This is a condition that when satisfied identifies that urgent action is required to improve the web content informativeness due to the lack of appropriate functionality that may cause serious issues within a typical user's experience of the website.

The IWCG can be applied alongside the WCAG guidelines for accessibility, though the focus is different: WCAG identifies system aspects at a page-by-page level, whilst IWCG applies across the entire site. This model is most useful where a consistent design style or pattern library has been used as a basis for all content, providing an easy mechanism to improve individual aspects of a website.

A.1. Guideline 1. Availability

Availability of information has been identified as a key aspect of information quality success within both prior literature and the case studies. Availability of information covers three key areas:

A.1.1. Alternative formats

Guideline

Provide alternative formats for users, e.g., download of information, links to apps and telephone contact as appropriate.

Description

This success criterion is concerned with the provision of information in alternative formats for the user. For example, during the think aloud studies, participants often wanted to access information in other ways, such as via phone, app, or download. Making the user aware of alternative information formats and how to find such materials would facilitate ease of use for many users.

Information quality attributes

Availability, information usability

Critical errors

Information is only presented in a single format with no opportunities for users to clarify questions.

Rating for 'Alternative formats'

Rating	**Criteria**
Rating 0	No alternative formats or contact information, or a critical error
Rating 1	Not applicable
Rating 2	Some alternative formats though these may be challenging to find, with unclear signposting.
Rating 3	Not applicable
Rating 4	Alternative information sources signposted throughout the site in a consistent manner.

Derivation

Providing information in alternative formats is an availability issue and occurred throughout the three case studies. In the University case study, one participant discussed the lack of assurance that fallback options were

available (section 8.5.3). This was followed in case study 2 where the lack of direct contact information such as an email address to discuss follow-up questions led to a participant raising the issue of questions not addressed by the informative site content (section 9.5.3).

These two case studies were followed by the health website, where the prominence of alternative contact information was noted by two participants (section 10.5.3). Rather than commenting on the lack of alternative channels, this case study provided an example where other information sources are made prominent using contextual contact details including opening times and references to other websites. This was noted as particularly helpful with non-urgent queries that could not be answered directly by the informative aspects of the site.

The fallback information quality attribute is not addressed within the prior literature, and this could be caused by two issues. Firstly, the majority of IQ frameworks within the field are based on the model provided by Wang & Strong (1996). This seminal work was authored before the widespread emergence of the Internet as an informative data source and focused on *data* rather than *information*. The distinction between these terms is provided by Ackoff (1999, pp. 170–172) and (Checkland & Poulter, 2006, pp. 112–113): data is *unprocessed symbols*, whereas information has been *processed to create meaningful facts for an intended purpose*. This distinction is highlighted within informative websites, where their purpose is to transfer information to the user, not data. Creating appropriate web page content requires an understanding of the user, external systems, and the way it may be interpreted, leading to potential gaps in the information provided.

The third case study differs from the first two as it has been created using a development framework by a large public sector organisation known for conducting in-depth usability studies. It is likely that these studies have uncovered such gaps, with the alternative means of contact provided as ways to satisfy unexpected user needs using existing resources. Smaller

organisations may not be able to provide such a variety of alternative contact mechanisms, however web sites should always provide at least one clear means for asking follow-up questions.

The second potential reason for *fallback* not being included as an attribute in prior studies is the focus of the models. Prominent models within the Information Systems field such as DeLone-McLean and TAM assume that all information is contained within the system and is accessible by the user. Web systems require a different level of understanding as pages may be accessed by novice and expert users alike with unknown goals, expectations, and prior domain knowledge. Not all questions or stages in a process can necessarily be addressed by the content authors, therefore a fallback channel when the information cannot be relied upon to contain the full extent of the organisation's knowledge base in a manner accessible to the user is required.

Conducting the first three action case studies has highlighted the need to include fallback as a discrete attribute separately from availability as a facet of designing web-based content rather than generic information systems.

A.1.2. Contextual navigation (stateless)

Guideline

Avoid non-stateless navigation.

Description

Ensuring that user navigation is stateless prevents assumptions being made by previous pages visited. Users rarely follow a single track through a website and are accustomed to design patterns that facilitate access to information through multiple means. By providing stateless content, users will not be stuck in loops, subsections or contextual areas that do not facilitate navigation to other areas.

Information quality attributes

Consistent representation, navigation

Critical errors

Navigation changes dependant on page states.

Rating for 'Contextual navigation (stateless)'

Rating	**Criteria**
Rating 0	Navigation of entire system relies on consistent states
Rating 1	Not applicable
Rating 2	Some stateless navigation with clear links to break states
Rating 3	Not applicable
Rating 4	Navigation fully stateless

Derivation

The impact of stateless navigation became apparent during the first case study, where confusion was created when pages changed based on prior selections. Whilst this can be an expected path for multi-step and wizard navigation patterns, when accessing informative content this results in an inconsistent experience not matching cognitive expectations.

At a view-based level, the WCAG guidelines define state as representing 'data associated with the component or user interaction possibilities' (W3C, 2018a) thus state-based changes should only be made to further represent data in a format accessible to the user. A parallel consideration can be made for state within the entire website: although not covered by WCAG, entire views should not change unless required to do so by the information represented. This promotes a level of consistency throughout several interconnected screens allowing users to navigate between sections of

comparable information on related pages without requiring any further consideration than the distinct and deliberate changes between such screens.

A.1.3. Link disambiguation (styling)

Guideline

Ensure link styling is clear and consistent across the site.

Description

Whilst the WCAG 2.1 guidelines already contain success criterion 2.4.9 stating that link purpose must be identifiable from text alone (W3C, 2018a), it does not contain any guidance on the specific styling of links. CSS styling provides the opportunity to radically alter the appearance of links, changing their context to users. ARIA roles can be used to identify any element as a link, however they limit the browser functionality for users (Coyier, 2020) and thus their access to information.

Information quality attributes

Consistent representation, navigation

Critical errors

Link styles inconsistent across site without clear purpose.

Rating for 'Link disambiguation (styling)'

Rating	**Criteria**
Rating 0	Links are styled inconsistently
Rating 1	Not applicable
Rating 2	Links are mostly styled consistently to annotate their purpose
Rating 3	Not applicable
Rating 4	All links follow a common design pattern with their purpose clearly identified

Derivation

This outcome has a similar derivation to stateless and contextual navigation represented by Contextual navigation (stateless). To provide a consistent experience that does not require users to re-learn interface and navigation elements. The WCAG guidelines covers some elements of link styling within §2.4 Navigable (W3C, 2018a), however these are at a mechanical link-by-link level rather than assessing the consistency across all links within a page or site.

Link disambiguation as a manifestation of consistent representation occurred within all three case studies (see §§8.5.2, 9.5.2 and 10.5.2), highlighting the importance of styling within users access to informative content. Whilst styling could be considered as a purely navigation impact, the ongoing user experience issues encountered during the think aloud studies demonstrated the ongoing repercussions caused by inconsistent representation of information structures across otherwise consistent screens. By changing the representation of the information on subsequent pages, users needed to change their understanding of the web content, thus requiring different access patterns to retrieve otherwise similar concepts.

This outcome introduces consistent representation among links as an important factor in web content design, enabling access to further specific areas as expected by the user's navigational patterns.

A.2. Guideline 2. Information architecture

The WCAG 3.0 guidelines focus on the use of specific elements within a specific page rather than taking a holistic overview of the entire site. This guideline introduces success criteria to be applied to a website and ensure that the information architecture is defined and accessible to the user with a consistent approach.

A.2.1. Alternative routes

Guideline

Provide natural alternative routes to information to account for different user models, promoting cross-navigation.

Description

Traditional website hierarchies follow a top-down approach, with a homepage leading to categories, sub-categories and finally content pages (Lynch & Horton, 2016). Whilst this may match the user model for some websites, often non-expert users do not have the same understanding of the information represented within a system. By facilitating alternative routes to information, users will be able to follow natural thematic links to explore the content.

Information quality attributes

Availability, fallback

Critical errors

No contextual links within content between pages or sections within the same page.

Rating for 'Alternative routes'

Rating	**Criteria**
Rating 0	Information follows a structured linear format
Rating 1	Not applicable
Rating 2	Some referenced links though inconsistent across site
Rating 3	Not applicable
Rating 4	Relevant links are referenced within content and pages

Derivation

At the beginning of the World Wide Web, pages were designed with a hierarchical approach due to studies such as McDonald & Stevenson (1996). This study determined that hierarchical structures provided a convenient balance between linear processes (such as those already understood by the user) and non-linear processes that required larger cognitive load. As the concept of an informative web has developed further, Information Architecture (IA) has emerged as a sub-field to promote the effective study and implementation of improved navigational patterns.

Information Architecture not only defines the navigational structures: it also includes the patterns used to deliver concepts to users. Rosenfeld et al. (2015, Chapter 4) determines that information structures should match the underlying real-world architecture that users are accustomed to: this promotes positive information patterns and provides appropriate structures that match user behaviour.

This outcome extends such patterns to suggest that promoting cross-links between different areas of content creates more opportunities for users to find their desired information more quickly and in a format appropriate for their search.

A.2.2. Consistent representation

Guideline

Maintain consistent information representation across a site and content areas.

Description

Whilst the WCAG 2.1 guidelines define multiple success criterion describing how not to define the appearance of various elements, there is no guideline suggesting consistent representation across a website. By applying consistent representations across all pages and elements, the user can explore a familiar environment, helping develop their mental model of the information and improving learnability of the system.

Information quality attributes

Consistent representation, presentation

Critical errors

Information presented inconsistently across pages.

Rating for 'Link disambiguation (styling)'

Rating	**Criteria**
Rating 0	Website designed without core components and style guide
Rating 1	Some consistency within site though no core components
Rating 2	Website navigation consistent though content elements appear differently on different pages
Rating 3	Navigation elements consistent with most content included in standardised formats
Rating 4	All website styles consistent, including navigational and in-page content areas

Derivation

Ensuring that information elements are consistently presented in the same manner helps users to navigate the site and its inherent content. Whilst the WCAG guidelines provide some guidance for consistency of call to actions, further requirements are required to maintain access to informative elements.

One key area where this needs to be considered is when the user is navigating between various pages of the same site or a family of linked sites, as per the first case study. Whilst there are often operational requirements for different navigational structures, there is a greater need for consistency of representation in informational cues. This requirement for uniformity across web presence is being addressed by organisations such as the Government Digital Service by the production of frameworks of navigation and information elements promoting positive user experiences.

A.2.3. Number of links

Guideline

Avoid providing users with large lists of links.

Description

The number of links presented on a page can have an impact on a user's attention and information seeking behaviour. By ensuring that the user is not overloaded with irrelevant links, it can be easier for the user to follow links and find website content.

Information quality attributes

Navigation, presentation

Critical errors

Large sets of links presented as lists.

Rating for 'Number of links'

Rating	Criteria
Rating 0	Huge number of links presented on page without content or organisation
Rating 1	Not applicable
Rating 2	Many links on page with contextualisation such as sub-headings and styling devices to create distinctions
Rating 3	Not applicable
Rating 4	Links only used where appropriate and embedded within text where possible

Derivation

The number of links presented to users can often cause anxiety and confusion when navigating information. This can be the case in both sites with content spread across multiple pages, and within single long pages.

Areas that may be particularly affected by large numbers of links include page footers and portals, where the use of information architecture devices such as sub-headings and styling devices can be used to provide additional structure. Ideally content should take the user on a journey through the information, starting with basic concepts building up to detailed information that may be of less importance to the whole audience.

A.2.4. Recency

Guideline

Ensure content is recent, and this is communicated with the user.

Description

More recent web content often aids information seeking behaviour and influences user choices (Westerman et al., 2014). By annotating information with dates of publication and updates, users may find it easier to access more relevant information.

Information quality attributes

Recency

Critical errors

No recency notifiers, e.g., date last updated, or date posted within content.

Rating for 'Recency'

Rating	**Criteria**
Rating 0	No recency notifier, e.g., uploaded, or updated date
Rating 1	Not applicable
Rating 2	Recency notifier included though content not updated in timely manner
Rating 3	Not applicable
Rating 4	Recency notifier included along with recent relevant updates to page

Derivation

As demonstrated during the think aloud studies, users place a reliance on up-to-date content being provided. Where there is no indicator of when content was authored, users may assume that it is no longer applicable or out of date, leading to a distrust of the information and behaviours to seek information elsewhere. This can be mitigated by providing the date when content was last revised, though note this can generate issues for sites with

infrequent content updates. If the user does not see any content that has been provided recently (for their notion of recent), then this may also have a negative impact on their information seeking behaviour.

A.2.5. Search results

Guideline

Provide clear and consistent search mechanisms for all content.

Description

Users often perform searches as a primary navigation mechanism of website content, however there is little guidance or standardisation on what should be searched and how the results should be displayed. Search mechanisms should clearly label what information is being searched, cover all website content in a single result set and display matches in a consistent manner ranked according to relevance.

Information quality attributes

Consistent representation, search

Critical errors

Site does not have an internal search mechanism.

Rating for 'Search results'

Rating	**Criteria**
Rating 0	No search mechanism
Rating 1	Basic search mechanism for page titles or metadata only
Rating 2	Search mechanism includes full content search
Rating 3	Full content search along with relevance filter
Rating 4	Full content search ranked by relevance, along with clear identification of most appropriate results for query

Derivation

Many content authors focus on the performance of their content through Search Engine Optimisation (SEO) and other means to drive traffic to their site, though internal search also requires consideration. When faced with large amounts of materials and/or links, users may rely on internal search features to help find the information they are seeking. Providing an appropriate full content search with a relevance ranking will help users to find the information they need. Where common synonyms or alternative nomenclature exists, the content author should ensure these are accounted for in both search queries and results.

A.3. Guideline 3. Interactivity

The WCAG 2.1 and 3.0 guidelines focus on accessibility of basic elements rather than the design and stylistic choices made by developers. The following four success criteria define a basis for progressive enhancement, maintaining existing functionality whilst adding additional features for website content navigation.

A.3.1. Consistent interactions

Guideline

Maintain consistent interaction patterns across entire site and content components.

Description

Where interactivity is provided, this should be maintained in a consistent manner to aid users navigating multiple pages and content areas. Inconsistent interactions can impair access to information.

Information quality attributes

Consistent representation, interactivity, navigation, presentation

Critical errors

Variety of interaction methods applied inconsistently across page.

Rating for 'Consistent interactions'

Rating	**Criteria**
Rating 0	Interaction methods vary within same page
Rating 1	Interaction methods vary across site
Rating 2	Standardised interactive elements hide key information from user
Rating 3	Interactive elements clearly identified with appropriate call-to-actions
Rating 4	Interactive elements provide affordance with reactive adaptation to user browsing preferences

Derivation

Whilst interactive elements can add interest to web sites and provide various means for users to interact with content, they can distract the user from finding the information they are seeking. Where interactive elements are provided, they should be designed to be consistent and intuitive to use, with standard icons and nomenclature to provide visual cues to the user. Structures which adapt to the user's browsing behaviour help enhance the experience rather than restrict the availability of informative content.

A.3.2. Expected layout

Guideline

Maintain layout consistency across pages and content areas.

Description

Maintaining consistent layouts across pages and content areas can aid user familiarity and facilitate easy access to information.

Information quality attributes

Interactivity, navigation, non-volatility, presentation

Critical errors

Content layout inconsistent between pages or sections of the same page.

Rating for 'Expected layout'

Rating	**Criteria**
Rating 0	No consistency of layout
Rating 1	Not applicable
Rating 2	Some consistent layout elements; however, these vary across pages
Rating 3	Not applicable
Rating 4	Fully consistent page layouts with easy access to information

Derivation

When navigating a series of pages within the same site, the layout of each page should behave in a uniform manner. This prevents users from having to re-learn or re-identify key features to find the information they are seeking. This was strongly evidenced in the first case study, where users accessed information across two different designs of website with opposing structures, creating a disjuncture in the browsing experiences.

A.3.3. Progressive enhancement

Guideline

Provide information as simply as possible with enhancements to build more complex pages.

Description

It is widely understood that web features should be designed and developed from the simplest to the most complex to aid compatibility and accessibility. This can be expanded to include content: content authors can start with short, simple snippets of information and build up to more complex pages. Taking this approach will aid the proves of writing content, as an outline can be provided initially and expanded upon once user needs are understood. Providing initial basic text will ensure that the content users seek is available in some format prior to the full provision of information.

Information quality attributes

Availability, fallback

Critical errors

Content not presented clearly with a reliance on complex formats or interactive features.

Rating for 'Progressive enhancement'

Rating	Criteria
Rating 0	Complex features with content presented in difficult to read formats
Rating 1	Not applicable
Rating 2	Some complex features within content
Rating 3	Not applicable
Rating 4	No complex features within site

Derivation

Progressive enhancement is the idea that simplest content features should be designed and built first, followed by those that are more complex. By starting with simpler ideas, the basics can be constructed to provide information without distractions of interactive or other presentational features.

When developing content, ensure that the information is presented primarily without complex features such as accordions or other content panels that may obscure the user seeking behaviour. Only then add the structures that can provide additional linkages and interest to the page, maintaining the inherent structure that users can easily find the information they are looking for.

A.3.4. Visibility

Guideline

Avoid hiding information within on-page navigational structures.

Description

During the initial case studies, there were occasions when the information users were seeking was available however it was hidden by the

enhancements provided by the designers to aid navigation. Hiding information on a page will impair access to the content, therefore developers should avoid devices such as accordions which require user interactions to access relevant information.

Information quality attributes

Consistent representation, navigation, visibility

Critical errors

Not all information visible on the page by default.

Rating for 'Visibility'

Rating	**Criteria**
Rating 0	Information hidden on page without clear identification to the user
Rating 1	Not applicable
Rating 2	Some information hidden on page with appropriate call-to-actions
Rating 3	Not applicable
Rating 4	No information hidden on page

Derivation

Throughout the three case studies, there were several occasions where users navigated to the correct areas of a page to find information yet were unable to see the answers they were seeking due to content obscurity. Ensuring that the main content is visible on the page before adding interactive features and signposting clearly the areas where content may be hidden, ensures it can still be accessed by users.

A.4. Guideline 4. Language

Guideline 3.1 of WCAG 2.1 (W3C, 2018a) already defines mechanised techniques for readable website content such as unusual words, reading level and provision of pronunciation information. This does not cover all of the language-related criteria that may apply to website content. The addition of Guideline 5.4 is intended to introduce success criteria related to the information quality attributes of an overall site, including content on multiple pages.

A.4.1. Ambiguous terms defined within text

Guideline

Define ambiguous terms within text using clear language.

Description

Whilst Success Criterion 3.1.3 already requires the definition of words used in an unusual way such as idiomatic or colloquial language, this criterion extends these requirements to include any ambiguous term used within text. For example, there was confusion during the first case study between *essays*, *assignments* and *coursework*. Explanations of how these terms are used should be provided contextually within the text so that user unfamiliar with the specific language can still understand the purpose and meaning within the content.

Information quality attributes

Language, understandability

Critical errors

Many terms used interchangeably with no in-context definitions.

Rating for 'Ambiguous terms defined within text'

Rating	Criteria
Rating 0	Ambiguous terms used interchangeably with no definitions
Rating 1	Not applicable
Rating 2	Ambiguous terms used though described through annotations and/or glossary
Rating 3	Not applicable
Rating 4	No ambiguous terms used

Derivation

With the case studies, participants encountered new or unfamiliar terms that were not defined within the text and a general audience could not have been expected to understand. Providing a glossary and/or in-place annotations is one work around for this issue where similar terms are often used interchangeably, though ideally content authors should ensure consistency across the entire site.

A.4.2. Clear use of language

Guideline

Write content using plain language and universal grammar.

Description

Writing content at the appropriate level of detail can be challenging, particularly where there are multiple audiences of expert and non-expert users. The UK Government Digital Service (2016) provide guidance that writing in plain English can benefit both those with high and low literacy, with 80% of experts in the legal domain preferring clear sentences to those written at an

expert level. Ensuring content is written plainly will increase accessibility of information for those with and without domain knowledge.

Information quality attributes

Interpretability, language, understandability

Critical errors

Many complex words within content.

Rating for 'Clear use of language'

Rating	**Criteria**
Rating 0	Confusing language used with complex words
Rating 1	Not applicable
Rating 2	Language generally accessible to intended audience though some complex concepts
Rating 3	Not applicable
Rating 4	All content explained in simple, plain language

Derivation

Content authors have long used automated metrics to assess the readability or reading age of text, for example the Flesch Reading Ease Score (1948). Whilst this can provide a measure of the educational attainment level required to understand the content, there are simpler mechanisms that can be used to make information accessible within the context. Content that is aimed at general audiences should have less confusing language or complex words, relying on plainer descriptions of key concepts. This widens the availability of the information to a wider number of users helping them seek answers to their own questions online.

A.4.3. Consistent use of terminology

Guideline

Use terminology consistently without synonymy.

Description

Synonymy has proven to be a challenge for users accessing information with the three exploratory case studies, particular for non-expert users unfamiliar with the domain. Ensuring that the language used throughout information in consistent across the entire site will improve user access to the web content and enhance information seeking patterns.

Information quality attributes

Consistent representation, language, understandability

Critical errors

Synonymous terms used interchangeably within content.

Rating for 'Consistent use of terminology'

Rating	**Criteria**
Rating 0	Synonymy present in content
Rating 1	Not applicable
Rating 2	Synonymy within content identified with in-text contextualisations
Rating 3	Not applicable
Rating 4	No synonymy in content

Derivation

All the case studies presented participants with some form of synonymy, though this was most present in the first task scenario. Whilst some users may

understand the analogous terms, others may not match comprehend the links between the concepts and assume that the information they are seeking is not available. In addition, where there are different terms used across different pages or sites, users may miss the links between the key themes and make assumptions about the applicability of the content.

A.4.4. Universal applicability

Guideline

Make sure language is appropriate for all anticipated audiences.

Description

The specific terminology used with website content can often include terms that can be applied in different ways depending on the context. Ensuring that the language used is universally applicable for expert and non-expert users as well as the range of website audiences will improve access to information and remove potential barriers for understanding content.

Information quality attributes

Accuracy, level of detail

Critical errors

Content includes many domain-specific terms explained in a way only expert users could understand.

Rating for 'Universal applicability'

Rating	**Criteria**
Rating 0	Language not universally applicable with domain-specific terms not explained
Rating 1	Not applicable
Rating 2	Some non-universal language with appropriate explanations
Rating 3	Not applicable
Rating 4	Language universally applicable with domain-specific terms explained clearly

Derivation

As encountered during the case studies, there are often occasions where professional nomenclature understood by expert users is different to that understand by those less familiar with the domain. For example, potential University applicants in the first example came across terms such as assignment that can be ambiguous without further explanation of what this means. Within the organisational context, there is likely a universal understanding of how the term can be applied and its consequences.

Similar examples were exposed in the third case study with the e-health website. Some terms such as painkillers or simple medical language were used within the content to provide simpler, more universal information although not technically consistent with the more specific medical terminology.

Appendix B. Ethical Approval documentation

This appendix includes key Ethical Approval documentation related to this research study.

B.1. Ethical Approval confirmation letters

The Ethical Approval for this research study was obtained in two stages. Both letters of confirmation are included below.

University of Salford MANCHESTER

Research, Innovation and Academic Engagement Ethical Approval Panel

Research Centres Support Team
G0.3 Joule House
University of Salford
M5 4WT

T +44(0)161 295 7012

www.salford.ac.uk/

11 May 2017

Dear Jessica,

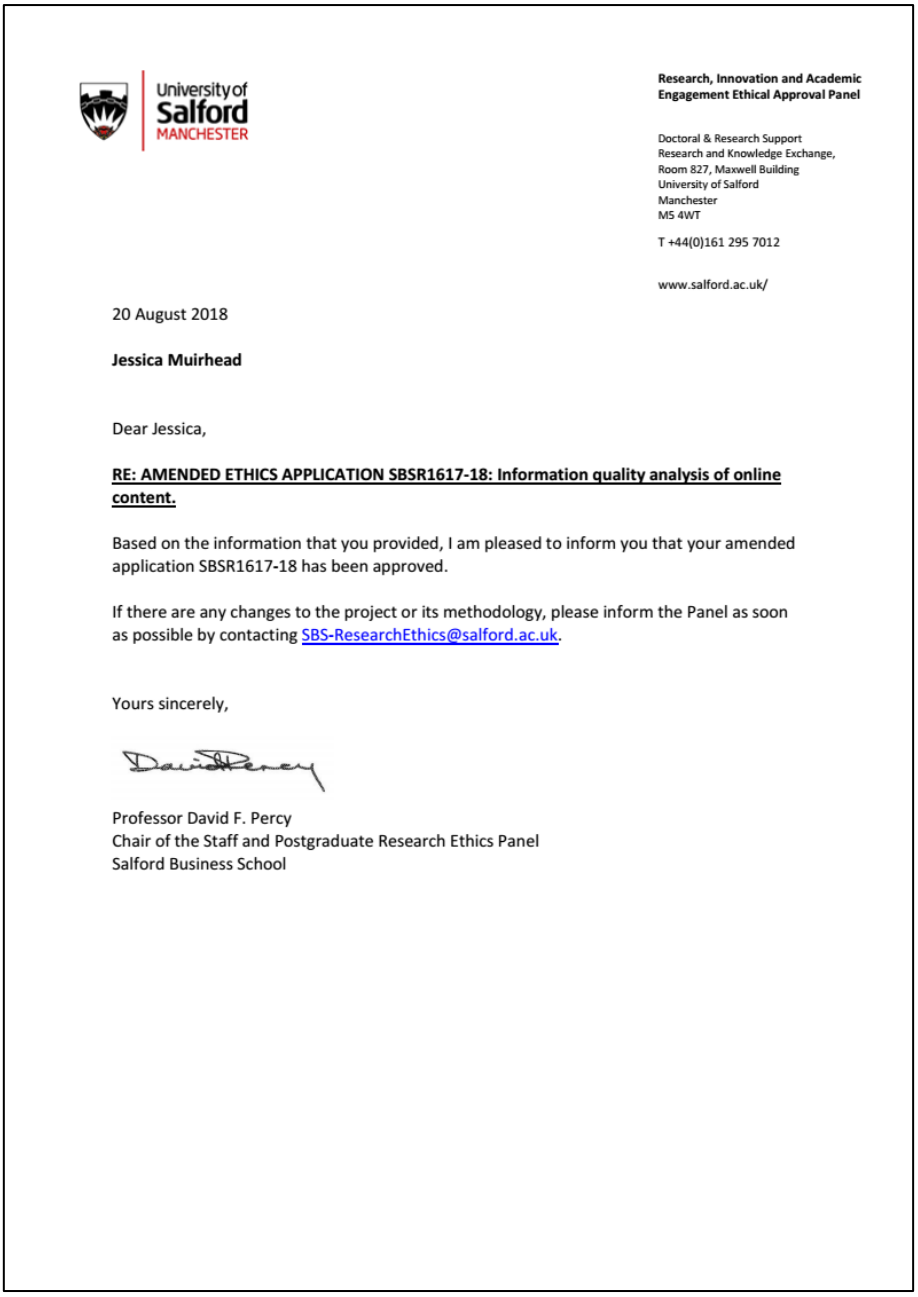
RE: ETHICS APPLICATION SBSR1617-18: How do user-centered techniques contribute to the design process to create online content that is informative and accessible?

Based on the information that you provided, I am pleased to inform you that your application SBSR1617-18 has been approved.

If there are any changes to the project or its methodology, please inform the Panel as soon as possible by contacting SBS-ResearchEthics@salford.ac.uk.

Yours sincerely,

Professor David F. Percy
Chair of the Staff and Postgraduate Research Ethics Panel
Salford Business School



University of Salford MANCHESTER

Research, Innovation and Academic Engagement Ethical Approval Panel

Doctoral & Research Support
Research and Knowledge Exchange,
Room 827, Maxwell Building
University of Salford
Manchester
M5 4WT

T +44(0)161 295 7012

www.salford.ac.uk/

20 August 2018

Jessica Muirhead

Dear Jessica,

RE: AMENDED ETHICS APPLICATION SBSR1617-18: Information quality analysis of online content.

Based on the information that you provided, I am pleased to inform you that your amended application SBSR1617-18 has been approved.

If there are any changes to the project or its methodology, please inform the Panel as soon as possible by contacting SBS-ResearchEthics@salford.ac.uk.

Yours sincerely,

Professor David F. Percy
Chair of the Staff and Postgraduate Research Ethics Panel
Salford Business School

B.2. Organisational invitation sheet

The following information sheet was sent to potential organisations to invite them to take part in this research.

University of Salford MANCHESTER

Salford Business School

Organisational Consent for Research Project

Information Quality of Online Content

I would like to invite you to assist me in conducting a research study. Before you decide you need to understand why the research is being done and what it would involve for you and for the participants. Please take time to read the following information carefully. Ask questions if anything you read is not clear or if you would like more information. Take time to decide whether or not to facilitate this research.

Who I am and what this study is about

I am a second year PhD student at the University of Salford researching the ways in which online content can be improved by applying user centred design techniques based on individual characteristics of information quality. This study will focus on identifying the key attributes that impact online content, and the way in which they can be applied to a website to produce a better experience for users. All research activities will be conducted in accordance with the University of Salford's ethical approval procedures, and full approval will be obtained prior to commencement of the study.

What I need your assistance with

During this research, it will be necessary to work with one or more websites to analyse current content and suggest improvements based on industry practices and academic literature. I hope to work with your organisation to make such changes to client website(s), providing an opportunity to measure the changes in user experiences provided by the techniques analysed.

What will taking part in the research involve?

The researcher will conduct user research studies prior to any research, and make content suggestions based on the academic literature and relevant theoretical models. After applying these changes to the online content, the same user research studies will be repeated to demonstrate the impact of the relevant information quality characteristics.

As part of the research, it would be useful to have access to relevant Google Analytics data to provide a benchmark and quantitative analysis of the online content, and confirm the results of the user research studies. In addition, it may be necessary to discuss and work with members of the team on the content improvement aspects to ensure these are appropriate to the client and congruent with the agencies' approach. I expect all modifications to content to be approved by the agency prior to any live amends.

1

SALFORD BUSINESS SCHOOL

Who will have access to data from research?

Any data produced from the research will be stored securely by the researcher in accordance with the University of Salford's ethical approval procedures. Data will be used solely for the purposes of PhD research, with participants remaining anonymous. It may be necessary to share select items of collected data with academic supervisors to ensure a rigorous research methodology is applied and to produce results relevant to the research community. All participants will need to give their explicit consent to these procedures prior to their involvement, and will be free to withdraw at any time.

What will happen to the results of the study?

The primary written output of this study will be the researcher's PhD thesis. In addition, there may be opportunities to publish segments of the work as conference and journal submissions. Where this does happen, the organisations and participants will be suitably anonymised unless explicitly agreed prior to publication.

Who should you contact for further information?

Researcher:

Jessica Muirhead
Salford Business School, University of Salford, The Crescent, Salford M5 4WT
j.muirhead@edu.salford.ac.uk 0161 295 2126

Main Supervisor:

Dr Maria Kutar
Salford Business School, University of Salford, The Crescent, Salford M5 4WT
m.kutar@salford.ac.uk 0161 295 3056

B.3. Participant invitation email

> The email below was sent to potential participants to invite them to take part in the think aloud studies.

This email is being sent to you on the understanding that you may be interested to take part in an academic research project.

Please ignore this email if you do not wish to take part – I won't contact you again.

If you do wish to take part, simply reply to this email saying "Yes! Count me in!"

Invitation to take part in academic research to improve (organisation's) website

About Me

My name is Jessica Muirhead, and I am a PhD student in the Centre for Digital Business at Salford Business School. The Centre for Digital Business is focused upon understanding the rapid changes new technologies are bringing to our society and the way they can be improved to deliver better user experiences.

About the project

I am working with (organisation) on a project to understand the ways in which online content can be improved to deliver more accessible and informative websites for users. As part of this study, I am asking participants to complete an activity to obtain their feedback on the current information on a web page. This will form part of a larger study to understand the evolution of online content.

Activity

I will ask you to complete an online activity to discover your experiences of using a website. The activity will involve both guided and unguided elements whilst being observed and discussing your use of the site.

Confidentiality

The contents of your responses will be completely anonymous, and no-one other than the researcher will ever know what you said. At the beginning of the task, you will be asked to consent to your responses being used for these purposes.

Questions?

If you have any questions about this research study, you can contact me by post: Jessica Muirhead, Maxwell 220, University of Salford, The Crescent, Salford, M5 4WT or by email: j.muirhead@edu.salford.ac.uk

So – if you want to take part simply reply to this email saying "Yes! Count me in!" I'll be in touch with a link to the questionnaire.

Best wishes

Jessica

Appendix C. Data analysis worked example

This appendix contains a worked example of the data analysis process as followed for participant [P1.2] during the first action case study based on the university website (see Chapter 8).

At the point of this worked example, the researcher had already categorised the first participant [P1.1]'s responses within the information quality attributes identified for the study.

C.1. Overview

For all the action case studies, the researcher applied the cyclical approach to data analysis as shown in Figure 6.19 on p. 190. This follows a four-step process for each new piece of data:

1. Identification of relevant data
2. Analysis of key themes
3. Incorporating data into narrative
4. Evaluating IQ attributes

The evaluation process involved taking each observation or utterance from a user study in turn, analysing key themes to incorporate the data into the existing narrative following the IQ attributes. Data was analysed directly from the researcher's notes generated during the activity, with collections of insights developed in a Microsoft Excel spreadsheet. This process led to both the identification of data relevant to the 11 existing attributes discussed in Chapter 4, and the three new attributes surfaced from the case study data (see Section 11.3.2).

To demonstrate the data analysis cycle, the remainder of this appendix describes the process of the application to the first case study (university

website) incorporating the second participant's data after the first participant's study has been considered. This shows the key practices and decisions made by the researcher when considering the evidence generated by the think aloud studies. Analysis of all participant data in all action case studies followed the same cycle though a full worked example is not included for brevity.

C.2. Initial data

The initial data for the first case study based on [P1.1]'s responses is summarised in the table below. This is taken as the starting state of data before the data analysis cycle begins.

Attribute	New?	Responses	Participant
Availability		Search feature required exact terminology	[P1.1]
		Information required in alternative formats	[P1.1]
Consistent representation		Change in navigation patterns between sites required re-learning	[P1.1]
Language		Some terminology forms barriers to accessing information, e.g., domain-specific terms such as UCAS	[P1.1]
Level of detail		Information far down page is sometimes missed	[P1.1]
Navigation		—	
Presentation		Same-page links styled as if they link to other content elsewhere on the site	[P1.1]
Recency		—	
Search		In-page search used to circumvent page structure	[P1.1]
Understandability		—	
Value added		Extra media in page can prove distracting when searching for information	[P1.1]
Visibility		—	
Fallback	✔	Would have preferred to use phone rather than online – alternative formats	[P1.1]
Information usability	✔	Commented that access to other formats would help them find information	[P1.1]

At this point, analysis of [P1.1]'s responses have revealed data relevant to 8 existing attributes, plus the two new attributes of *fallback* and *information usability*.

C.3. Applying data analysis cycle to [P1.2]'s responses

The next stage in the process involves incorporating [P1.2]'s responses into the data set using the data analysis cycle. Step 1 involves *identifying relevant data*, and the researcher performed this step during and shortly after the think aloud studies. Based on observations and the participant's utterances, several important data items with a potentially significant impact on the informativeness of online content were generated, though at this point not categorised by IQ attribute.

The researcher stepped through each data item sequentially following the order the data was obtained in the study notes. Each of these items are discussed below.

1. Synonymous terminology

The first item of data for analysis was an observation and user comment that they found the terms *tech* and *digital* confusing within the website. Despite a common interchangeability within the content and the relevant sector, the user did not understand that these are often synonymous terms (within the context of the website). This provided further evidence for the language attribute and established the first evidence of the previously identified understandability attribute.

2. Live chat window

As the participant was exploring the main university website, a live chat window opened in the lower right of the screen aiming to provide a fallback mechanism for the user to ask questions. This covered the key content the participant needed to find for the task scenario, leading to confusion and a need to spend longer on task than they otherwise would have. This was classified as a visibility issue, providing the first evidence for this attribute within the first action case study.

3. Entry Requirements detail

During the task scenario, participant [P1.2] navigated to the entry requirements section of the page. This provided an accordion to hide information until accessed. When the participant clicked to open the device, no content was visible in the area other than a number. This did not provide the information the participant was seeking, hence was

categorised as a level of detail issue. This provided a second piece of evidence for the level of detail attribute, in addition to the observations from [P1.1].

4. Plethoric use of links in page footer

Whilst searching for links to navigate to content relevant to the fourth question in the task scenario (see Figure 8.3), the participant struggled to find the choice they required within the footer of the page. The researcher prompted the participant to describe the issues they were encountering, with a response that they were frustrated that they could not find the link they needed within the page, thus ended up searching through many links placed in the footer. This provided further evidence of availability issues within the case study (a reliance on quick links due to navigation not possible elsewhere) and navigation issues within the site.

5. Change in navigation patterns between sites

When switching between the main university website and subsite to complete task 4, participant [P1.2] mentioned that they found the experience jarring due to the complete change in the style, layout, and structure of the page. This was due to the shift from a mainly externally facing platform to information maintained in a format closer to an intranet (though publicly accessible online). As the systems are maintained separately, none of the features carry across from one site to another. This experience led to the participant needing to re-learn how to navigate the content and find information, further corroborating the insights provided by participant [P1.1] for the *consistent representation* attribute.

After analysing the data and incorporating the five elements described above, [P1.2]'s insights provided additional insights for seven information quality attributes: availability, consistent representation, language, level of detail, navigation, understandability, and visibility. Some of this data improved upon existing insights (e.g., the change in navigation between sites), whereas other data provided fully new insights not previously considered.

C.4. Data at end of cycle

The table below shows the data at the end of the cycle, incorporating both [P1.1] and [P1.2]'s responses. The next stage of data analysis involved the inclusion of participant [P1.3]'s responses, and so on, for all of the data collected during this action case study.

Attribute	New?	Responses	Participant
Availability		Search feature required exact terminology	[P1.1]
		Information required in alternative formats	[P1.1]
		Reliance on quick links in the footer due to navigational links not being clearly labelled elsewhere	[P1.2]
Consistent representation		Change in navigation patterns between sites required re-learning	[P1.1, P1.2]
Language		Some terminology forms barriers to accessing information, e.g., domain-specific terms such as UCAS	[P1.1]
		Synonymous terms such as tech and digital prevent access to content	[P1.2]
Level of detail		Information far down page is sometimes missed	[P1.1]
		Lack of explanation for some content with minimal information	[P1.2]
Navigation		Too many links provided in footer – unable to find appropriate choice	[P1.2]
Presentation		Same-page links styled as if they link to other content elsewhere on the site	[P1.1]
Recency		—	
Search		In-page search used to circumvent page structure	[P1.1]
Understandability		Interchangeable use of terms technology and digital did not match participant's expectations from outside the sector	[P1.2]
Value added		Extra media in page can prove distracting when searching for information	[P1.1]
Visibility		Live chat window covered content, hiding key information required to complete task	[P1.2]
Fallback	✔	Would have preferred to use phone rather than online – alternative formats	[P1.1]
Information usability	✔	Commented that access to other formats would help them find information	[P1.1]

Note that the data from participant [P1.2] did not cover every IQ attribute, though more clarification was provided for availability, consistent representation, language, and level of detail. The analysis of this think aloud study led to new insights into navigation, understandability, and visibility. At this point, no evidence had been obtained to suggest the final new attribute, *interactivity,* nor *recency* as an existing attribute identified in the literature. Both attributes were considered later in the action case study based on data from further participants.

C.5. Summary

The cyclical data approach first outlined in Section 6.7.10 provides a framework for analysis of data obtained from the participant think aloud studies throughout this research project. Data collected from either utterances by the participants or observations of use of the case study are categorised and combined to form a richer picture of insights into each of the information quality attributes. This also led to the discovery of three new attributes within this research: those of fallback, information usability, and interactivity.

The data analysis cycle as applied above provides a method for thematic analysis of responses without requiring a preconceived notion of information to be considered. The information quality attributes identified from the literature in Chapter 4 provided an initial taxonomy, with insights providing evidence of the new attributes.

Appendix D. Application of model

During Chapter 12 of this thesis, the researcher applied the Informative Web Content Guidelines (IWCG) to the case study page. Some areas within the IWCG provided the highest possible rating of 4, indicating broad alignment with the principles of the model. The researcher did not consider improvements for these areas as any changes will not have as much impact as those areas rated lower.

The examples below are not exhaustive of all the modifications made as a result of the application of the model, however they demonstrate a subset of the changes made to the website to increase its potential informativeness.

D.1. Guideline 1. Availability

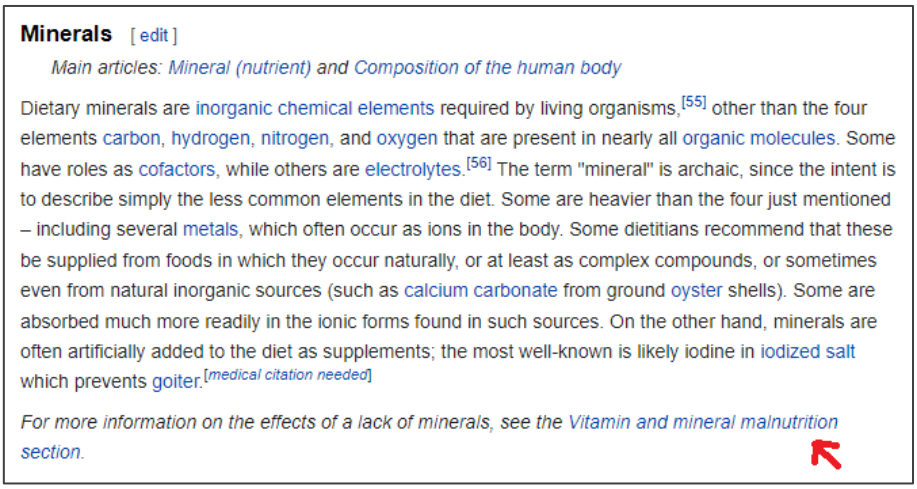
D.1.1. Alternative formats

The researcher added additional links to information in different formats and other authoritative resources, with prominent links at the top of the page. This appears as a banner above the page content:

D.2. Guideline 2. Information architecture

D.2.1. Alternative routes

Additional links between page sections have been added at the end of the relevant content to aid users in exploring the page via a non-hierarchical architecture. Examples of such links are shown below:

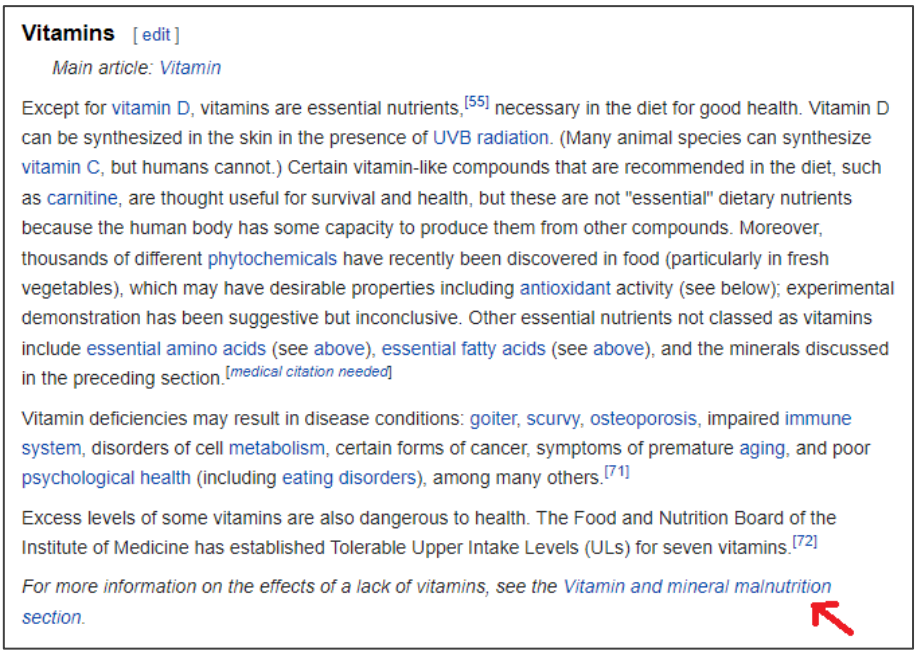


Minerals [edit]

Main articles: Mineral (nutrient) and Composition of the human body

Dietary minerals are inorganic chemical elements required by living organisms,[55] other than the four elements carbon, hydrogen, nitrogen, and oxygen that are present in nearly all organic molecules. Some have roles as cofactors, while others are electrolytes.[56] The term "mineral" is archaic, since the intent is to describe simply the less common elements in the diet. Some are heavier than the four just mentioned – including several metals, which often occur as ions in the body. Some dietitians recommend that these be supplied from foods in which they occur naturally, or at least as complex compounds, or sometimes even from natural inorganic sources (such as calcium carbonate from ground oyster shells). Some are absorbed much more readily in the ionic forms found in such sources. On the other hand, minerals are often artificially added to the diet as supplements; the most well-known is likely iodine in iodized salt which prevents goiter.[*medical citation needed*]

For more information on the effects of a lack of minerals, see the Vitamin and mineral malnutrition section.



Vitamins [edit]

Main article: Vitamin

Except for vitamin D, vitamins are essential nutrients,[55] necessary in the diet for good health. Vitamin D can be synthesized in the skin in the presence of UVB radiation. (Many animal species can synthesize vitamin C, but humans cannot.) Certain vitamin-like compounds that are recommended in the diet, such as carnitine, are thought useful for survival and health, but these are not "essential" dietary nutrients because the human body has some capacity to produce them from other compounds. Moreover, thousands of different phytochemicals have recently been discovered in food (particularly in fresh vegetables), which may have desirable properties including antioxidant activity (see below); experimental demonstration has been suggestive but inconclusive. Other essential nutrients not classed as vitamins include essential amino acids (see above), essential fatty acids (see above), and the minerals discussed in the preceding section.[*medical citation needed*]

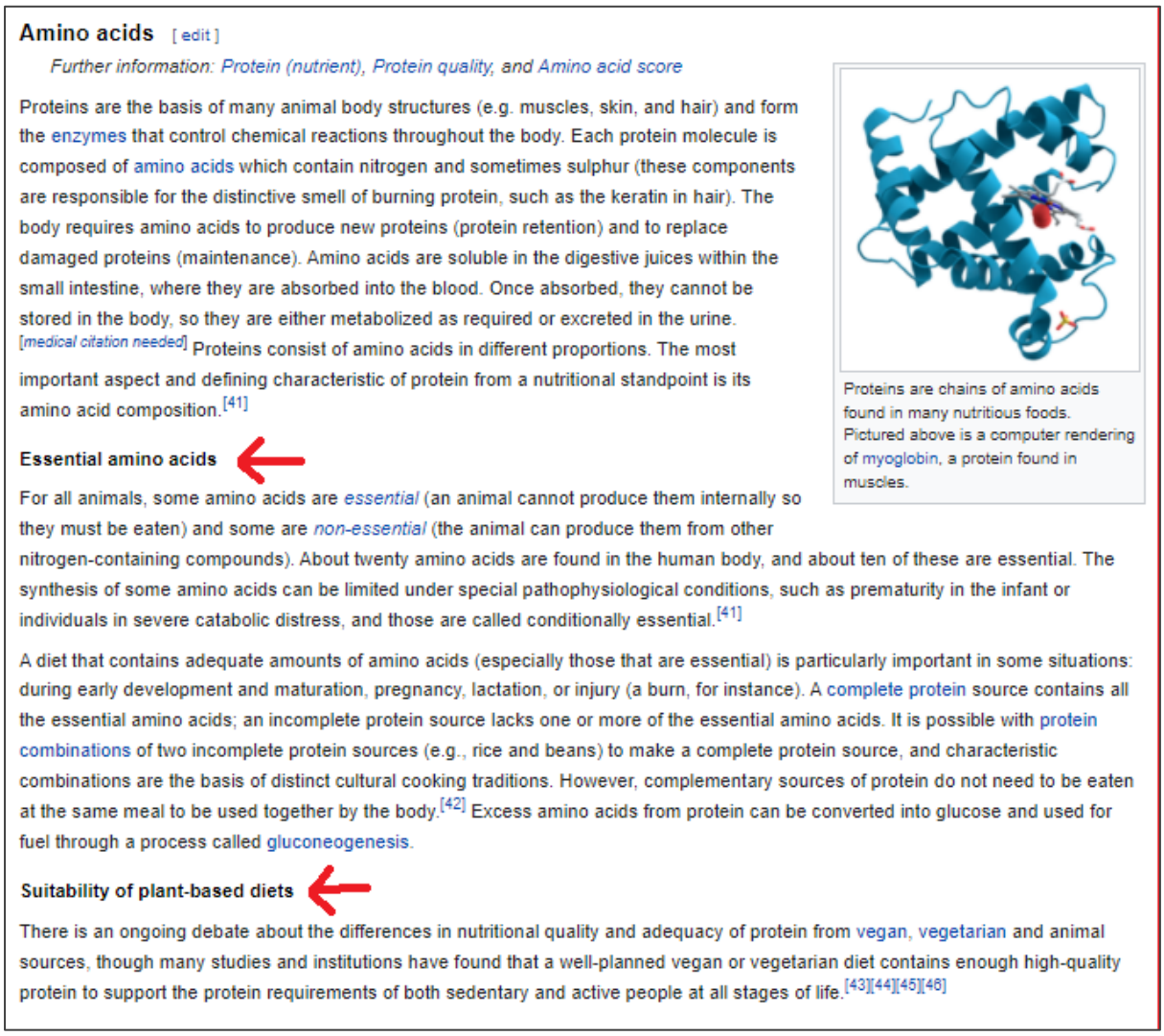
Vitamin deficiencies may result in disease conditions: goiter, scurvy, osteoporosis, impaired immune system, disorders of cell metabolism, certain forms of cancer, symptoms of premature aging, and poor psychological health (including eating disorders), among many others.[71]

Excess levels of some vitamins are also dangerous to health. The Food and Nutrition Board of the Institute of Medicine has established Tolerable Upper Intake Levels (ULs) for seven vitamins.[72]

For more information on the effects of a lack of vitamins, see the Vitamin and mineral malnutrition section.

D.2.2. Consistent representation

The major change to consistent representation is the structuring of content, with the addition of additional sub-section headings in large bodies of text, for example two new headings inserted within the amino acids section to restructure content into a more consistent and easy-to-find format:



Amino acids [edit]

Further information: Protein (nutrient), Protein quality, and Amino acid score

Proteins are the basis of many animal body structures (e.g. muscles, skin, and hair) and form the enzymes that control chemical reactions throughout the body. Each protein molecule is composed of amino acids which contain nitrogen and sometimes sulphur (these components are responsible for the distinctive smell of burning protein, such as the keratin in hair). The body requires amino acids to produce new proteins (protein retention) and to replace damaged proteins (maintenance). Amino acids are soluble in the digestive juices within the small intestine, where they are absorbed into the blood. Once absorbed, they cannot be stored in the body, so they are either metabolized as required or excreted in the urine. [*medical citation needed*] Proteins consist of amino acids in different proportions. The most important aspect and defining characteristic of protein from a nutritional standpoint is its amino acid composition.[41]

Proteins are chains of amino acids found in many nutritious foods. Pictured above is a computer rendering of myoglobin, a protein found in muscles.

Essential amino acids

For all animals, some amino acids are *essential* (an animal cannot produce them internally so they must be eaten) and some are *non-essential* (the animal can produce them from other nitrogen-containing compounds). About twenty amino acids are found in the human body, and about ten of these are essential. The synthesis of some amino acids can be limited under special pathophysiological conditions, such as prematurity in the infant or individuals in severe catabolic distress, and those are called conditionally essential.[41]

A diet that contains adequate amounts of amino acids (especially those that are essential) is particularly important in some situations: during early development and maturation, pregnancy, lactation, or injury (a burn, for instance). A complete protein source contains all the essential amino acids; an incomplete protein source lacks one or more of the essential amino acids. It is possible with protein combinations of two incomplete protein sources (e.g., rice and beans) to make a complete protein source, and characteristic combinations are the basis of distinct cultural cooking traditions. However, complementary sources of protein do not need to be eaten at the same meal to be used together by the body.[42] Excess amino acids from protein can be converted into glucose and used for fuel through a process called gluconeogenesis.

Suitability of plant-based diets

There is an ongoing debate about the differences in nutritional quality and adequacy of protein from vegan, vegetarian and animal sources, though many studies and institutions have found that a well-planned vegan or vegetarian diet contains enough high-quality protein to support the protein requirements of both sedentary and active people at all stages of life.[43][44][45][46]

D.2.3. Number of links

The number of links in the page was an issue at both the top and the bottom, with many links to related content and sources. Whilst reducing the number of links may have removed useful content for the user, the researcher provided further categorisation to aid participants in navigating to the

desired content. The example below shows the changes made within the *See also* section:

See also [edit]

General health [edit]

- Health
- Dieting
- Healthy diet
- Food group

Dietary professionals

[edit]

- Dietitian
- Nutritionist
- Food Studies

Types of food [edit]

- Diet food
- Fast food
- Functional food
- Junk food
- Food supplement
- Ultra-processed food

Healthy eating advice

[edit]

- 5 A Day
- Canada's Food Guide
- Food guide pyramid
- Healthy eating pyramid
- MyPyramid
- Nutritional rating systems
- Nutrition scale

Substances [edit]

- Dietary supplement
- Food fortification
- Nutraceuticals
- Probiotic
- Prebiotic (nutrition)

D.2.4. Recency

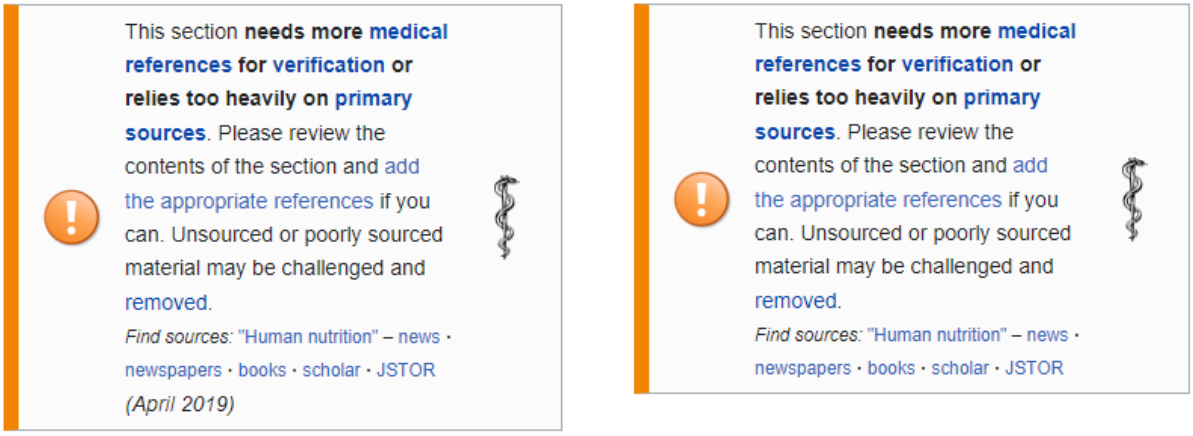
Through the application of the model, recency received a rating of 2 indicating areas for further improvement. The researcher identified two ways in which this could be applied:

1. Page history

2. In-page devices annotating lack of sources

To prevent users from the impression of out-of-date content, the researcher removed dates from the editor annotations:

Before *After*



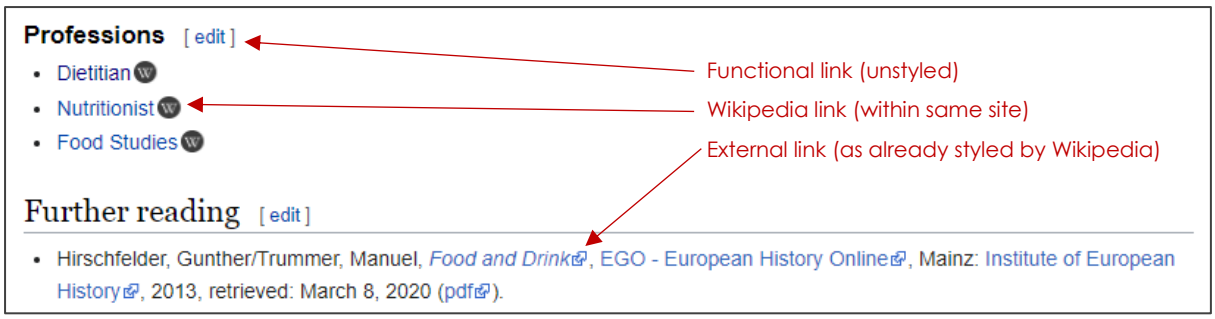
D.2.5. Search results

The model produced a rating of 3 for the search results due to a lack of relevance indicator or contextualisation during browsing. Due to the lack of control over the entire website during the action case study (due to use of a local mirror of the page), technical limitations prevented the researcher from improving this page. Further application of the model could provide additional simulation in this area to demonstrate benefits to users.

D.3. Guideline 3. Interactivity

D.3.1. Consistent interactions

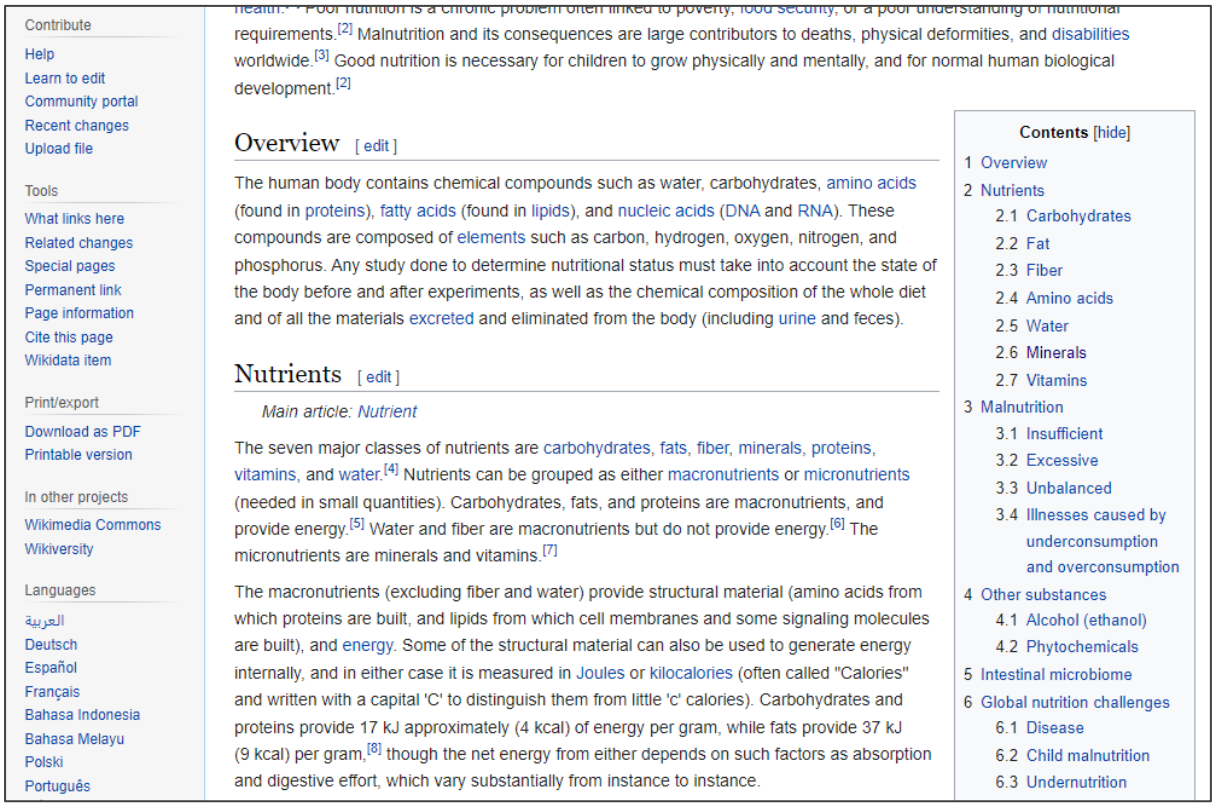
To provide consistent interactions across the page, links should be styled appropriately for their purpose. Wikipedia already uses an external link icon to signify links to other sites and denotes content that may be behind a paywall by a padlock icon. The researcher introduced an additional Wikipedia icon to highlight which links navigate to content within the same site. This is distinctive from functional links (such as those for edit operations) and anchor links within the same page to help users navigate content more easily.



D.3.2. Visibility

As identified by the IWCG model, visibility of information is a key area of concern, particularly where there are devices such as accordions that may hide content from the user.

To aid with visibility within the case study website, the researcher moved the table of contents to the right-hand side of the screen (for tablet devices and larger), reducing its width to make more of the page content visible for the user whilst still providing the secondary navigation in a more visible location.



D.4. Guideline 4. Language

D.4.1. Ambiguous terms

The IWCG suggests that ambiguous terms should be described in-context, to provide appropriate definitions. The below example shows refined clarifications within the context of the page:

women, plus an increase in proportion to the higher energy requirement, equal to 300 mL/day.[51] To compensate for additional fluid output, breastfeeding women require an additional 700 mL/day above the recommended intake values for non-lactating women. Dehydration and over-hydration – too little and too much water, respectively – can have harmful consequences. Drinking too much water is one of the possible causes of hyponatremia, i.e., low serum sodium.[51][54]

D.4.2. Clear use

The use of clear language when describing terms that may be unfamiliar to users is essential for informative websites. The researcher clarified several such terms, including the dietary fiber example below.

Before

Fiber [edit]

Main article: Dietary fiber

Dietary fiber is a carbohydrate, specifically a polysaccharide, which is incompletely absorbed in humans and in some animals. Like all carbohydrates, when it is metabolized, it can produce four Calories (kilocalories) of energy per gram, but in most circumstances, it accounts for less than that because of its limited absorption and digestibility.

The two subcategories are *insoluble* and *soluble* fiber.

After

Fiber [edit]

Main article: Dietary fiber

Dietary fiber is a form of carbohydrates with multiple units of sugar (known as a polysaccharide). In some animals and humans, dietary fiber may not be fully absorbed by the body. As with all other carbohydrates, when it is digested by the body it can produce at most four Calories (kilocalories) of energy per gram. In most circumstances limited absorption and digestability reduce the amount of energy obtained from dietary fiber.

The two subcategories are *insoluble* and *soluble* fiber.

D.4.3. Consistent use of terminology

The IWCG suggest that terminology must be used consistently within the same page. The researcher changed several terms in the article to ensure language is used more consistently, such as the example below where the word *sugar* is substituted throughout the introductory text.

Before

> Molecules of carbohydrates and fats consist of carbon, hydrogen, and oxygen atoms. Carbohydrates range from simple monosaccharides (glucose, fructose, galactose) to complex polysaccharides (starch). Fats are triglycerides, made of

After

> Molecules of carbohydrates and fats consist of carbon, hydrogen, and oxygen atoms. Carbohydrates are a form of sugars, ranging from simple monosaccharides (one-unit sugars) to complex polysaccharides (multi-unit sugars, commonly known as starches).

D.4.4. Universal applicability

The principle of universal applicability specifies that content should be authored for a general rather than expert audience. The researcher made several adjustments to the page content, including the example shown below:

Before

Carbohydrates [edit]

Main article: Carbohydrate

Carbohydrates may be classified as monosaccharides, disaccharides or polysaccharides depending on the number of monomer (sugar) units they contain. They are a diverse group of substances, with a range of chemical, physical and physiological properties.[13] They make up a large part of foods such as rice, noodles, bread, and other grain-based products,[14][15] but they are not an essential nutrient, meaning a human does not need to eat carbohydrates.[16] The brain is the largest consumer of sugars in the human body, and uses particularly large amounts of glucose, accounting for 20% of total body glucose consumption.[17] The brain uses mostly glucose for energy; if glucose is insufficient however, it switches to using fats.[18]

After

Carbohydrates [edit]

Main article: Carbohydrate

Carbohydrates are a common source of energy found in both natural and processed foods. They make up a large part of foods such as rice, noodles, bread, and other grain-based products,[14][15] but they are not an essential nutrient, meaning a human does not need to eat carbohydrates.[16] Carbohydrates are a form of sugar – largest consumer of sugars in the human body is the brain, which uses particularly large amounts of glucose, accounting for 20% of total body glucose consumption.[17] The brain uses mostly glucose for energy; if glucose is insufficient however, it switches to using fats.[18] There are three different types of carbohydrates used by the body, categorised by the number of sugar *units* they contain: monosaccharides (one unit), disaccharides (two units), or polysaccharides (three units or more) They are a diverse group of substances, with a range of chemical, physical and physiological properties.[13]

Appendix E. Researcher achievements

This appendix contains a list of academic publications and presentations by the researcher during the period of doctoral studies.

Muirhead, J. (2021). Preventing underage alcohol purchasing online using payment card details. Institute of Alcohol Studies (IAS). https://www.ias.org.uk/wp-content/uploads/2021/12/IAS-Preventing-underage-alcohol-purchasing-online-using-payment-card-details.pdf

Muirhead, J., & Grout, V. (2020). Effective age-gating for online alcohol sales. Alcohol Change UK. https://s3.eu-west-2.amazonaws.com/files.alcoholchange.org.uk/documents/Effective-age-gating-for-online-alcohol-sales-Final-Report.pdf

Kreps, D., Rowe, F., & Muirhead, J. (2020). Understanding Digital Events: Process philosophy and causal autonomy. Proceedings of the 53rd Hawaii International Conference on System Sciences, 6133–6142.

Kreps, D., & Muirhead, J. (2019). Introduction to Understanding Digital Events. In Understanding Digital Events: Bergson, Whitehead, and the Experience of the Digital (pp. 1–13). Routledge.

Muirhead, J. (2019). Assessing Online Content Quality through User Surveys and Web Analytics. Proceedings of the 2019 Conference on Human Information Interaction and Retrieval, 433–436.

Kreps, D., & Muirhead, J. (2018). 'Infomateriality': Whitehead and digital experience amongst the over 65s. European Association for the Study of Science and Technology (EASST) Conference, Lancaster, UK.

Muirhead, J. (2018a). A framework for user-centred design technique selection for quality online content. UK Academy of Information Systems (UKAIS) Doctoral Consortium, Oxford, UK.

Muirhead, J. (2018b). Information quality analysis of online content. British HCI Doctoral Consortium, Belfast, UK. http://hci2018.bcs.org/prelim_proceedings/papers/Doctoral%20Consortium%20Track/BHCI-2018_paper_197.pdf

Muirhead, J. (2017a). Constructionist approaches to teaching hands-on app development. EngageHEI, Salford, UK.

Muirhead, J. (2017b). How do user-centred techniques contribute to the design process to create online content that is informative and accessible? UK Academy of Information Systems (UKAIS) Doctoral Consortium, Oxford, UK.

Muirhead, J. (2016). User-centred techniques for informative and accessible content. Psychology of Programming Interest Group (PPIG) Doctoral Consortium, Cambridge, UK.