DOCTOR OF PHILOSOPHY

Disentangling Digital Preservation Risk
An Interdisciplinary Exploration and Solution

Pennock, Maureen

Award date:
2024

Disentangling Digital Preservation Risk: An Interdisciplinary Exploration and Solution

By Maureen Elizabeth Pennock

ORCiD: 0000-0002-7521-8536

Thesis submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy at the University of Dundee

 October 2023

Table of Contents

List of Figures

List of Tables

Abbreviations used

ACM	Association for Computing Machinery
AIT	Austrian Institute of Technology
AI	Artificial Intelligence
API	Application Programming Interface
ARA	UK Archives and Records Association
ASCII	American Standard Code for Information Interchange
BL	British Library
BPML	Business Process Modelling Language
CCSDS	Consultative Committee for Space Data Systems
CD	Compact Disc
CEDARS	Curl Exemplars in Digital Archives
CHARM	Conceptualising and CHaracterising digital preservation risk: A Reference Model
CLIR	Council on Library and Information Resources
CPU	Central Processing Unit
CRISP	Crowdsourcing Representation Information to Support Preservation
CRL	Center for Research Libraries
CTS	Core Trust Seal
DCC	Digital Curation Centre
DESRIST	Design Science Research in Information Systems and Technology
DiAGRAM	Digital Archiving Graphical Risk Assessment Model
DNA	Deoxyribonucleic acid

DNB	Deutsche Nationalbibliothek
DPC	Digital Preservation Coalition
DPC-RAM	Digital Preservation Coalition Rapid Assessment Model
DPE	Digital Preservation Europe
DRAMBORA	Digital Repository Audit Method Based On Risk Assessment
DRM	Digital Rights Management
DSA	Data Seal of Approval
DSR	Design Science Research
DSRM	Design Science Research Methodology
DSRP	Design Science Research Process
DVD	Digital Video Disc *or* Digital Versatile Disc
EAP	Endangered Archives Programme
EC	European Commission
EPUB	Electronic Publication (format)
ERPAnet	Electronic Resource Preservation and Access network
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
FEDS	Framework for Evaluation of Design Science
FFMA	File Format Metadata Aggregator
GML	Geography Mark-up Language format
HDD	Hard Disk Drive
IA	Internet Archive

ICA	International Council on Archives
IDCC	International Digital Curation Conference
IDP	International Dunhuang Project
IEEE	Institute of Electrical and Electronics Engineers
IFLA	International Federation of Library Associations
IJDC	International Journal of Digital Curation
INESC-ID	Instituto de Engenharia de Sistemas e Computadores: Investigação e Desenvolvimento em Lisboa
INSPECT	INvestigating the Significant Properties of Electronic Content over Time
iOS	iPhone Operating System
IPS	Integrated Preservation Suite
IP	Information Package
ISO	International Standards Organisation
IT	Information Technology
JASIST	Journal of the Association for Information Science and Technology
JCDL	Joint Conference on Digital Libraries
JPEG	Joint Photographic Experts Group format
KB-NL	Koninklijke Bibliotheek van Nederland
LAN	Local Area Network
LDL	Legal Deposit Library
LOCKSS	Lots of Copies Keeps Stuff Safe
MP3	Motion Pictures Expert Group (MPEG) Audio Layer 3 format

MPT	Minimum Preservation Tool
NARA	U.S. National Archives and Records Administration
NDSA	National Digital Stewardship Alliance
NPLD	Non-Print Legal Deposit
OAIS	Open Archival Information System
OCLC	Online Computer Library Centre (though typically now known only by its acronym)
OPF	Open Preservation Foundation
PASIG	Preservation and Archiving Special Interest Group
PDF	Portable Document Format
PERICLES	Promoting and Enhancing Reuse of Information throughout the Content Lifecycle taking account of Evolving Semantics
PNG	Portable Network Graphics format
PORRO	Preserved Object and Repository Risks Ontology
PREMIS	Preservation Metadata: Implementation Strategies
R&D	Research and Development
RAID	Redundant Array of Inexpensive Disks *or* Redundant Array of Independent Disks
RAS	Risk Assessment Spreadsheet
RI	Representation Information
RIF	Risk Identification Framework
RLG	Research Libraries Group
RSS	Really Simple Syndication

SAA	Society of American Archivists
SCAPE	SCAlable Preservation Environments
SPOT	Simple Property Oriented Threat
SRA	Society for Risk Analysis
SSD	Solid State Drive
TIFF	Tag Image File Format
TPDL	Theory and Practice of Digital Libraries
TRAC	Trustworthy Repositories Audit and Certification
UK	United Kingdom
UKOLN	UK Office for Library and information Networking (though typically now known only by its acronym)
ULCC	University of London Computing Centre
UML	Unified Modelling Language
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICODE	Universal Character Encoding
URL	Uniform Resource Location
US	United States
USB	Universal Serial Bus
UTF-8	Unicode Transformation Format (8 bit)
VPN	Virtual Private Network
VRC	Virtual Remote Control
WAN	Wide Area Network
XHTML	Extensible HyperText Mark-up Language

Acknowledgements

I am hugely grateful to the people who have supported me over the course of this PhD. It's been quite the journey. If I'd had a crystal ball then I probably would have waited a few years. I'm glad I didn't though. What a ride.

Most of all, I am grateful to my family – my husband Dom, and our children, Millie, George and Will. Your encouragement gave me the self-belief to get started. Your support has kept me going. Your patience has been exceptional. I wouldn't have got here without you. Thank you, from the bottom of my heart.

I'm grateful to my supervisors, Nancy McGovern, Patricia Whatley, and Craig Gauld. Your compassion and support has been very much appreciated. A particular thank you goes to Nance, for asking me the difficult questions that made me look again at what I'd done, and realise I could make it better.

I'm grateful to the British Library for giving me the opportunity to undertake this research. The Library is made special by two things: the collection, and the people that work with it. Several people at the Library have helped me along the way, including my line managers and colleagues, but most of all, my team. You have all helped, in different ways.

I'm grateful to all the friends and wider family who've supported me as well. A big thanks to Amy Rose and Sally Oldfield especially. That seemingly never-ending question… how's the PhD going… yep! Got there in the end.

Thank you, all.

Declaration

I, Maureen Elizabeth Pennock, hereby confirm that this thesis and the corresponding practical submission is my own work. I have consulted all references cited unless otherwise stated. This work has not previously been submitted or accepted for any other higher degree award.

Maureen E. Pennock, October 2023

Dedication

To Dom, for everything.

Abstract

Memory institutions such as the British Library face the important challenge of preserving their digital collections for future generations. Disciplinary efforts to address this challenge are extensive but demonstrate significant inconsistency and uncertainty about how the field understands risk, as well as what it considers to be a valid response. Moreover, they are often not easily aligned with wider, organisational risk management practices. This research, undertaken at the British Library as a practice-based PhD, addresses that problem by asking the question 'how can the nature and complexity of digital preservation risk be more thoroughly and consistently represented, so as to support the foundations for a more flexible yet comprehensive preservation planning risk response?'

A design science research methodology provided the framework for the research. Requirements for a new solution were established through a thorough review of the problem space. Risk science provided the structure for a new, meaningful conceptual definition of digital preservation risk, distinguishing between the concept of risk and its characterisation. Analysis of the risk source concept against this definition led to the design and population of a new conceptual reference model for digital preservation risk: CHARM. A series of methods were designed that demonstrate different ways to use the model, aligning with international risk management standards and practices. The outputs of these methods can subsequently inform a comprehensive preservation planning risk response.

This research makes a significant and original contribution to knowledge with the field's first reference model of digital preservation risk. It demonstrates that by drawing on risk science, digital preservation risk can be more consistently expressed, more thoroughly represented and more clearly demonstrated to stakeholders than before. Through its contribution of a new model for defining the digital preservation risk domain alongside three methods for applying it, the research represents a clear improvement-based and prescriptive contribution to knowledge. It provides not only a deeper understanding of the problem but also a new solution for responding to digital preservation risk, designed at the British Library but relevant to all in the wider community who share their challenge.

Chapter One: Introduction

Research Problem

The modern world is digital, characterised by the bits and bytes through which we live, work, play, and communicate. Digital information and digital content have become the cornerstone of the modern cultural memory and scholarly record, both of which are fundamental to the ongoing development of society. Memory organisations such as the British Library now acquire many millions of digital files, every single year. This digital content, just like non-digital elements of collections, must be proactively managed and preserved for future users, lest it becomes damaged, deleted or lost. The practice of 'Digital Preservation' is the solution to this problem.

Digital Preservation is an applied field that seeks to ensure the longevity and survival of intangible digital information so that it can be made available to future users without significant changes to its intellectual meaning. Though with its feet firmly in the archival and library domain (Hedstrom, 1984; U.S. National Archives and Records Service, 1984; Kenney and Personius, 1992; Tylers, 1995), contemporary thinking and development of digital preservation practice draws on and integrates practices from other disciplines ranging from computer science to information systems, data science, management science, and economics, amongst others (Antunes *et al*., 2011; Innocenti, 2012; Maemura, Moles and Becker, 2017; Pennock and Coufal, 2017). Many of these also represent applied fields – the domain of information systems, for example, applies and develops theory in order to solve problems relating to organisational use of information technology. Modern day thinking about digital preservation is thus highly interdisciplinary, drawing on these related fields to explore the various challenges of digital preservation and how they may be resolved.

The term ‘digital preservation’ means different things to different people. A common misconception is that digital preservation is about digitisation, i.e. the creation of digital representations of physical items, on the assumption that this will support preservation of the original artefact. This perhaps originates from an assumption that memory institutions are still predominantly focused on collecting physical artefacts such as books, paintings, papers and so forth. Modern memory institutions increasingly collect both analogue and digital content, including that which is ‘born digital’ as well as digitised. The term ‘born digital’ describes an artefact that is created in a digital environment for use in the same. Its origins are independent of any specific physical artefact. For contemporary acquisitions in some national libraries and archives, born digital content may even be preferred over physical (British Library, 2020, p. 3) and the balance between analogue and digital in collecting institutions increasingly tips towards the digital. The focus of ‘digital preservation’ is these digital collection objects themselves, regardless of their origins. As Conway states:

> *‘Digitisation for preservation creates valuable new digital products, whereas digital preservation protects the value of those products regardless of whether the original source is a tangible artefact or data that were born and live digitally.’* (Conway, 2010, pp. 64-65).

In short, digital preservation maintains the value of an object for as long as needed, regardless of its original source. It is a broad ranging endeavour that encompasses many different types of challenges - Lavoie and Dempsey (2004) suggest there are at least ‘thirteen ways’ of thinking about digital preservation, each way considering a particular set of uncertainties or options. This reflects the interdisciplinary nature of the field and the many different lenses through which to consider and support digital preservation - from organisational and cultural, to economic, human, and technological.

A key perspective that cuts across all of these lenses is risk management. The theme of risk is pervasive in digital preservation literature and thinking, with digital preservation variously described as a ‘risky business’ (Underdown, 2019), ‘a series

of risks and strategies to mitigate them' (Johnston, 2020, p. 193), and even 'a risk management activity' (Ross and McHugh, 2006). Technological obsolescence is often cited as one of the main risks to digital longevity and digital preservation (Conway, 1996; Garrett and Waters, 1996; Curtis *et al*., 2007; Pearson and Webb, 2008; Todd, 2009; Ryan, 2014), reflecting the particular dependencies between digital file formats and the technological environments in which files are created. The rapid pace of technological change particularly in the 1990s caused concern for organisations with a need to provide reliable access to authentic digital materials after their original environments had become outdated and obsolete. The short time-frame in which obsolescence could manifest caused particular alarm (Hedstrom, 1997/98, p. 191), especially the suggestion that digital records lasted 'forever or five years', whichever came first (Rothenberg, 1995, p. 42). For memory institutions with a significantly longer timeframe, this was an exceedingly short window of opportunity in which to act.

Strategies for responding to the risk of obsolescence often focus on specific technological approaches, such as migration of objects to newer file formats or emulation of an old environment on a modern computer (Granger, 2000; CEDARS, 2002; Potter, 2002; van der Hoeven, Lohman and Verdegem, 2007; Rechert, von Suchodoletz and Welte, 2010; Cochrane *et al*., 2019). Others explore how software preservation can support these strategies (Matthews *et al*., 2009; di Cosmo and Zacchiroli, 2017) or develop technical knowledge bases and tools to help identify technologies and their technical dependencies prior to implementing such strategies (McGath, 2013; May, Pennock and Russo, 2019; Spencer, 2022). Whilst these approaches focus on a technology-driven solution to the problem, others take a more measured approach to informing selection of an appropriate strategy. Re-usable risk or threat assessment frameworks are one such type of approach and several have been developed particularly to explore format-based risks (Bennett, 1997; Lawrence *et al*., 2000; Stanescu, 2004; Rosenthal *et al*., 2005; Rog and van Wijk, 2008; Barateiro *et al*., 2010; Vermaaten, Lavoie and Caplan, 2012; Graf and Gordea, 2013; Pennock, Wheatley and May, 2014; Johnston, 2018). There is however often relatively little evidence of these frameworks being re-used outside

of their originating institutions, indicating a degree of uncertainty about their broader utility.

Other types of risks relating to more organisational matters, such as legal issues, policies and budgets, can also affect the longevity of digital material (Garrett and Waters, 1996; Ross, 2000, pp 18 - 19; McGovern, 2007; Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2010; Corrado, 2022). A number of assessment frameworks have been developed that look beyond the format to address and assess capabilities in these wider areas. The Trustworthy Repository Audit and Certification Checklist (TRAC) was one of the first to tackle this successfully, grouping assessment metrics into three main areas: organisational infrastructure, digital object management, and technologies, technical infrastructure, and security (Task Force on Digital Repository Certification, 2007). Similar groupings and metrics are evident in several subsequent repository assessment tools (ISO, 2012b; nestor, 2013; Core Trust Seal, 2020; Digital Preservation Coalition, 2021), indicating a good degree of confidence in the appropriateness of the coverage for this type of assessment.

Assessment tools of this sort make frequent reference to trust or trustworthiness and often identify risks within some of their metrics, though it is rare that they are badged specifically as risk assessment tools. Some function as an audit and/or certification tool (ISO, 2012b; Core Trust Seal, 2020) whereby compliance is considered to be evidence of success in digital preservation (Giaretta *et al.*, 2019). Others operate as a form of maturity model whereby assessors benchmark their current capabilities against a scaled set of maturity levels (Dollar and Ashley, 2014; Digital Preservation Coalition, 2021; NDSA, 2022). Most of these are not, however, risk assessment tools in and of themselves. Reservations have begun to be expressed about whether conformance with this type of assessment standard, even certification, is adequate for purposes of risk mitigation (Frank, 2022, p. 45). A small number of assessment tools have been developed that expressly aspire to support risk assessment (DCC and DPE, 2007; The National Archives, 2023b), though these too can prove challenging to use in a way that clearly and consistently identifies different types of risks.

At the heart of the problem is a lack of clarity into the concept of digital preservation risk and its various meaningful manifestations. Digital preservation risk appears to be everywhere – the PORRO ontology of digital preservation risk for example identifies over 600 'risk classes' that collectively represent risk-related elements including institutional requirements, object-centric characteristics, and associated risk cause or effect factors (McHugh, 2016). Other tools and solutions frequently explore the potential risks to different outcomes and processes without clearly or consistently defining what is meant by the term 'risk' (e.g. Vermaaten, Lavoie and Caplan, 2012; The National Archives, 2023a). Those that do define the term often do so in generic terms or without substantive, practical contextualisation (e.g. Barateiro *et al.*, 2010, p. 6; Dappert, 2013, p. 117; McHugh, 2016, p.3). A report on the State of the Art in Digital Preservation in 2018 went so far as to directly ask 'What future problems and risks are we trying to solve or mitigate?' (Rieger, 2018). Much of the thinking about digital preservation risk remains heuristic rather than clearly evidence-based (Altman and Landau, 2020, p.2) and the seemingly precautionary position of many memory institutions means that even in cases where insufficient evidence of a theoretical risk has accumulated, it continues to feature in risk assessment frameworks. Format obsolescence for example is considered by many to be less of a problem than originally envisaged (Rusbridge, 2006; Gollins, 2009; Rosenthal, 2010; Jackson, 2012; van der Knijft, 2013), but it continues to assume a central role in assessment frameworks (Graf and Gordea, 2013; Johnston, 2018; Friedrich, 2019). From a practical and experiential perspective, the nature of digital preservation risk remains grounded in uncertainty.

This uncertainty is compounded by the vocabulary that the community variously uses to describe and explore digital preservation risk, which is inconsistent and often poorly explained. As a result, it is difficult to determine exactly what is meant by the phrase 'digital preservation risk'. This causes many problems: if risk is not well understood, then it is very difficult to meaningfully manage; moreover, if it is not understood well by the people responsible for managing it, it is even more difficult to convincingly communicate it to other audiences. This poses a particular

challenge in an institutional setting, where digital preservation risk requires management alongside other types of risk, sometimes with competing priorities. The lack of clarity also makes it challenging to establish acceptable levels of risk – it is impossible to completely eliminate risk, but it should certainly be controlled so that it is within institutional tolerance levels. How can this be done when it is poorly and inconsistently understood? From a consistency perspective, digital preservation risk remains a 'social construct' (Frank, 2018, 2020) rather than something that the community has clearly explored and agreed upon.

Inconsistent and poorly explained descriptions of risk inhibit meaningful disciplinary advances in thinking about solutions for managing digital preservation risk. Despite the apparent plethora of risk tools that the community has developed, the concept of 'Digital Preservation Risk' remains somewhat amorphous and represents a 'wicked problem' in research. Wicked problems are those that are 'ill-formulated' and where the overall ramifications are 'thoroughly confusing' (Buchanan, 1992, p. 15, citing Rittel, 1967). The problem of digital preservation risk arguably qualifies against these criteria. This lack of clarity hampers meaningful discussion about digital preservation risk both within the community and within memory organisations where effective communication with non-digital preservation experts is essential to 'make the case' for digital preservation investment. Digital preservation risk often remains something of a spectre rather than something that can be meaningfully and transparently managed, particularly within the wider context of other business activities. Moreover, a disconnect between the risk responses (i.e. the 'solutions') and the risks themselves has the potential to undermine the vital work done to manage and preserve digital collections for reliable re-use, as they are not always clearly based in evidential need.

Risk is clearly a significant problem in digital preservation. The literature on digital preservation risk is extensive. The assessment frameworks are many, and the corresponding solutions are various. Similar themes are nonetheless repeated time and time again, without necessarily demonstrating advancement in the overall disciplinary understanding of risk. Overall, digital preservation risk is a widely explored yet poorly understood phenomenon.

This thesis responds to the problem that the scale and multi-faceted nature of digital preservation risk is poorly and inconsistently understood. As a nebulous and amorphous concept, it is difficult to manage and even more difficult to clearly communicate, particularly in a manner that is holistic and transparent to non-experts. In an applied field, this is first and foremost a knowledge problem, though it ultimately requires a practical solution underpinned by a solid and consistent understanding of the challenge that is digital preservation risk.

Research Motivation

The motivation for this research comes from the author's role at the British Library, where they lead the Digital Collection Management and Preservation Department. This includes a responsibility for assessing risks relating to the longevity of the digital collection, so that it is reliable and accessible for future readers.

The author's responsibility for assessing digital preservation risk has led to practical experience with many of the existing disciplinary tools or methods for exploring and managing digital preservation risk. These experiences, alongside the practical day-to-day work of delivering a digital preservation service, led to the author's experience that not only is the disciplinary understanding of digital preservation risk uncertain, but that the current solution space has not enabled a sufficiently nuanced or relational exploration of risk in a manner that aligns with established corporate risk management practices. Whilst existing solutions are all of value in their own right, this integration is essential: digital preservation is not the only business function in an organisation, and digital preservation risks must often be assessed against other types of risks and priorities in order to determine viable organisational responses.

The primary motivation for this research is therefore the need to better understand the nuance of digital preservation risk, so that digital preservation risk can be more effectively managed and communicated in an organisational setting.

The secondary motivation for the research is a desire to expand disciplinary thinking around the concept of preservation planning as a risk response, driven by work underway at the British Library to develop a preservation planning system

known as the Integrated Preservation Suite (IPS). The disciplinary concept of preservation planning as a risk response is all too often limited to technical and format related risks (Becker *et al*., 2009; Graf, Gordea and Ryan, 2014; Johnston, 2018; Friedrich, 2019; Skødt, 2022). The templates or processes for exploring and managing these are often inconsistent with standard risk management approaches and they further the perception that the main risk to digital longevity is format obsolescence. This suggests a gap in disciplinary explorations of how preservation planning processes might look beyond the format and respond to a wider range of digital preservation risks. A formal, nuanced understanding of digital preservation risk is essential for the practice to move forwards and for the concept of preservation planning to evolve and address risks beyond those related to technological obsolescence. Work at the British Library takes a wide perspective on preservation planning (Day *et al*., 2014a; Day *et al*., 2018; Pennock, Day and Samaras, 2019; Pennock, 2020), based on our organisational awareness and understanding of the wider risk landscape and the need for information on our digital collection that transcends a technical characterisation or profile. The nature of digital preservation risk is much broader in spectrum than simply format-based risks. A narrow focus, whilst predominant within the community, will limit the relevance of the IPS platform for developing risk responses at the Library.

Both of these motivations represent a need to implement improved digital preservation risk management and preservation planning activities at the British Library. The motivation for this research is thus very practical in nature. It responds to a business need for a way to align digital preservation risk assessments of various different types with corporate solutions for the same. It also explores the broader relationship between risk management and preservation planning. This will not only inform the future development and direction of the Library's IPS platform but simultaneously contribute to wider disciplinary discussions and thinking about the nature and scope of preservation planning in practice. The research originates from the British Library but its range and scope makes it relevant to all types of memory institutions with a responsibility for preserving digital collections.

Research Question

The research question for this thesis addresses the core problem that whilst digital preservation risk is a major cause for concern for memory institutions, the concept of digital preservation risk is poorly and inconsistently understood. Disciplinary discussions and explorations of digital preservation risk are extensive but inconsistent across the digital preservation risk landscape, in terms of both coverage and language. Risk assessment frameworks often focus on technological and obsolescence related issues despite indications that format obsolescence is less of a concern than originally thought, whilst larger repository assessment frameworks that are suggestive of risk management use the language of trust and trustworthiness rather than risk. This inconsistency and lack of clarity in our current disciplinary understanding inhibits the standardisation of digital preservation risk management practices across different scenarios. It not only poses challenges for alignment with wider institutional risk management processes but also makes it difficult to see how the concept of preservation planning can apply beyond the format to function as a more coherent digital preservation risk response and treatment framework.

A more thorough understanding of digital preservation risk would enable institutions to consider and respond more consistently and transparently to different types of digital preservation risks. The research thus summarises the problem statement and research motivation by seeking to answer the following question:

How can the nature and complexity of digital preservation risk be more thoroughly and consistently represented so as to support the foundations for a more flexible yet comprehensive preservation planning risk response?

By answering this question, the thesis lays the foundations for practical improvements to the British Library's approach for managing digital preservation risk. Sharing this with the wider digital preservation community supports disciplinary advancement in thinking about digital preservation risk, which can in turn inform subsequent solutions and improvements to solutions. Digital

preservation is always ‘shooting at a moving target’ (Hofman, 1999) so solutions developed today must be extensible and flexible enough to remain relevant for as long as feasibly possible. The approach taken by this research bears that in mind.

Digital Preservation Concepts and Models

As an emergent field, digital preservation is still in the process of establishing its individuality. Related fields and concepts such as digital curation (e.g. Beagrie, 2006; Abrams, 2015; Dobreva and Duff, 2015; Dollar, 2016; Poole, 2016; Higgins, 2018), digital archiving (e.g. Steenbakkers, 2005; Leggett, 2021), digital stewardship (e.g. Saachi, 2015; Langley, 2019), and digital sustainability (e.g. Bradley, 2007; Lazorchak, 2011), all cover similar ground. There are clear inconsistencies in how each is or should be understood within the community (Cunningham, 2008; Dallas, 2016), and it is not uncommon for these terms to be used interchangeably (e.g. Hedstrom, 2003; Gollins, 2009; Digital Preservation Coalition, 2015a; Underdown, 2019; Blumenthal *et al*., 2020; Abrams, 2021; Post and Chassanoff, 2021). The recent Digital Archive and Preservation (DAP) Framework acknowledges this particularly in relation to digital preservation and digital archiving, suggesting a long term or ‘over-time’ perspective as one potential disambiguator (McGovern, 2022). Despite this however, no single source has as yet been widely accepted that authoritatively differentiates between them all. This terminological confusion reflects the relative immaturity of the disciplinary vocabulary and the inconsistency acknowledged in the problem statement, further underlining the importance of clarity over key terms and concepts used when seeking to advance disciplinary knowledge.

There is no single and firmly established definition of digital preservation, though recurring themes prevail in different definitions throughout the professional literature. The British Library for example, defines digital preservation broadly as ‘the combination of actions and interventions required throughout the digital content lifecycle to ensure continued and reliable access to authentic digital materials’ (British Library, 2017, p. 2). UNESCO summarises it simply as ‘all those processes aimed at ensuring the continuity of digital heritage materials for as long as they are needed’ (2023). The Digital Preservation Coalition takes a similar

perspective, defining digital preservation as 'the series of managed activities necessary to ensure continued access to digital materials for as long as necessary' (2015a). The American Library Association definition goes a step further and describes digital preservation in more detail as a combination of 'policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change', noting also the ultimate goal of 'accurate rendering of authenticated content over time' (2008). This more detailed definition acknowledges not just the high-level goal but also specific challenges and potential responses.

Definitions in other works (e.g. RLG-OCLC, 2002; Pennock, 2006c; Brown, 2013; Owens, 2018) consider similar themes: ongoing processes, managed activities, future access, and authenticity. Recent efforts to identify attitudinal principles for evaluating success in digital preservation – arguably a reflection of its purpose – have focused particularly on the concepts of authenticity and accessibility, alongside integrity and usability (Abrams, 2021, 2023). Collectively, all represent important concepts in digital preservation and the challenge of ensuring digital longevity.

Models are a helpful tool with which to explore the relationships between these concepts and the domain in general. Modelling is a long-established tool in engineering and the natural sciences (Rosenblueth and Wiener, 1945; Hutten, 1954) that has since expanded across many other disciplines and interdisciplinary research endeavours (Hesse, 1976; Friedman, Friedman and Pollack, 2008; Cabot Vallecillo, 2022). The term 'model' can be associated with a broad spectrum of potential manifestations (e.g. Rothenberg, 1990; Duit and Glynn, 1996), with different forms suited to different types of research or purposes - one form is not necessarily better than any other (Giere, 2001).[1] Regardless of form, models typically function as abstract or simplified representations of ideas, objects or concepts. They support many different research and communication activities, from

[1] The term 'diagram' is often considered synonymous with 'model' and the two used interchangeably. For the purposes of this thesis, a distinction is drawn between them in that whilst a model represents a state of simplified and abstract representation, a diagram may represent all component parts without abstraction or simplification.

testing and refining a theory, to prototyping a solution, assessing capabilities, understanding relationships, simulating an environment, or providing clarity on core functional components. Different types and representations of models range from the material to the mathematical, the conceptual to the procedural, the graphical to the descriptive, and the computational to the theoretical.

Conceptual models are particularly valuable in information systems design research as a graphical way to identify and represent relationships between different entities and properties prior to the development of a system (Frank, 1999; Rothenberg, 1990; Olivé, 2007; Fettke, 2009). Reference models are a type of conceptual model not necessarily intended for a specific system implementation. They are characterised instead by their high level of abstraction, their reusability and utility, and their comprehensive representation of a given domain, which can be used as a reference from which to generate other conceptual or implementation models (Fettke, Loos and Zwicker, 2005; Thomas, 2007; Winter, Gericke and Bucher, 2009, p.2). Their use and significance can vary across domains (Gray and Rumpe, 2021), with their function or purpose ranging from exploration of new concepts and requirements verification to semantic reference catalogues or common ontological vocabularies (Lee, 2005, p. 53).

As mechanisms for understanding an emerging interdisciplinary domain, models are valuable research tools with which to explore and represent different concepts in digital preservation (e.g. Hedstrom, 2002; Ross, 2006). The range of model types developed within and around the field is extensive, exploring a broad range of topics, functions, and challenges. These include (but are not limited to) detailed conceptual domain-level models (Candela *et al.*, 2007; ISO, 2012a; Dappert, 2013; Abrams, 2015), illustrative high-level domain models (McGovern, 2007; Moulaison Sandy and Corrado, 2018), cost and sustainability models (Slats and Verdegem, 2005; Wheatley and Hole, 2009; Abrams, Cruse and Kunze, 2012; Grindley, 2013; L'Hours *et al.*, 2014), graphical lifecycle and process models (Higgins, 2008; Choudhury, Huang and Palmer, 2020), ontological and metadata models (Constantopoulos and Dritsou, 2007; Mikelakis and Papatheodorou, 2012; Bakhshandeh *et al.*, 2013; PREMIS Editorial Committee, 2015), risk and threat

models (Rosenthal *et al*., 2005; Vermaaten, Lavoie and Caplan, 2012; McHugh, 2016; Barons *et al*., 2021), capability and maturity models (Kenney and McGovern, 2003; Dollar and Ashley, 2014; Digital Preservation Coalition, 2021; NDSA, 2022), and many more. Each helps to advance thinking in a specific aspect of the field, taking a range of different graphical or descriptive forms. The use of formal modelling languages such as the Unified Modelling Language (UML) or Business Process Model and Notation language (BPMN) is not uncommon particularly in publications with a technological and domain or system-level perspective, though their suitability typically depends on the type of model, the problem area under investigation, and the target community served by the model.

As risk is an inherent feature of digital preservation, it follows that all digital preservation models represent either implicitly or explicitly an element of risk. The potential for risk across many different aspects of digital preservation suggests a particular value for domain-level models in representing risk or a risk response framework. The illustrative domain model of a three-legged stool for example (McGovern, 2007) identifies three core aspects of digital preservation – organisation, technology, and resource - to which equal attention must be paid else the stool falls over. Risk management is not explicitly addressed but is arguably required to prevent the stool from falling. The more extensive ISO 14721 domain Reference Model for an Open Archival Information System (OAIS) (ISO, 2012a) takes a different approach and explicitly incorporates risk analysis into one of its six functional entities, that of 'preservation planning'.

The Reference Model for an Open Archival Information System was initially published as a draft for discussion by the Consultative Committee for Space Data Systems (CCSDS) in the late 1990s, becoming ISO standard 14721 in 2003.[2] The preservation planning functional entity in ISO 14721 addresses the risk of obsolescence to ensure that digital content remains accessible, understandable, and usable to its community (ISO, 2012a, p. 4-2; Lavoie, 2014). Activities to achieve this include monitoring the designated community and technologies for issues that

[2] The current version of ISO 14721 was published in 2012. For a full account of the early developmental history of OAIS, see Lee (2005).

could cause obsolescence or prevent access to holdings, responding to changes and emergent risks (from both internal and external environments) that could affect preservation activities, developing preservation policies and strategies, and producing migration plans. ISO 14721 now forms the basis of much of the digital preservation community's shared vocabulary around information packages and functional stages, though whilst it uses the term 'preservation planning' extensively, it does not define the term 'preservation plan' (Pennock, 2020). Preservation plan templates within the community have nonetheless been developed, generally with a particular focus on addressing or minimising expected format and obsolescence-based risks (e.g. Becker *et al.*, 2009; Graf, Gordea and Ryan, 2014; Johnston, 2018; Friedrich, 2019). This is broadly in line with the purpose of the preservation planning functional entity as obsolescence-avoidance, though it reflects a level of disconnect between perceptions of digital preservation risk when the OAIS Reference Model was first published and contemporary thinking about the broader and more expansive nature of digital preservation risk as an organisational and holistic concern.

The literature of the domain thus represents a broad range of model types from which to learn, and an array of themes or concepts to consider when seeking to achieve greater clarity in the overall digital preservation risk landscape.

Risk Management and the Concept of Risk

The human practice of managing risks has a long history (Covello and Mumpower, 1985; Hay-Gibson, 2008), though its development as an academic field is more recent with its roots primarily in the mid-twentieth century (Dionne, 2013). There is much internal discussion on both the status and name of the risk field (Cumming, 1981; Beck, 2004; Thompson, Deisler and Schwing, 2005; Yeo, 2019), culminating most recently in promotion of the term 'risk science' to represent and formalise academic explorations into risk. Risk science is an emergent field of thinking evolved from risk analysis (Yeo, 2019, p. 6) and dominated by a relatively small number of authors (e.g. Krewski *et al.*, 2014; Hansson and Aven, 2014; Aven, 2016; Westphal *et al.*, 2017; Aven and Thekdi, 2022; Hao, Li and Wu, 2023; Ylönen and Aven, 2023). It seeks to provide a systematic and objective understanding of risks

through the development and application of models and methods for assessing and predicting risks. The term ‘risk science’ represents development of the most reliable (i.e. epistemically most warranted) and contemporary knowledge on risk concepts, assessment, communication, and management (Aven and Thekdi, 2022, p. 312).

Just as science may be classed as basic or applied, with basic science representing the core knowledge base of a particular domain and applied science using that knowledge to devise solutions, so too may risk science. Generic (or basic) risk science is concerned with development of ‘concepts, principles, approaches, methods, and models for understanding, assessing, characterising, communicating, managing, and governing risk’, whilst applied risk science supports scientific knowledge generation for specific scenarios and activities, often from an interdisciplinary perspective (Aven, 2020, p. 1889). Both can exist independently in a risk research endeavour though may also be interactive: applied risk research can highlight challenges not convincingly addressed by existing approaches and therefore lead to development of new approaches that represent generic risk science.

Risk Management represents the practical application of risk science principles, approaches, methods and concepts. It is the process by which risk is identified, analysed, and where appropriate, mitigated so that it remains within acceptable levels. Risk management is a well-established and widely documented practice. There are various risk management frameworks and standards, including the so-called ‘Orange Book’ of risk management processes and concepts for the UK public sector (UK Government, 2023) as well as other national or pan-national frameworks from the UK, Europe, Australia and New Zealand and beyond. Whilst there is some degree of variation between these standards, they are often informed by or consistent with the global risk management standard from ISO, represented by the ISO 31000 risk management framework of standards documents. This framework includes the core standards document ISO 31000 (ISO, 2018), a vocabulary guide (ISO, 2009), and ISO 31010 on risk assessment techniques (ISO, 2019). Collectively these define a framework of widely accepted processes, terminology, and methods for identifying, assessing and managing risks.

The ISO risk management process represents the 'application of management policies, procedures and practices to the activities of communicating, consulting, establishing the context, and identifying, analysing, treating, monitoring and reviewing risk' (ISO, 2018, p.8). By this description, the risk management process is perhaps more constructively conceptualised as a set of interrelated processes rather than a single end-to-end activity.

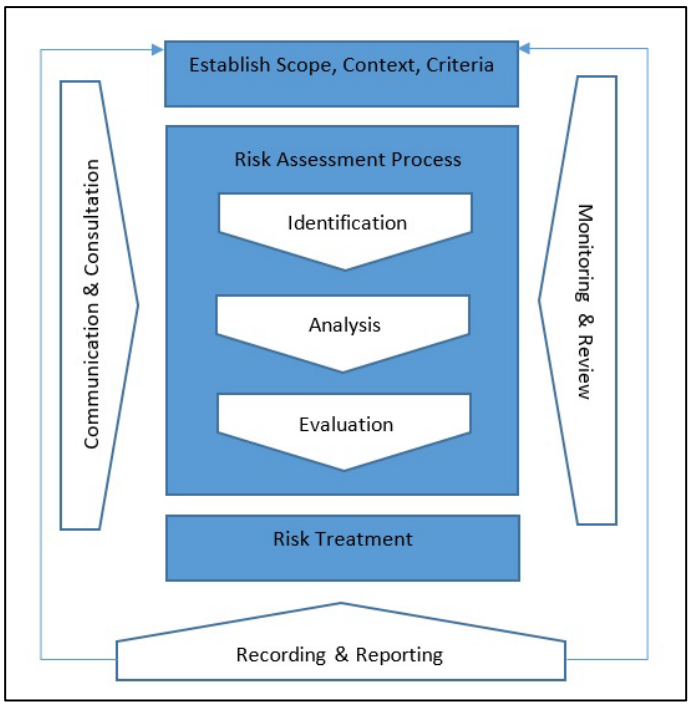

Figure 1: ISO 31000 Risk Management Process

Within the Risk Management Process sits a Risk Assessment Process, comprised of three main stages: *identify* sources of uncertainty that may result in a risk; *analyse* uncertainties in terms of source, likelihood, consequence and any other relevant characteristic, and; and *evaluate* analysis to inform next steps or treatment options. This risk assessment process itself is relatively linear, though as the environment in which risk is managed is rarely static, the overall risk management process is iterative.

Academic literature on risk and risk science often explores concepts in a different manner from ISO 31000, which defines the term 'risk' primarily as the 'effect of uncertainty on objectives' (ISO, 2009, p. 1). An effect is a deviation from what is otherwise expected, either positive or negative. In an uncertain world however,

anything and everything may potentially be perceived as a risk. The Society for Risk Analysis (SRA) observes that the ISO definition is open to interpretation, and suggests a longer definition that also specifies the negative or undesirable nature of an outcome (Society for Risk Analysis, 2018). Moreover, whilst the ISO definition works at an abstract level, it is important to recognise a distinction between the concept of risk and the practical manifestation of risk. The ISO definition of risk is sufficiently abstract that it arguably presents risk as a concept. Yet when it comes to risk management, risk needs to be described in more detail so that it can be analysed, evaluated, and treated. It needs more context and precision in order to be meaningfully manageable.

This thesis draws particularly on risk science thinking as a foundation for defining what is meant by the term 'digital preservation risk'. Ylönen and Aven observe that 'much of the confusion observed in practice concerning risk can be tracked back to the concept of risk being mixed with its measurement or characterisation', arguing the value of a distinction between conceptual descriptions of risk and descriptions of practical, manifest risk (2023, p. 592). Greater contextualisation, at both conceptual and practical levels, can provide a mechanism through which to address this. A precise framing or description of risk can also go some way to help counteract distorted risk perception, whereby risks are interpreted in a way that is inconsistent with reality. This is particularly helpful in digital preservation given its predominantly theoretical underpinnings when it comes to risk. Clarity of terminology is essential in any scientific field and especially when exploring largely conceptual matters, yet terminological imprecision is evident in much of the digital preservation risk literature. It is arguably a significant contributor to the overall uncertainty within the field about how to deal with digital preservation risk.

Methodological Considerations

The methodology underpinning this research is that of Design Science. The practical nature of this methodology is appropriate to the practical context of this research, particularly given its suitability for so-called 'wicked problems' and the characterisation of digital preservation risk as such. Design science research is characterised by its blend of knowledge, creativity, and practicality, based around

the fundamental principle that 'knowledge and understanding of a problem domain and its solution are achieved in the building and application of the designed artefact' (Hevner *et al.*, 2004, p. 75). It is a pragmatic endeavour that starts with a problem and uses different methods as necessary to devise and deliver a usable solution (Romme, 2003; Goldkuhl, 2012; Thuan, Dreschler and Antunes, 2019). A design science research methodology is an inherently iterative one that bounces between problem and solution space until both are sufficiently well understood that the knowledge generated through this process can be represented in one or more usable artefacts.

The usable artefacts generated through this research are discussed in this thesis but also contained in a companion submission that represents the artefactual and practical outputs of the research. This two-part submission is a requirement of the practice-based PhD format, an applied and creative form of doctoral research with a specific practical output. The practice-based approach is suitable to the nature of this research given the practical context from which it originates and within which the solution will be used. A practice-based PhD is sometimes also referred to as an 'artefact and exegesis' PhD.[3] In an artefact and exegesis PhD, the artefact represents a practical output, whilst the exegesis represents an accompanying written narrative. Artefact and exegesis PhDs are most common in the creative arts though the form has more recently gained traction for PhDs in other disciplines. Brabazon, Hunter and Quinton, for example, extol the value of an artefact-exegesis PhD for representing creative elements of research from disciplines such as science, engineering, and medicine (2022, p. 51). MacKrell, McDonald and Gammack (2017) argue the suitability of an artefact-exegesis structure particularly for design science research and PhDs in information systems, where the artefact can comprise anything from software, algorithms and workflows, to methodologies and policies. As with design science, the artefact-exegesis approach represents an inherently iterative methodology whereby the two parts develop alongside each other, each informing the contents of the other until the final form of both is settled.

[3] This term has particular traction in Australasia but is less frequently used in Europe.

Different forms of artefact and exegesis relationships exist, reflecting various priorities and expectations about the relationship between the two. For the purposes of this research, a research question model is employed whereby the two parts are 'conceptualised as independent answers' but can also be integrated to form a coherent whole (Milech and Schilo, 2004). The artefacts in this research communicate a solution to the research question for a practitioner audience, whilst the written thesis explores them and the research question in a wider academic context.

Contextual Considerations

The research outlined in this thesis was produced whilst the author was Head of Digital Preservation and Digital Collection Management at the British Library. The massive scale and diversity of operations and digital collection content at the British Library, an organisation with over 1500 members of staff, presents an ideal microcosm in which to explore the disciplinary nature of the challenge and generate outputs that are relevant to other memory institutions regardless of their own scale of collecting.

The Library's digital collection is vast and exceptionally heterogeneous. Content dates back to the 1980's when the Library first started to receive deposits of content that included floppy discs. The Library's digitisation programme started in the late 1990's and the collection now includes extensive amounts of digitised material produced at the Library's own digitisation studios or elsewhere, including through large scale commercial digitisation agreements and initiatives such as International Dunhuang Programme and the Endangered Archives Project.[4] Moreover, as a national and Legal Deposit Library (LDL), it receives content with many different rights associations and licences, from hundreds of different publishers. Annual digital acquisition rates now exceed those of the physical

[4] The Endangered Archives Programme (EAP) funds digitisation of internationally endangered physical archives. The British Library administers the programme and received copies of digitised content from sources around the world for preservation in its digital repository. The International Dunhuang Project (IDP) is a collaborative endeavour to digitise and make available materials relating to Dunhuang and the archaeological sites of the Eastern Silk Road. The British Library manages images and metadata for several IDP partners that do not have their own infrastructure.

collection. The Library is also home to the UK Web Archive and the national Sound Archive, both of which contain millions of individual files and objects. From born digital to digitised, licenced and purchased, to legal deposit, dataset, image, sound file, web archive and more, the digital collection represents a phenomenal testbed of varying content types, formats, data structures, technological dependencies, ages, descriptions, rights and constraints against which to develop and test the artefacts presented in this thesis.

Several events occurred during the course of this PhD that provided further insight into the complex nature of digital preservation risk, beyond those presented by the inherent complexities of the British Library and its digital collection. Two were specific to the Library's core technological infrastructure. The first was a digital repository migration project, requiring the installation of a brand new commercial repository system on a different storage architecture and the migration of digital collections from the old system to the new. The second was an infrastructure renewal programme, changing the underlying infrastructure hosting the new digital repository application and used for onsite file storage. This posed considerable constraints to technological operations during the implementation and update process, including the deployment and integration of the new repository system. Both initiatives presented opportunities for refinement and testing of the research presented in this thesis by presenting practical examples of new risk scenarios.

Moreover, the occurrence of a black swan event during this PhD in the form of a global pandemic led to an emergency prolonged lockdown of the sort that few had foreseen, with massive and significant impacts on operational capabilities and capacities. This gave another perspective on risks and risk sources in a digital preservation setting, as well as how quickly organisations might need to respond. Library staff were sent home and processes that were designed for onsite working had to be conducted remotely across an infrastructure with limited bandwidth for external access, designed to minimise and tightly control external access not facilitate it at scale. Depositing organisations faced similar challenges and there were major interruptions to acquisition, management and processing workflows, necessitating rapid response risk assessments to address this emergency situation.

A further event with global relevance occurred during the course of this PhD when Russia declared war on The Ukraine. Many nations made declarations of support and cyber security attacks across the UK began to rise rapidly. These served to highlight the relationships between digital preservation risk management, cyber security, and disaster planning, confirming the value of an integrated digital preservation risk management approach over a functionally isolated risk management practice.[5]

A final and still uncertain external event came in the form of the meteoric rise of Artificial Intelligence (AI) tools in early 2023. This has the potential to impact significantly on processes and workflows at the Library as well as the wider environment, and the frequent errors in AI outputs mean that it is not without its risks. The impact of this has yet to be felt but is the subject of a watching brief. All of these events have informed development of this practice-based, empirically grounded yet creative-led research.

Limitations of the Research

This research uses the British Library as the source organisation against which to identify and model digital preservation risk. The justification for this limitation in the approach is primarily the driver for the research, namely to ensure that the outputs can be used by the British Library to help improve digital preservation risk management practices. The wider context of the research and the research question requires however that the outputs are more broadly usable than just at the British Library, reflecting upon and representing the nature and complexity of digital preservation risk for the discipline and wider community in general. Several reasonable assumptions are therefore made about the suitability of the British Library context as a transferable and representative model source, specified here for reasons of transparency and rigour.

[5] Post-submission addendum: The Library fell victim to a cyberattack shortly after the submission of this thesis. For examples of media articles around that attack, see for example Knight, 'The Disturbing Impact of the Cyberattack at the British Library' (The New Yorker, 19 Dec. 2023) and Cooke, 'Writers left in a bind by British Library cyber-attack, but it remains a closed book' (The Observer, 6 Jan. 2024). Discussions of the impact of that event and the subsequent organisational response are not within the scope of this research and are not explored further.

Firstly, this research assumes that the British Library digital collection is sufficiently broad in both depth and breadth as to be representative of issues affecting most different types of digital collections, particularly those found in libraries and archives. With well over 2 petabytes of content in the collection, many millions of individual files, a vast array of different content types from different sources, a plethora of different technological dependencies, and multiple different conditions of use associated with it, this seems a reasonable assumption to make. It is further supported by the author's experience and knowledge of how this compares to digital collections at other, similar memory institutions. Some institutions may nonetheless collect content such as scientific research datasets – which the Library collects relatively few of – that the author currently has relatively little experience with. Efforts have therefore been made to remain considerate of different collection types but agnostic wherever possible in descriptions of terms. Illustration and examples are nonetheless frequently drawn from the Library and Archive domain, as this represents the author's main concern.

Secondly, it assumes that the British Library's functional structure is – at least at a high level - broadly reflective of the areas found in most other institutions. This is significant as it represents the institutional context in which risks manifest and thus against which the research presented in this thesis has been developed and tested. This assumption is validated to a degree by review of other assessment frameworks that define similar structural arrangements, particularly with regards to organisational and technological elements. Effort has nonetheless been taken to describe organisational and technological functions in sufficiently homogenous terms so as to be recognisable to most institutions, regardless of their own specific structures.

Thirdly, it assumes that valid feedback was received from colleagues during the course of refining and testing the research outputs, and this was not unduly influenced by the author's relatively senior position in the Library. Colleagues were encouraged to speak freely and frankly when providing feedback, to ask probing questions and to provide examples of situations in which proposed constructs or aspects of the model would not work. Feedback was sought from not only digital

preservation colleagues, but also those in various different positions across the Library with whom there was no direct departmental or line management responsibilities in place, as well as from colleagues more senior to the author. Further feedback was sought from a small number of external parties, however, due to the requirement for this research to represent original knowledge, external sharing of ideas was limited until very late in the process. Ideally there would have been wider community input to the research during earlier stages of development, as well as external testing, but the field and the subject area is still very much under development. This has instead been factored into a plan for post-submission updates.

The scope of the research presented in this thesis is also limited. The original aim of the research was to devise and deliver a toolkit for holistic management of digital preservation risk and preservation planning. Exploration of the problem area however revealed that there was much more uncertainty about the concept of digital preservation risk than was originally expected. As a result, significantly more attention was needed first on the conceptual foundations of digital preservation risk. A decision was therefore made to focus the practical element primarily on a reference model of digital preservation risk rather than a full toolkit, with guidance on viable methods in the form of those used at the British Library but ultimately allowing users to develop their own method for implementation, as appropriate to their needs. Questions around, for example, how to trigger a risk assessment, or what a holistic preservation planning template might look like, are therefore not addressed by this research.

A further limitation relates to this thesis' adoption of the risk science concept. Risk science is a relatively new area and young field that is dominated by a small number of authors. There are reservations in some quarters over whether this work truly constitutes a 'science' (Yeo, 2019, p. 6) and it is possible that these ideas have yet to be fully peer-reviewed within the risk community. Risk science is nonetheless a term used frequently in contemporary explorations of risk concepts and it is therefore adopted for this thesis. The principles and the methodology applied are

sufficiently rigorous that they should remain valid regardless of the nomenclature used to describe the field.

The cut-off date for the literature and solutions review was summer 2023. This was sufficient time for release of the iPRES 2023 conference programme, which was reviewed to ascertain whether any significantly new work was scheduled in relation to this thesis that the author was not already aware of.[6] There are also language limitations to this research, with the majority of the literature and solutions review focused on English language content. It is possible that relevant work has been published in other languages that has had relatively little wider exposure in the international community and that the author is not aware of.

Structure of the thesis

This thesis is structured into eight main chapters. Chapter one – this chapter – sets the scene for the research. Memory institutions such as the British Library face an important challenge in preserving their digital collections for future generations. Solutions have been developed in the field over the past twenty five years to address this challenge, though disciplinary understanding of digital preservation risk remains inconsistent, inhibiting integration of solutions with standard risk management practices and wider conceptual thinking around preservation planning. This thesis addresses that problem by asking the research question 'how can the nature and complexity of digital preservation risk be more thoroughly and consistently represented, so as to support the foundations for a more flexible yet comprehensive preservation planning risk response'.

Chapter two presents the design science research approach used to progress the research, with the core methodological tenets of creativity, pragmatism, scientific rigour, and utility. It also acknowledges the author's prior work in this area, establishing the personal knowledge base from which this research originates.

Chapter three considers the wider disciplinary knowledge base upon which the research builds. It identifies four main types of assessment frameworks devised

[6] iPRES is the digital preservation community's main international annual conference.

within the community over the past twenty-five years, focused either on formats, collections, systems, or organisations. It summarises the main solutions developed in each category, and analyses them in light of the wider digital preservation risk literature to more precisely identify the limitations of current solutions.

Chapter four represents the initial foray into production of a potential new solution. It summarises the analysis of previous approaches in order to more precisely characterise the problem area and consider what this might imply for the research. This forms the basis of the requirements for the new solution, focused in particular on clarity and precision of terms, a comprehensive scope, and transparency of process.

Chapter five is the main chapter, presenting a new, innovative approach to thinking about digital preservation risk. It uses an approach from the field of risk science to develop a meaningful conceptual definition of digital preservation risk, framed in relation to target values, objectives, and sources of uncertainty. It then uses this as the basis from which to develop a series of abstract models on digital preservation risk, culminating in the Digital Preservation Risk Source Model. This reference model identifies and populates a series of risk originating entities, classes of risk sources, risk factors associated with each class, and risk source instance types. Collectively, the models represent a new reference framework with which to subsequently explore and characterise individual manifestations of digital preservation risk.

Chapter six presents three methods for using the risk source model to help identify and respond to risks in a practical digital preservation setting, aligned with a standard risk management process. The first two use the model to support the identification stage of the risk assessment process, whilst the third builds on these to extend the approach to risk analysis and evaluation. The chapter reflects on how model and methods have been used at the British Library, before discussing them in relation to a preservation planning process. In this way, chapters five and six represent a solution to the research question of how the nature and complexity of digital preservation risk can be more thoroughly represented so as to support the

foundations for a more flexible yet comprehensive preservation planning risk response.

Chapter seven evaluates the research against the requirements stipulated in chapter four, as well as two other important sets of criteria associated with the research methodology. The first explores the construction of the outputs as design science artefacts, whilst the second considers the rigour of the research process as design science. The chapter demonstrates how the requirements have been met, reflecting the utility of the solution to answer the research question and the solid construction of the artefacts, as well as the rigour of the research process.

Chapter eight reflects on how the novel application of a risk science approach and production of a reference model for digital preservation risk answers the research question and represents a significant, original contribution to knowledge. It acknowledges the breadth, depth, and ambition of the research, with a new way of thinking about – and responding to – digital preservation risk. The solution contributes a one to two level capability increase as measured by the digital preservation maturity model of Kenney and McGovern (2005), and a clear improvement-based and prescriptive contribution to knowledge (Gregor and Hevner, 2013) that not only develops a new solution for the British Library and the wider community, but also provides a deeper understanding of the problem at hand.

As a practice-based PhD, this thesis is accompanied by a set of practical outputs. These represent the main artefact-based knowledge contributions made in the thesis, extracted and re-packaged for consumption and reuse by the wider community. Models are collated and published in the CHARM Reference Model for Conceptualising and Characterising Digital Preservation Risk, whilst Methods are presented in the CHARM How-To Guide, along with a series of suggested templates. These practical outputs are all published separately.

Chapter Two: Methodology

Introduction

The methodology used to progress this research blends interdisciplinarity with creativity to rigorously address the research question and better represent the nature and complexity of digital preservation risk. This chapter presents that methodology, the philosophical and developmental perspectives that may have implicitly influenced the course of the research, and the overall developmental approach. Clarification of epistemological assumptions contextualises and frames the research so that its outputs and conclusions can be understood in the context in which they were developed.

In creative research, new knowledge is generated through innovation, imagination, and original ideas. This thesis therefore represents neither qualitative nor quantitative research, and has no need for a typical data collection and analysis exercise. The methodology borrows from and integrates concepts and methods from other fields as appropriate to the research questions, the motivation for the research, and the quest for a practical output. The chapter presents these and outlines their application to the research before concluding with a reflection on the author's prior experience in the field to establish the personal contributory knowledge base upon which the research builds.

Methodological Background

Acquisition of an agreed disciplinary research paradigm is considered by many to indicate the maturity of a given scientific field (Kuhn, 1962, p. 11). Kuhn's definition of the term paradigm is open to interpretation (Dick, 1995, p. 223), though a broadly acceptable contemporary perspective is that of 'a set of basic and taken-for-granted assumptions which underwrite the frame of reference, mode of theorising and ways of working in which a group operates' (Saunders, Lewis and Thornhill, 2019, p. 140). Once defined, these assumptions provide the wider philosophical context for the work and help frame it, providing essential context in which the findings of the research are understood. Kuhn's use of the term paradigm

can also be interpreted not just as a set of underlying beliefs and values but also methods, in other words a 'disciplinary matrix' or methodology (Schwandt, 2001, pp. 183-184). Even more concisely, a research paradigm may be construed simply as 'the combination of research questions asked, the research methodologies allowed to answer them and the nature of the pursued research products' (van Aken, 2004, p. 220). In practice the terms paradigm and methodology might thus be used interchangeably.

There is as yet no standard or agreed disciplinary methodology for conducting digital preservation research. An explicit methodology presents and justifies the author's overall underlying assumptions, approach, and findings, so that the research and its conclusions can be understood in the context within which they were developed. In much of the research literature however, particularly in relation to certain types of assessment frameworks, methodologies are not always explicitly presented and the underlying rigour to the research or proposal is thus unclear (Maemura, Moles and Becker, 2017, p. 1619). Details on full methodologies such as McGovern's use of a constructive research methodology (2009), the overall philosophy and underlying assumptions identified by Barwick in their thesis on game preservation (2012), or the pragmatic methodological principles employed by Abrams (2023), are relatively infrequent. Ross (2012) acknowledges that the development of digital preservation knowledge would benefit from research that is 'more rigorous, methodologically founded, repeatable, verifiable, contextualized, and more effectively reported, that [...] could conform better to the "scientific paradigm' (p. 60). Arguably however, the field has yet to meaningfully explore and agree its research paradigm in substance.

Multiple paradigms or methodologies within a given field can nonetheless be valuable if one accepts the viability of different disciplinary epistemological beliefs, an arguably valid stance particularly within interdisciplinary fields. Feyerabend, for example, argues that 'anything goes' so long as it advances knowledge (Feyerabend, 1970, p. 26). From this pluralist methodological stance, each paradigm represents a 'different perspective on organisational reality' (Saunders, Lewis and Thornhill, 2019, p. 132). Viewed in this way, and particularly given the

interdisciplinary nature of digital preservation research, it is more important that the methodology is rigorous and clear than that it follows a particular form.

The methodology utilised in this research is presented here to demonstrate its rigour, the context in which the research was developed, and the assumptions that underpin it. As an interdisciplinary endeavour, the methodology borrows from and integrates concepts and methods from other fields that are appropriate to the research questions. The artefact and exegesis structure of the PhD is one such element. Artefact and exegesis PhDs are most common in the creative arts though the form has more recently gained traction for PhDs in wider fields including science, engineering, and medicine (Brabazon, Hunter and Quinton, 2022) as well as information systems research (MacKrell, McDonald and Gammack, 2017). The artefact-exegesis structure represents an inherently iterative methodology that weaves between development of artefact (practical output) and exegesis (thesis), with each iteration in one informing advances in the other. The underlying practical motivation for this research and the need for a usable, practical output make the practice-based structure a suitably appropriate approach.

The creative, practical and problem-solving aspects of this research were deciding factors in shortlisting appropriate methodologies and processes through which to develop the research. This led to formulation of the methodology first as a research framework and process, before subsequently framing that in light of epistemological perspectives. Three potentially viable research frameworks were considered, on the basis that each was either related to the problem field of digital preservation and risk management or supported the practical and creative element of the expected output. These were Management Science, Constructive Research, and Design Science.

Management Science research is focused on solving real world problems, particularly to support business related problem solving and decision making (Anderson *et al.*, 2009, p. 2). Research approaches for management science are often quantitative in nature, using mathematical probabilistic and deterministic models to analyse and evaluate different courses of action (Heinze, 1982), though modern management science research can be more diverse (Taherdoost, 2022).

Constructive Research is a creative methodology from the field of Management Accounting, developed to support problem solving not through mathematical and analytical modelling but through creation of new organisational procedures or models (Kasanen, Lukka, and Siitonen, 1993). Constructive Research builds on existing theory to generate new knowledge that can be applied in the form of a practical and usable solution. Design Science also represents a creative approach to problem solving and has a specific focus on development of products to attain goals and serve human purposes (March and Smith, 1995; Simon, 1996). Like Constructive Research, Design Science Research builds on existing knowledge to create new, innovative and practical artefacts that solve real-world problems. Design Science has trans-disciplinary application in domains ranging from information systems and management research to operations management, engineering, architecture, business, economics, and other information technology-related fields (vom Brocke, Hevner and Maedche, 2020a, p. 2; Dreschler, Gerber and Hevner, 2022).[7]

All three options are similar in nature, emanating from different fields but representing practical problem-solving methodologies associated with established and often similar processes. This has led to observations that constructive research and design science are essentially one and the same thing (Dresch, Lacerda and Antunes, 2015, p. v), with the latter also increasingly used to advance management science research (Ebneyamini, 2022). Of the three, design science represents the most flexible yet widely utilised and extensively cited methodology with applied relevance across a broad range of interdisciplinary endeavours. A Design Science approach was therefore selected as the core methodology through which to drive this research forwards.

The Design Science Methodology

Design science research is characterised by a blend of knowledge, creativity, innovation, scientific rigour, and practicality (Iivari, 2007; Hevner and Chatterjee,

[7] The Conference proceedings of the annual DESRIST conference on design science research further exemplify the broad use of DSR across various different domains.

2010; Baskerville *et al*., 2019; vom Brocke *et al*., 2020).[8] The development of new knowledge is a key differential between routine design and design science, with the latter building on and extending existing knowledge to create a newly innovative solution for a real world problem (Hevner and Chatterjee, 2010; vom Brocke, Weber and Grisold 2021). Design science research is based around the fundamental principle that the knowledge and comprehension of a design problem and a corresponding solution are acquired through the construction and use of an artefact, i.e. that which is man-made and of use to humans (Hevner *et al*., 2004; Iivari, 2007). Put most simply, the goal of design science research is the generation of new knowledge in order to demonstrably and practically solve human problems.

The tradition of Design Science can be traced back to the 15th century in the form of Da Vinci's creative problem solving processes (Dresch, Lacerda and Antunes, 2015, p. 51). More recently however, it is the work of Herbert Simon, winner of the 1978 Nobel Prize in Economic Sciences for his research into how people make decisions, which is more widely cited as the starting point for design science thinking. First published in 1969 with several subsequent updates and reprints, his seminal treatise on 'The Sciences of the Artificial' distinguishes between natural science and the science of the artificial, i.e. of human artifice, arguing the value of the latter in academic works and exploring the science of the design process by which such artefacts might be produced (Simon, 1996). A further significant milestone in design science thinking was reached in 1995 with Salvatore March and Gerald Smith's publication 'Design and natural science research on information technology' (March and Smith, 1995). Building on the work of Simon but explicitly using the term 'design science' instead of 'science of the artificial', they observe that whereas natural science seeks to explain and understand the world around us, design

[8] Design Science Research has been described as both a methodology and a paradigm. The Design Science framework used in this thesis for example is referred to as both a methodology and a paradigm by vom Brocke, Hevner and Maedche (2020b), Peffers *et al.,* (2006), and Venable, Pries-Heje and Baskerville (2017). Venable, Pries-Heje and Baskerville also reference several other authors that have published papers 'concerning DSR as a research method and paradigm (e.g. Hevner *et al*., 2004; March and Smith, 1995; Nunamaker, Chen and Purdin, 1990; Walls, Widmeyer and El Sawy, 1992)'. Baskerville (2008) and Weber (2010), on the other hand, explicitly assert that design science is not a methodology but a paradigm. It is not the place of this thesis to resolve this tension, but this mixed use of terms is nonetheless acknowledged in order to demonstrate that all perspectives have been considered in defining the thesis' overall approach.

science attempts to create valuable and utilisable products that serve human needs and requirements. The concern of design science is thus utility rather than truth (p. 253), the latter being predominantly the concern of natural sciences and of theory.[9] This focus on utility remains a defining factor of design science research today.

From a philosophical perspective, a design science approach is often one in which knowledge is developed to support action in an epistemologically pragmatic way (Romme, 2003, pp. 559 - 563). For a pragmatist, research 'starts with a problem and aims to constitute practical solutions that inform future practice', using different methods as relevant and appropriate to the research question in order to deliver practical solutions and outcomes (Saunders, Lewis and Thornhill, 2019, p. 146). Goldkuhl describes pragmatism as a 'research paradigm that is concerned with knowledge for action and change', observing that 'in a pragmatist perspective, knowledge is developed through a continual interplay between action and reflection' (2012, p. 92). Design science methodologies typically demonstrate this interplay with an iterative element in which solutions are built, evaluated, and refined to whatever extent is necessary.

Recent reviews of design science literature have identified significant levels of pragmatism in both foundation-level, conceptual literature about the field, and the practical approaches taken by researchers to design problem-solving solutions (Deng and Ji, 2018; Thuan, Dreschler and Antunes, 2019). Whilst other epistemological perspectives can also be identified in the literature (Iivari, 2007; Venable, Pries-Heje and Baskerville, 2017; Goldkuhl, 2020), pragmatism is thus a widely occupied epistemological stance for design science research. A pragmatic approach to development is arguably well-suited to an initiative in which the desired outcome is utility rather than absolute truth. Pragmatism is appropriate for a field in the early stages of development such as digital preservation, particularly where answers are sought to problems that are not always yet clearly understood. A pragmatic approach is also well suited for addressing so-called 'wicked problems'

[9] Winter (2008) interprets this as opposition to the inclusion of theory in design science research, though it is perhaps better positioned as an argument against employment of any one specific theoretical or reasoning method for design science developments.

in design science (Dreschler and Hevner, 2022, p. 9), i.e. those problems that are ill-formulated, with a myriad of interdependent factors, where the overall ramifications are somewhat confusing and where there is no clear or definitive answer (Buchanan, 1992, p. 15, citing Rittel, 1967). As indicated by the exploration of the research problem in chapter one, the question of how to resolve digital preservation risk arguably qualifies as a wicked problem.[10]

March and Smith proposed that, in practice, a design science research framework is built around two dimensions: Research Activities and Research Outputs. The *Activities* dimension represents four types of scientific endeavours that collectively represents the work required to develop and demonstrate an application of design science: Build, Evaluate, Theorise, and Justify. The Build and Evaluate processes are specific to design science research and demonstrate whether an artefact can be developed that addresses the problem space, and whether it does the job it was designed to do. The Theorise and Justify processes are natural science methods that explore the how and why or whether an artefact did or did not work in a given environment, as well as what might be learned going forwards.

The *Outputs* dimension represents the actual artefacts delivered by the research. March and Smith identified four main types of design science research outputs, otherwise known as artefact types: constructs, models, methods, and instantiations (also known as implementations). *Constructs* (or concepts) identify a shared vocabulary for a problem domain and establish the language by which both problem and solution are defined and presented. *Models* express relationships between constructs in terms of the problem/solution space, with their main concerns being utility and usability rather than truth. *Methods* are the sets of steps involved to apply a solution to the problem space, using constructs and models as appropriate. Finally, *Instantiations* operationalise and implement a solution into a system that addresses and satisfactorily resolves the initial problem.

[10] For more on wicked problems, see Rittel and Webber's 1973 paper, 'Dilemmas in a general theory of planning', which identifies ten defining characteristics of a wicked problem. These include the notions that 'there is no definitive formulation of a wicked problem', 'wicked problems have no stopping rule', and 'there is no immediate and no ultimate test of a solution to a wicked problem'.

There is a logical connection between all four types of artefacts, though the lines between each are often not clearly drawn. Winter, Gericke and Bucher for example, argue that models and methods may be considered two sides of the same coin depending on their manifestation. They suggest that reference models and generic methods (particularly when presented as procedural models) can both be considered a form of prescriptive design modelling (2009, p. 7). Similarly, the utility of a model presented without a corresponding construct to explain the terms used would certainly be questionable. Nonetheless it is not necessary to produce all four artefact types as part of a design science process. This is evident in much of the literature where the outputs do not represent all four artefacts but instead focus on only one or two, in particular models (Winter, 2008; Thuan, Dreschler and Antunes, 2019).[11] Researchers may thus focus on production of as few or as many artefacts as relevant to the problem/solution space defined, whilst remaining true to the underlying principles of design science research.

These four output types form the core set of artefacts proposed by March and Smith in 1995, though the role of an additional *Theory* artefact has since been explored as a valid design science research output (Gregor, 2006; Gregor and Hevner, 2013; Baskerville *et al*., 2018; Iivari, 2020). Definitions of theory vary from testable hypotheses to narratives, insight, enlightenment, and conceptual explanations (Gregor, 2006; Sarker, 2007; Wacker, 2008; Kuechler and Vaishnavi, 2012), though in essence theory fundamentally attempts to explain *why* things happen. From a natural sciences perspective, this represents the truth of why things are as they are. In design science, the nature of theory remains open to debate. Some (Gregor, 2006, p.628; Gregor and Hevner, 2013, p. 339) suggest that design science theory generates new knowledge that answers the question of *how* to resolve a problem before demonstrating that the proposed solution works. Iivari (2020) argues for a more critical and expansive interpretation, including the incorporation of theory into metarequirements and metadesign for an artefact, or the development of design theory based on solid kernel theories originating from

[11] In their analysis of over 100 design science research publications, Thuan *et al*. (2019) found that in over 70% of the papers only a single artefact was produced.

other disciplines.[12] [13] This debate aside, a better interpretation of theory from a design science perspective is perhaps 'how' rather than 'why'. Representations of design science research theories may be embodied in one of the four artefact types proposed by March and Smith or may have an independent manifestation with a form (such as a formula or a verbal statement) as appropriate to the type of theory explored.

The knowledge generated by design science research can take different forms, from descriptive and explanatory knowledge, to prescriptive design knowledge that demonstrates how something *can* be done (Winter, Gericke and Bucher, 2009; Gregor and Hevner, 2013; Dreschler and Hevner, 2022). Broadly speaking, design science artefacts that are mainly abstract or referential such as reference models and generic methods represent transferable prescriptive knowledge, whilst highly contextualised artefacts represent applied, situational knowledge. Both can represent valid new contributions to an established knowledge base. The significance of the knowledge generated is dependent on the level of maturity in the prior work upon which the new contribution builds. This prior work, or interpretations of the prior work, is described variously as kernel theory, justificatory knowledge, or pre-existing, practical and descriptive knowledge, not just as represented in documented outputs from a given field but also in the practical experience of practitioners (Kuechler and Vaishnavi, 2012; Gregor and Hevner, 2013; Knutas, Pourzolfaghar and Helfert, 2019; Dreschler, Gerber and Hevner, 2022). Whilst new, prescriptive or applied knowledge typically remains the primary knowledge output of design science research, the relationship between that and kernel knowledge indicates that both should be acknowledged in a design science research endeavour (vom Brocke, Hevner and Maedche 2020a; Dreschler and Hevner, 2022; Prat *et al*., 2022). The practical experience of this author

[12] Gregor and Hevner suggest that, in line with Merton's 1968 treaty on Social Theory and Social Structure, design theories are those which occupy a 'middle range' in the hierarchy of theoretical relevance - operating at a level beyond the hypotheticals of day-to-day research but falling significantly short of grand unifying theory that aims to explain everything.

[13] The literature on the nature of design theory is extensive. It is not within the scope of this research to explore it further here, instead readers are referred to Iivari's excellent and thorough review, 'A Critical Look at Theories in Design Science Research' (2020).

represents a significant amount of kernel theory and practical knowledge gained through experience, upon which is built the new knowledge represented in this thesis. The author's prior knowledge is therefore presented later in this chapter, followed in chapter three by analysis of literature and solutions from the wider field. Both contribute to the pre-existing knowledge that forms the foundations for this work.

Application of the Methodology

Several different versions of design science processes are available that support the design science framework, principles, and guidelines as outlined above. These are largely consistent with one another, though with minor differences across objectives, domain, and implementation methods (Venable, Pries-Heje and Baskerville, 2017, p. 7). The one selected as the basis for this research is the Design Science Research Process (DSRP) (Peffers *et al.*, 2006). This has achieved most prominence within the design science research community according to citations (Venable, Pries-Heje and Baskerville, 2017, p. 8) and aligns most easily with existing problem solving processes within the British Library.

The DSRP model is a six-stage process model in the following nominally sequential order, based on a standard problem solving approach:

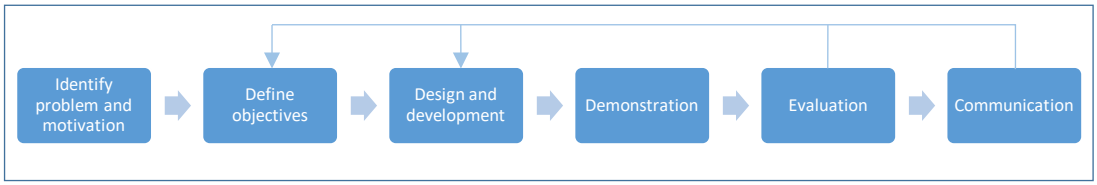

Figure 2: Simplified version of the DSRP model, adapted from Peffers et al. (2006)

The process is flexible. In theory, any stage can function as the starting point for the research and there is no expectation or requirement for researchers to proceed through stages in sequential order (Peffers *et al.*, 2006, p. 92). In such cases, scientific rigour is achieved by retrospectively addressing each stage and subsequently establishing relationships between their outputs, rather than working through them in sequence. Stage numbering is thus indicative rather than sequentially determinative. In stage one the specific research problem is defined and the value or importance of a solution established. Objectives are specified in

stage two, inferred logically from the problem definition, followed by determination of requirements and creation of the artefact(s) in stage three. The efficacy of the artefact(s) to solve the problem is demonstrated in stage four through their use in an appropriate activity, whilst stages five and six focus on evaluation and communication respectively. The flow of arrows in the diagram represents the potentially iterative nature of the process, whereby latter stages may lead back to either a review of objectives or changes to the design and development, or both. This recursive development may occur multiple times before delivery of a usable artefact or set of artefacts. Overall the process represents an iterative and hermeneutic endeavour through which knowledge is developed based on a recursively dialogical interface between the problem and the solution that leads to an enhanced understanding and coherence of both parts.[14]

The DSRP model is focused exclusively on design and build research activities. The other two activities proposed by March and Smith, namely theorise and justify, do not feature in the process model. These activities explore the question of *why* artefacts work (or not) within the scope of the research they were designed to support. This is explored to a degree in the evaluation chapter, though is not explored in significant detail as the focus of the research is on the production of practical outputs rather than theorising. It is nonetheless considered an opportunity for further research, as outlined in the conclusion.

Each stage of the research is described in this thesis in the locations below:

Stage	**Function**	**Location in thesis**
1: Identify problem and motivation	Define specific research problem and justify value of a solution	C1: Introduction and Research Problem Statement
2: Define objectives	Infer objectives from the problem definition and knowledge of what is possible/feasible	C3: Analysis, and C4: Requirements for a Solution
3: Design and Development	Establish requirements and create artefacts	C4: Requirements, and C5: Deconstructing

[14] For a consideration on the hermeneutical nature of design that predates much of the modern design science thinking, see Snodgras and Coyne (1997) *Is Designing Hermeneutical?* This relates to hermeneutic thinking in the philosophical or methodological sense, not the biblical sense.

		Digital Preservation Risk
4: Demonstration	Demonstrate use of artefact to solve one or more instances of the problem	C6: Demonstrating the Solution
5: Evaluation	Observe and measure how well artefact performs	C7: Evaluation
6: Communication	Communicate artefact to relevant audiences	C8: Conclusion, also publication of thesis and availability of artefacts

Table 1: DSR stages as represented in this thesis

The process was successfully used to develop a series of artefacts that represent a solution to the problem space and answer to the research question of how the nature and complexity of digital preservation risk can be more thoroughly and consistently represented so as to support the foundations for a more flexible yet comprehensive preservation planning risk response. These take the form of a glossary of terms, models of digital preservation risk, and a series of methods for using the models and assessing digital preservation risk. The artefacts are also represented in the external documents that represent the practical component of this PhD. An underlying, nascent theory on how to represent and explore digital preservation risk is embedded into the design of the models, whilst the utility of the models is demonstrated in the methods.

As this research is undertaken as a practice-based PhD, artefacts are represented in both this written thesis and the corresponding practical outputs. The distribution of artefacts across both elements is as follows:

Artefact type	**Artefact produced during this research**	**Location in thesis**	**Location in practical outputs**
Theory	Theory of how to explore digital preservation risk	C5: Deconstructing Digital Preservation Risk	Embedded in approach
Construct	Glossary of Key Terms	C5: Deconstructing Digital Preservation Risk; Appendix A	The CHARM Reference Model Glossary

Model(s)	Risk Models	C5: Deconstructing Digital Preservation Risk	The CHARM Reference Model document
Method(s)	Risk Identification Methods, Risk Identification Framework and Risk Assessment Template	C6: Demonstrating the Solution	The CHARM 'How-To' Guide; externally published exemplar method templates

Table 2: DSR Artefacts produced for this research and their locations

The artefacts were developed in an iterative and interactive manner, with each one providing insight into the other. Each iteration provided greater insight into not only workable forms of a solution but also the problem and its context. This is consistent with the established knowledge interplay between problem and solution in design science research endeavours (Dreschler and Hevner, 2022, p. 16). It is represented here in three main developmental phases, representing mainly stages three and four of the research process, demarcated by 'break points' in the research when limitations of a given approach were realised or reached.

The first phase used mainly experiential and descriptive knowledge to attempt simultaneous generation of both model and method, based around different OAIS functional entities such as ingest or preservation storage. The idea here was that risk assessments could focus on a specific stage or OAIS function, leading to a risk model oriented around these stages. The approach was based on first selecting a stage, then identifying risks relevant to that stage, describing these in a spreadsheet for assessment and identifying key parts of the risk descriptions that could be modelled. This phase did not progress beyond the design stage of DSRM as the process led to multiple inconsistencies in the descriptions of different types of risks both within and across different functional areas. As a result, it was not possible to generate a model that consistently represented risk across different functional entities. This experience suggested that the model and method should be defined in two different stages, separating the question of 'What is Digital Preservation Risk' from 'How can we represent Digital Preservation Risk'.

The second phase addressed this by using a risk science approach to establish a meaningful and contextualised definition of digital preservation risk, identifying different conceptual aspects of risk that could be further explored through a new risk model. This approach established the underlying theory of the research and introduced concepts such as target values and risk sources. These concepts formed the basis of a classification for different types of risk and aspects of risk in a typical digital preservation setting, leading to the first high-level version of a holistic digital preservation risk model. This addressed the shortfalls of the efforts in phase one and formed the basis of a new method for risk assessment. However, testing of the method indicated that the model needed more detail, particularly around the concept of risk sources, in order to consistently inform the production of risk assessments.

The third phase developed the concept of risk sources in more detail to explore and model the risk factors associated with each source and their potential manifestations (subsequently termed 'instances') in practical digital preservation settings. This took some time to finalise, particularly in terms of determining whether different aspects of risk should be considered a source, a factor, or an instance. Once the function of each entity type was clearly differentiated, a complete model was generated and populated using background and kernel knowledge from the author's experience and the literature review. Its final form was determined through an iterative process of generating risk statements, checking them for consistency, and revising terms or allocations where needed. This was then further developed into a method for using the risk source model in the form of a spreadsheet-based risk assessment, which was used in a series of different risk assessment scenarios. These tested different aspects of the model and proved the overall concept to be both usable and useful. A further method developed the model into a question-based risk identification framework. This translated it into a series of questions for non-expert users to explore the range of different risk sources that may need attention in a given scenario. It was a useful further test of consistency across the different entities and terms used, and led to minor refinements.

This iterative process led to production of the final versions of the practical outputs: the CHARM Reference Model and the corresponding 'How-To' Guide. Based on refinement of an underlying theory for representing risk, and having been devised and tested in a real world and practical environment, the results as presented in this thesis are thus considered to be both empirically and theoretically founded.

Author's prior experience

This research builds upon the author's existing knowledge, generated through a career of over twenty years in digital preservation. That knowledge contributed to the foundations for the research explored in this PhD and is therefore represented here as part of the background methodology.

The author's career in digital preservation started with the Dutch Digital Preservation Testbed (Testbed Digitale Bewaring) project (2001 - 2003), which explored the viability of different technical approaches for preservation of digital archival records (Potter, 2002). This work, inspired by a RAND report for the Dutch National Archives and Ministry of the Interior on carrying authentic, understandable and usable records through time (Rothenberg and Bikson, 1999), provided cutting-edge insight into the newly developing field of digital preservation. The project explored the impact of technical preservation approaches such as migration and emulation on the authenticity and integrity of four different types of archival content: emails, spreadsheets, text documents, and databases.[15] The author played a key role in this project, implementing the practical experiment process used by the project and contributing to production of the project's published recommendations on all four content types as well as other project reports.[16] The author's subsequent work with the EU co-funded ERPAnet (Electronic Resource Preservation and Access network) project (2003 – 2004) led to production of a series of UK and Netherlands-based case studies on organisational responses to

[15] The project website is no longer available, though project outputs can be retrieved from the Internet Archive at https://web.archive.org/web/20061010042318/http://www.digitaleduurzaamheid.nl/index.cfm?paginakeuze=185 (accessed 16 September 2023).

[16] The experiment process used by the project strongly influenced creation of the PLATO preservation planning tool released by the SCAPE project some years later.

digital preservation.[17] Case studies explored, amongst other things, organisational awareness of risks to digital longevity and practical actions taken (or otherwise) to address them. Case studies also provided insight into the lifecycle of digital content, a concept that was further developed whilst the author worked at UKOLN (2005 – 2008) for the Digital Curation Centre (Pennock, 2006a, 2007). Other research undertaken whilst at UKOLN focused on repository infrastructures and specific content types such as email and web records, considering not just general preservation good practice but also highlighting potential risks and stressing the importance of coordinated lifecycle management (Pennock, 2006b, 2006c; Pennock and Kelly, 2006).[18]

Since joining the British Library in 2008, the author's research has explored a more diverse range of digital preservation challenges across a broader set of digital collection types. These include web archives (Pennock, 2013), blog content (Pennock and Davis, 2009) and Twitter content (Hockx-Yu, Johnson and Pennock, 2012), handheld media content including software executables, games, audio visual content, databases and viruses (Pennock *et al*., 2016; Day *et al*., 2016; Pennock, Day and Samaras, 2019), eBooks and eJournals (Pennock and Day, 2018), emerging formats such as mobile eBook apps (Pennock and May, 2019) and web-based interactive narratives (Day *et al*., 2018). Practical experience at the Library not represented in the author's publications record has also involved sound content, different types of digitised materials from newspapers to monographs and endangered archives, born digital electoral registers and personal digital archives, and geospatial mapping data. This work made clear that similar risks exist across all aspects of the digital collection and across the lifecycle, from initial capture and acquisition to processing, storage, and preservation, but that the precise nature of the risk and corresponding mitigation action can vary from collection to collection

[17] The author led on the production of seven case studies in total and contributed to several more, though individual authorship of case studies is not attributed. Case studies are still available online at https://www.erpanet.org/studies/index.php, as is the set of interview questions used to collect the data for each case study.

[18] Technology Watch briefing papers included Fedora, DSpace and ePrints, all published in 2006. They since have been removed from the DCC website as they are out of date, though copies can be retrieved via the Internet Archive at https://web.archive.org/web/20091005005032/http://www.dcc.ac.uk/resource/technology-watch

and context to context. In short, whilst conceptual risks can be shared across collections and contexts, the manifestation of risk is highly contextual and a mitigation action for one type of content will not necessarily work for another.

Building on early Testbed experiences, the author has also initiated work at the British Library to further develop the British Library's conceptual understanding and practice of preservation planning. This involves a combination of desk-based research and practical activities to establish an understanding of the digital collection and different types of preservation risks that may affect it. Inspired by work from the National Library of Australia (Webb *et al*., 2013), the team maintains a series of Digital Collection Profiles that describe acquisition processes and clarify the preservation intent associated with different collections that must be supported through preservation planning (Day *et al*., 2014a; Day *et al*., 2014b; Day and Pennock, 2022). A criteria-based framework for format sustainability assessments has been devised and is in place to help identify potential risks that may influence a preservation planning or content acquisition process (Pennock, Wheatley and May, 2014). Risks and potential preservation planning issues for new types of content are also addressed through an ongoing and responsive programme of research, considering for example emerging formats such as mobile apps and interactive web-based narrative content (Day *et al*., 2018; Pennock and May, 2019). Most significantly in terms of institutional capability, the programme led to the conceptualisation and development of the award-winning Integrated Preservation Suite (IPS) project.[19] This is delivering a web-based platform that supports scalable preservation planning for highly diverse digital collections through a set of interlinked components including a software repository, a technical knowledge base, and a documentation database (Pennock and May, 2018; May, Pennock and Russo, 2019). This knowledge, alongside our experience with working at scale across multiple different digital content types, led to a policy around preservation planning as a response to risk-based triggers rather than the pre-emptive preservation planning often practised elsewhere (Pennock, 2020). This policy is

[19] Best Paper Award, iPRES 2019: 'The Integrated Preservation Suite: Scaled and automated preservation planning for highly diverse digital collections' (May, Pennock and Russo, 2019).

justified at the Library particularly due to the sheer scale and diversity of the collection, as significant amounts of resource may otherwise be expended on preservation plans to mitigate risks that may simply never manifest as expected.

The author's experience with preservation plans includes exploration of authenticity criteria, known also as significant properties (Rothenberg and Bikson, 1999; Wilson, 2007). These represent features or characteristics of intellectual content that convey meaning such as colour in a heat map, or italicised text to emphasise particular portions of a document. They can vary from object to object depending on the type of content and the nature of the features used. Pre-determined authenticity criteria or significant properties enable evaluation of a preserved object to assess whether or not changes introduced because of a preservation action (such as migration) are acceptable. Significant Properties are explored in the author's Testbed work as well as through contributions to the INSPECT project on significant properties (Potter, 2002; Knight and Pennock, 2009). They manifest at a high level in British Library's Digital Collection Profiles, which identify the main characteristics of collections that need to be preserved. These include for example structural relationships between cells in a data-centric collection, internal navigational functionality of an archived website, or intellectual content in journal articles (Day *et al.*, 2014a). It is clear from our experiences however, that detailing significant properties in a single collection object (such as an eJournal article) or small collection (e.g. all articles in a journal, or all journal issues in a title) is quite different from detailing them across a whole content type (for example all eJournals from a publisher in a given domain, from a single publisher, or a collection of eJournals from a range of publishers). Significant properties are likely to be significantly more expansive in the latter compared to the former. It is not yet clear precisely how this should inform a preservation planning process, aside from a suggestion that plans using significant properties for validation should be tightly scoped so that the number of evaluation properties remains manageable.

In addition to the conceptualisation of IPS, this author has also initiated and contributed to work on a number of other tools to support preservation. These go

some way to mitigate risks around, for example, failure to capture content, insufficient technical information about content, and failure to implement basic preservation storage requirements that may otherwise result in loss. The 'ArchivePress' solution, developed in conjunction with colleagues at the University of London Computing Centre (ULCC), represents a simple approach to capture of blog content by using RSS feeds that can subsequently be ingested to and preserved in a repository (Pennock and Davis, 2009). This functions as an alternative approach to archiving an entire website, in something akin to a 'the medium is not the message' solution. 'Twittervane' crowdsourced the selection of web sites for archiving in order to help mitigate the risk of selector bias in curated collections (Hockx-Yu, Johnson and Pennock, 2012). It worked by searching for specific terms related to a special web archiving collection in tweets from the Twitter Streaming API. URLs from tweets were extracted and expanded from their shortened form, with the most frequently shared URLs automatically input to the online selector tool for archiving. The 'CRISP' prototype was created to crowdsource and capture technical Representation Information (RI) to support preservation (Pennock, Jackson and Wheatley, 2012). CRISP was designed to operate as a community resource for representation information, providing a web-based form and Twitter handle that anyone could use to nominate a link to an online RI website. The website would subsequently be harvested by the British Library's web archiving tool for inclusion in a Representation Information database. Whilst enthusiastically received, take-up was low and it became clear that the Library would need to populate an RI database itself, which it is subsequently doing in the form of the IPS technical knowledge base. Most recently, the Minimum Preservation Tool solution (MPT) has been developed as a low cost alternative to a monolithic and costly digital repository (Pennock *et al.*, 2021). This publicly available tool supports configuration of basic preservation storage functions not usually met by a typical corporate IT environment, such as replication, check summing, and scheduled fixity checking, reducing the risk of loss in environments where these functions are not otherwise supported.

The author's involvement with large-scale digital preservation capability assessments has provided an altogether different perspective on digital preservation risk. An ISO 16363 self-assessment conducted by the author in 2015 painted an early holistic picture of the risk landscape. It identified a need for greater coordination at a policy and documentation level, including representation from other departments such as legal, finance, and IT Infrastructure.[20] A subsequent external audit of the Library's digital preservation capability for non-print legal deposit material highlighted the practical importance of file integrity assurance and timely workflows, alongside the importance of effective governance arrangements, clarity over roles and responsibilities, and the simple need for sufficient numbers of staff (Digital Preservation Coalition, 2017, p. 2).[21] Informal use of the Data Seal of Approval (DSA) assessment model in 2018 (Data Seal of Approval, 2013) provided further insight into mechanisms for evaluation, particularly in terms of maturity scales and scores.

These assessments illustrated the important role of organisational management practices in supporting and enabling digital preservation. Successful digital preservation practice does not exist in a silo: it must operate within the management practices of its parent organisation. Management activities undertaken by the author over a ten-year period as Head of the Digital Preservation department at the British Library provided critical insight into approaches that help embed digital preservation work into wider organisational activities. This has included everything from business case development and business planning, to recruitment, staff retention, and skills development (Pennock, 2018), as well as production and coordination of new Library policies and strategies (e.g. British Library, 2017). All of these activities represent different aspects of digital preservation risks – the risk of insufficient budget to support preservation, or failure to retain staff to operate systems, failure to keep skills up-to-date and manage systems, failure to ensure policy compliance, or failure to deliver strategic

[20] The report was not published externally but the experience led to production of a 'How-To Guide' shared at the International Digital Curation Conference (IDCC) (Pennock & Smith, 2016).
[21] This audit was commissioned collectively by the six UK and Ireland Legal Deposit Libraries to support a scheduled review of the 2013 Non Print Legal Deposit (NPLD) Regulations.

objectives. Add these to the more technically-focused digital preservation risks relating, for example, to bit-stream integrity, rendering inaccuracies, or simple loss of files, and the complexities of digital preservation risk management starts to become clear.

These activities illustrate the relationship between the author's prior experience in digital preservation and the work outlined in this thesis. These experiences, alongside responsibility for managing corporate digital preservation risk within the Library's strategic risk register, represent the author's pre-existing and practical knowledge base that underpins the research subsequently described in this thesis.

Chapter Three: Review and Analysis of Current Approaches

Introduction

Over the past thirty to forty years, digital preservation has transitioned from a predominantly archival uncertainty into an emerging interdisciplinary academic and applied field. Publications from both academics and practitioners, reflecting on or promoting developments in theory, practice, or tools, can be found in various, mainly online locations. This 'digital first' mode of communication is consistent with the digital focus of the field and typically takes the form of online magazines or journals, commissioned reports or project deliverables, and within peer-reviewed publications and conference proceedings.

This chapter begins with an overview of those sources, acknowledging the diverse ways in which digital preservation research is shared and illustrating the extent of the review undertaken in support of this research. It then focuses specifically on a review of relevant literature and approaches developed within the domain to date that variously explore different aspects of digital preservation risk. Many relevant frameworks have been developed and these are analysed in light of the wider literature to identify the current state of the art and the state of the practice in thinking about, managing, and responding to digital preservation risk. Critical analysis takes place at the end of the chapter, synthesising lessons learned, recurrent themes, and problem areas to be addressed by a new solution.

Knowledge sources and background

Dissemination of digital preservation research occurs across many different publication modalities, representing both the applied and interdisciplinary nature of the field. In addition to scholarly journal articles and conference papers, this review encompasses many different sources of research that may or may not have been through a formal peer review process. This includes conference presentations, online magazine articles, monographs, commissioned reports, project papers, and blogposts. The absence of formal peer review in such cases should not be taken to

indicate questionable quality, simply the emergent and developing nature of the field.

As an interdisciplinary, niche subset of archival, library, computer and information science, the field is relatively small and there is currently no dedicated, regularly issued and peer-reviewed journal for digital preservation. The closest equivalent is the International Journal of Digital Curation (IJDC), published 1 – 2 times a year, which hosts research articles and selected papers predominantly associated with the annual International Digital Curation Conference (IDCC). As a result, academic digital preservation publications tend to appear sporadically across a range of different disciplinary and peer-reviewed journals. These include Archivaria (the Journal of the Association of Canadian Archivists), Alexandria (the Journal of National and International Library and Information issues), JASIST (the Journal of the Association for Information Science and Technology), Archival Science, and the Journal of Documentation. Semi-formal online magazines and journals such as Code4Lib, RLG Diginews and D-Lib are also valuable sources of digital preservation articles, though their peer review process is not always wholly clear and the latter two are now defunct.[22]

Conference papers and presentations are a more consistent source of shared digital preservation knowledge. The iPRES international digital preservation conference series provides the most substantial, consolidated, and regularly accruing source of peer-reviewed digital preservation conference papers.[23] Annual publications of conference proceedings include the majority of accepted submissions including short and long papers as well as panel and poster descriptions. The Preservation and Archiving Special Interest Group (PASIG) meetings and IDCC also focus specifically on preservation and/or curation, though provide neither full peer-

[22] RLG Diginews ceased publication in 2007 and D-Lib in 2017. The D-Lib site is still online and accessible at https://www.dlib.org/dlib.html, whilst the RLG Diginews site can be retrieved from the OCLC website at https://www.oclc.org/research/publications/newsletters/diginews.html. Both can also be still accessed via the Internet Archive (IA) though some parts of the sites may be missing due to the way in which the IA captures website.

[23] The conference is usually known as designation 'iPRES', though the variation 'iPres' sometimes features in conference proceedings.

reviewed nor comprehensive published conference proceedings.[24] Papers on digital preservation occasionally feature in related peer-reviewed conference series, such as the ACM/IEEE Joint Conference on Digital Libraries (JCDL), the International Conference on Theory and Practice of Digital Libraries (TPDL), and the Archiving conference from the Society for Imaging Science and Technology. In addition, conferences and events organised by professional library and archives organisations such as the International Federation of Library Associations (IFLA), the International Council on Archives (ICA), the UK Archives and Records Association (ARA), and the Society of American Archivists (SAA), occasionally also feature relevant digital preservation presentations.

As an emergent field, a substantial amount of research is represented in project reports such as those co-funded by the European Commission, national funding agencies, and smaller independent research organisations. Both the European Commission and Jisc, for example, invested heavily in digital preservation research in the 2000's and funded several projects across the UK and Europe (Pennock, 2008; Strodl, Petrov and Rauber, 2011). These projects contributed substantially to the published literature of the field through the release of many project reports, as well as journal articles and conference papers.[25] The Research Libraries Group (RLG), Educopia, the National Digital Stewardship Alliance (NDSA), the Digital Preservation Coalition (DPC) and the Council on Library and Information Resources (CLIR) have also commissioned many reports over the past three decades that make valuable contributions to the field.

---

[24] IDCC presenters are on occasion invited to develop submissions into papers for publication in the associated International Journal on Digital Curation (IJDC), though the conference typically focuses significantly more on research data management than on digital preservation per se.

[25] Ironically, many of the original project websites are now unavailable. Concerns about the loss of project outputs from Jisc-funded project websites during the 2000s led to the archiving of Jisc-funded project websites by the UK Web Archive. This service is provided by the British Library and sites archived in this way are publically available from https://www.webarchive.org.uk/. There is no known comparable arrangement for websites from European Commission co-funded projects and many are no longer available on the live web. Brexit is one reason for this, as projects registered from the UK but using a .eu domain vanished overnight when the Brexit transition period terminated UK ownership of EU domains. The Internet Archive (IA) has copies of most of them, though they are often incomplete due to the particular way in which the IA crawls and captures websites. For more information on web archiving and the challenges of different approaches, see the DPC Technology Watch Report on Web Archiving (Pennock, 2013).

Blogs and blogposts provide a further glimpse into the work undertaken by practitioners, from institutions around the world. The DPC blog for example regularly features contributions from its membership of well over one hundred organisations, often on a weekly or twice-weekly basis. The Open Preservation Foundation (OPF) hosts another popular blog that features regular posts on the experiences of OPF members around the world. Both have featured blog posts relevant to and acknowledged in this thesis.

Finally, there is a small but growing number of digital preservation monographs. Two stand out for their award-winning contributions: 'Practical Digital Preservation', from Adrian Brown of the UK Parliamentary Archives (Brown, 2013), and 'The Theory and Craft of Digital Preservation' from Trevor Owens of the USA Library of Congress (Owens, 2018).[26] Both authors have significant experience working in the field at national-level memory institutions.

Publications from these different sources acknowledge and explore the many challenges of digital preservation that can pose a threat to the longevity of digital content. The archival field was amongst the first to raise the alarm, acknowledging uncertainty about how processes and principles to manage and preserve analogue content would translate to an intangible and transient digital environment (Hedstrom, 1984; U.S. National Archives and Records Service, 1984; Cox, 1992; Tylers, 1995 Bearman, 1989; Gavrel, 1990; Hedstrom, 1991). A new way of thinking was required and a new paradigm needed, in order to ensure that content generated in the newly digital world would be reliably managed and made available for future generations (Duranti, 1995; Hedstrom, 1995; Cook, 1997). This new paradigm became known as Digital Preservation.

Technological obsolescence was widely cited as an early concern that affected both archives and libraries (Rothenberg, 1995; Conway, 1996; Hedstrom, 1997/98). This

[26] Adrian Brown's 'Practical Digital Preservation: A How-To Guide' won the annual International Digital Preservation Award for Teaching and Communications in 2014 https://www.dpconline.org/events/digital-preservation-awards/digital-preservation-awards-2014, whilst Trevor Owens' 'The Theory and Craft of Digital Preservation' won the Association for Library Collections and Technical Services (ALCTS) outstanding Award Publication in 2019 http://www.ala.org/news/member-news/2019/02/owens-book-wins-alcts-outstanding-publication-award.

was variously described in relation to storage media, formats, software, and hardware, referencing concerns about how to access files over time and ensure that content could be displayed and experienced properly by users. Other dependency-related technological challenges were also identified, such as degradation and deterioration of storage media, though non-technical challenges were also acknowledged as important to truly ensure persistence over time (Garrett and Waters, 1996; Ross, 2000). These different challenges are variously reflected in a range of models developed within the field over the past two decades. Many of the models represent solutions to specific challenges, such as economic and cost models (Slats and Verdegem, 2005; Wheatley and Hole, 2009; L'Hours *et al*., 2014). Others function as more generic representations that represent challenges in association with specific processes, for example lifecycle management (Higgins, 2008). The three-legged stool model (McGovern, 2007) identifies three key areas of a digital preservation endeavour in which challenges can manifest: organisational infrastructure, technological infrastructure, and resourcing. Other domain-level models explore similar areas though with different structures, for example embedding resourcing as an organisational challenge alongside such matters as legal issues, policies, and governance. The wider literature of the field explores these variously as both challenges and risks. Regardless of their framing, all must be tackled as part of a holistic response to ensuring digital preservation.

Two key types of approach quickly emerged to address these challenges: risk management, and certification of digital archives. Several different frameworks have since been developed along these lines within the field, to address various challenges and risks and thus facilitate the longevity of digital content. The remainder of this chapter reviews these frameworks to chart their development and establish the current state of the practice in responding to digital preservation risk. In order to represent as many relevant solutions as reasonably practical, this chapter includes different types of approaches from risk management to threat modelling, maturity modelling, and audit/certification. A brief overview of each framework is provided and observations made regarding high-level similarities or

differences between them, as well as potential issues with their application and any sources of contention. This identifies the gaps in our current knowledge and capabilities that this research must address.

For the purposes of analysis and for ease of discussion, solutions are grouped into one of four categories. This categorisation emerged as a result of the analysis and is not currently an established classification structure; as such, it represents a new way to think about the focus of digital preservation risk assessment and management activities. The first is a format-based approach, in which a methodology supports identification, assessment and/or evaluation of preservation risks or threats associated with a given file format. The second is an object-focused approach whereby the objective is to understand risks or threats associated with a collection or type of object, in which a range of both technical and non-technical criteria are considered. These often include format-related risks but are not exclusive to them. The third is a system-focused assessment, in which a technical repository system is the primary subject of the threat or risk analysis.[27] The fourth is an organisation-focused assessment, where the goal is to consider the whole spectrum of preservation activities both organisational and technological. These typically have the end goal of demonstrating that a given approach meets expectations around best or good practice for digital preservation rather than to operate as risk assessment approaches. They are nonetheless included here as they are often considered to imply a degree of risk management.

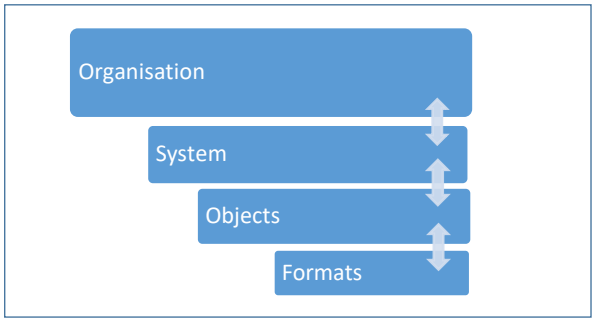

Figure 3: Hierarchical representation of assessment groupings

[27] The term 'repository' is used variously within the community to refer to both the technical infrastructure of a system (as in an 'institutional repository') and the memory organisation itself whose responsibility it is to undertake digital preservation (as in an 'archival repository'). This thesis cites works that use both definitions. Where the meaning of the term as cited is not obvious from the context, this is addressed within the text.

These four different types of approach successively represent something of a hierarchy: *organisations* manage *systems*, used to hold collection *objects*, represented by *formats*.[28] As a result, the lines between each type of assessment are blurred rather than definitive and criteria from one category may also be found in another, but viewed or defined from a different perspective.

Format-focused frameworks

Many of the risk assessment frameworks and models developed over the past few decades have a primary focus on the file format in which objects are encoded, exploring predominantly format-related risks including obsolescence. This section covers six such examples: the OCLC Format Durability framework (Stanescu, 2004), the File Format Evaluation framework from the Koninklijke Bibliotheek van Nederlands (Rog and van Wijk, 2008), the File Format Metadata Aggregator (FFMA) from the Austrian Institute of Technology (AIT) (Graf and Gordea, 2013), the British Library Format Sustainability Assessment framework (Pennock, Wheatley and May, 2014), the File Format Risk Assessment framework from the U.S. National Archives and Records Administration (NARA) (Johnston, 2018), and the Format Assessment framework from the Danish Rigsarkivet (Danish National Archives) (Skødt, 2020).

The earliest of these, the OCLC format durability framework (Stanescu, 2004), identifies six classes of risk, three of which relate to technological dependencies or obsolescence. Thirteen specific risks are also identified (though not explicitly in relation to the classes), assessment of which could help track a format's 'inevitable march to obsolescence' (p. 2). Proposed risks – also termed 'risk factors' - relate to, amongst other things, complexity, adoption, digital rights, licensing arrangements, and backwards compatibility. The framework advocates measurement of each factor using a probability/impact assessment, though the detail of such an approach is not fully explored and it is unclear from the literature review whether this framework was ever put into practice.

[28] This relationship might also be described the other way around: formats represent objects, held in collections, managed in systems, owned by organisations. The diagram thus represents the arrows as bi-directional.

The Koninklijke Bibliotheek van Nederland (KB-NL) approach, 'Evaluating File Formats for Long-term Preservation' (Rog and van Wijk, 2008), proposes twenty-four so-called 'sustainability criteria' against which to assess formats, grouped into seven classes. Classes cover similar ground to several of the risks proposed by Stanescu though with more detail, for example, digital rights management is broken down into five individual criteria covering password protection, copy protection, digital signature, printing protection and content extraction protection. The method proposes a scoring schedule and a weighting that favours standardisation, technological independence, adoption in the cultural heritage sector, and absence of compression. The higher the final score for a format, the better suited it is deemed to be for preservation.

The File Format Metadata Aggregator (FFMA) (Graf and Gordea, 2013) from the Austrian Institute of Technology (AIT) in association with several co-funded projects, also proposes a weighted approach to scoring file format risks. FFMA is comprised of twenty-one weighted risk factors, including some for which scores are automatically calculated by querying registries of information freely available on the web.[29] It is a rule-based system that results in a 'high', 'middle', or 'low' preservation risk rating for formats and indicates which features of a given format are the main source of risk. A revised version of the FFMA (Graf *et al*., 2014) increases the number of risk factors to twenty-eight and allocates an endangerment indicator rating to each factor, based on Ryan's academic study into perceptions of file format endangerment (2014). The highest indicators of endangerment include, for example, availability of rendering software and specifications, support for backwards compatibility, and legal restrictions, whilst factors around compression, viruses, and developer support receive a lower rating.

The British Library's format sustainability assessment approach (Pennock, Wheatley and May, 2014) avoids scoring and takes a more exploratory approach with a view to better understanding, rather than comparatively evaluating, potential risks of a

[29] These include the digital preservation service PRONOM from the UK National Archives, as well as DBPedia and Freebase, two general online sources of format and related technical information from outside of the digital preservation community.

given format within the BL. The framework supports production of 'evidence-based recommendations around use of a specific format, including whether or not a format is suitable as a Preservation Master' (p. 145).[30] It identifies ten main sustainability categories considered to represent areas of potential risks to format longevity and ongoing use, including development status, levels of adoption, software support and legal issues. The assessment framework remains in use today and informs – but does not determine - decisions about preferred formats and specifications for acquisition, as well as providing curators with insight into issues likely to pose challenges later down the road. Assessments are freely available on the Digital Preservation Coalition wiki for community consultation and re-use.[31]

The file format risk assessment framework from the US National Archives and Records Administration (NARA) assesses and scores formats against a set of sustainability factors, in order to identify preferred formats for preservation and propose migration pathways for formats that fail to achieve the desired score (Johnston, 2018). The first release of this framework includes thirty-nine criteria distributed across nine categories largely based on the framework of format sustainability factors from the US Library of Congress, including adoption, self-documentation, licensing, and digital rights management (U.S National Archives and Records Administration, 2023).[32] Criteria are weighted and scored, resulting in a final risk rating for each assessed format as 'high', 'medium', or 'low'.

Finally, the format assessment framework from the Danish National Archives (Skødt, 2019, 2020) assesses and scores twenty different criteria to enable comparison and subsequent selection of preferred formats for preservation. Criteria explore a similar range of characteristics seen in previous frameworks, from usage and documentation to interoperability, legal issues and compression, as well as more unusual metrics around support for significant properties, searchability and dissemination. Criteria are weighted, favouring support for 'good future prospects',

[30] The term 'Preservation Master' here refers to the best available edition of an object that meets both preservation needs and user needs, enabling creation of derived files with minimal loss.
[31] See https://wiki.dpconline.org/index.php?title=File_Formats_Assessments for assessments published to date. Work is currently underway to turn these into a dataset format that will enable automated integration of the assessments with the Integrated Preservation Suite project.
[32] See https://github.com/usnationalarchives/digital-preservation for publicly available assessments.

significant properties, and uncompressed formats. Descriptive text gives some indication on how to interpret these factors, though it is not always clear whether there is sufficient guidance to support assessors in allocating appropriately consistent scores.

Overall these six frameworks cover very similar ground, though with wildly varying numbers of criteria and the occasional 'wild card' not seen elsewhere. Both qualitative and quantitative approaches are used, with some also utilising scoring ranges, automatic allocation of values from online sources, and preferential weighting. A number of other resources are available online that might also be considered relevant to this section. These include the PRONOM database from the UK National Archives and the Library of Congress Sustainability of Digital Formats website, both of which use very similar criteria to many of those in the format assessment frameworks reviewed above.[33] The Library of Congress site for example provides the basis of information used in the NARA framework, whilst PRONOM also links formats to software, vendors, and potential migration paths. Both are thus a form of assessment frameworks, though arguably their goal is primarily to support knowledge sharing rather than to function as risk identification and assessment frameworks.

Object-focused frameworks

Object-focused frameworks provide another opportunity to explore format-related risks but in the context of a specific type, instantiation, or collection of objects, and alongside other potential risks associated with the wider environment in which the object or set of objects may exist. This section covers seven such examples: the British Library's Preservation Scorecard approach (Bennett, 1997), the CLIR migration assessment framework (Lawrence *et al.*, 2000), the Virtual Remote Control assessment framework from Cornell University (Kenney *et al.*, 2002; McGovern *et al.*, 2004), a British Library framework for risk assessment of handheld media (McLeod, 2008), the Simple Property Oriented Threat (SPOT) model

[33] PRONOM is available online at https://www.nationalarchives.gov.uk/PRONOM/ whilst the Library of Congress Sustainability of Digital Formats website can be found at https://www.loc.gov/preservation/digital/formats/.

collaboratively developed by OCLC, Florida Digital Archive, and Statistics New Zealand (Vermaaten, Lavoie and Caplan, 2012), the object validation approach from the German National Library (DNB) (Hein and Schmitt, 2013), and the change model from the PERICLES project (Waddington *et al.*, 2016).

The British Library has two examples in this category: the Preservation Scorecard and the Handheld Media Risk Assessment Framework. The Scorecard is a lightweight, simple methodology that reflects four key principles: avoidance of obsolescence, use of 'enduring' file formats, ensuring value, and capturing provenance (Bennett, 1997). Objects are scored across four categories - content type, format type, media type, and environment type - using a scoring framework that prioritises characteristics thought to represent good preservation potential. The British Library Handheld Media Assessment Framework represents another scored approach that identifies twenty-three risks across eight ranked categories, relating mainly to obsolescence, degradation, and policy failures (McLeod, 2008). Risks associated with storage media are ranked most urgent, followed by obsolescence of formats, hardware, operating systems, and then software. Policy-related risks are ranked least urgent. Application of the framework across a collection or set of collections indicates areas where the risks are greatest and thus can inform prioritisation decisions about where to focus resources.

The Simple Property Oriented Threat (SPOT) methodology takes a different approach in the form of an unscored threat model framework (Vermaaten, Lavoie and Caplan, 2012). Threat models are widely used in the area of cybersecurity and can be descriptive or graphical, though the literature on SPOT uses the term 'threat' and 'risk' interchangeably without differentiation between the two.[34] The SPOT model identifies six properties deemed to represent successful digital preservation – namely availability, identity, persistence, renderability, understandability, and authenticity - and suggests a number of threats to each property that, if not properly managed, may affect the longevity of objects in a collection. Threats relate (for example) to hardware and software availability, media management, metadata

[34] A descriptive model is helpful, for example, for defining different types of threats, whilst a graphical model can function as an effective way to illustrate relationships between threats.

practices, and knowledge of the user community, though wider contextual threats such as economic or legal matters are explicitly out of scope. The literature briefly describes how the model might be used to support risk assessment in two of the authors' institutions, though with insufficient detail to confirm its rigour or reusability.[35]

Taking a wholly different direction, a workflow-oriented approach from the Deutsche Nationalbibliothek (DNB) automatically assesses files during ingest against just five hierarchical criteria: file integrity, format identification, restrictions (e.g. digital rights management), metadata extraction, and format validation (Hein and Schmitt, 2013). Files must 'pass' the first check (file integrity) before proceeding to the next, and so on. The more that are passed, the greater the preservation level and 'risk management probability that the deposited publication can be preserved' (p. 315). The authors note that many files achieve only level two, observing a level of immaturity in the capability of tools used. These may have improved since the paper was published though no update was located in the course of the literature review.

The Virtual Remote Control (VRC) project from Cornell for risk assessment of web sites takes a much wider perspective, considering not just the object itself but also the wider physical, technical and organisational context in which the site is hosted (Kenney *et al*., 2002; McGovern *et al*., 2004). This acknowledges that objects do not exist in isolation, that their dependencies are greater than those associated with the file format in which they are primarily encoded, and that organisational efforts are required to undertake the management process. The solution proposes not a set of criteria as seen in most other frameworks, but a series of subjects for consideration. These range from server vulnerabilities to patching frequencies, software maintenance, backups, environmental controls, and geographical threats, as well as risks relating to linked but externally hosted websites. A sequence of technological and organisational steps are defined and mapped against a standard

[35] Staff at the British Library attempted to use the SPOT model prior to developing their new format sustainability assessment approach but found it insufficiently detailed, leading to uncertainty about what to assess and how to manage overlaps between threats across different properties.

risk management process to demonstrate a suggested implementation method. The change monitoring approach from the Pericles project takes a similarly wide perspective, with the underlying premise that change in a digital object's 'ecosystem' is a risk to the object's longevity and reusability (Waddington *et al.*, 2016). It identifies seven types of changes, including those relating to semantics, policy and requirements, as well as technology and dependencies. However, the risk aspect of the solution appears under-explored and it is unclear from the available project literature how the model may align with a risk assessment or management process.[36]

The final framework in this section supports assessment of risks to an object during a file format migration (Lawrence *et al.*, 2000). This explores a broad range of organisational risk categories such as staffing, legal and costs, alongside more technological risks from hardware/software dependencies to digital rights management, and fixity. It provides a clear method for undertaking the assessment, in the form of a series of questions on a range of subjects. These explore, for example, the relationship between target and source formats, the availability of conversion software, systems used to hold objects, metadata availability, and security measures. Resulting risks are assessed in terms of Probability/Impact, though the framework cautions against generating single scores for decision making on the basis that the probability of risk is hard to quantify, and risk-measurement scales - like risk definitions - are highly contextual (p. 24, citing Williams, Walker, and Dorofee, 1997). This focus on assessing risks during a migration is perhaps an early instantiation of a risk-based preservation planning approach.

Almost all of these object-focused frameworks demonstrate greater awareness of the wider collecting environment and its implications for digital longevity than those that focus predominantly on formats. They represent a broad range of different approaches and methods, from questionnaires to automated workflows,

[36] Use of the model to preserve software-based art is explored in a 2016 iPRES conference paper (Rechert *et al.*, 2016), though the focus is primarily on a technical assessment of the object rather than a structured risk assessment process.

and although they frequently cover similar ground, they often use different terms and terminology for the threats and risks they identify.

System-focused frameworks

System-focused frameworks are framed around the technological system used by organisations to manage and preserve objects. They consider a wide range of potential problem areas at both format and object level, mainly from a technological perspective though with some considerations for organisational challenges that may impact on a system if not addressed. Relatively few assessment frameworks have been devised at this level and only two examples are considered relevant here: the LOCKSS Threat Taxonomy (Rosenthal *et al*., 2005), and a risk management approach from the Portuguese research, development and innovation centre INESC-ID (Barateiro *et al*., 2010).

The LOCKSS (Lots of Copies Keeps Stuff Safe) Threat Taxonomy (Rosenthal *et al*., 2005) identifies thirteen threats that may adversely affect an organisation's ability to manage its digital preservation system so that its contents remain available to users over time. These include service failure and obsolescence in various forms, as well as threats around use or abuse of the system, natural disasters, and organisational/economic failure. It explores each one in terms of a threat/requirements pairing, in which each threat is countered by one or more suggested requirements. It thus operates primarily as a 'bottom up approach' to system design rather than a risk or threat assessment model, though it covers several of the same core technological risk criteria seen in the frameworks discussed above.

Similar ideas about the relationship between threats and requirements are evident in the methodology from the Portuguese research, development and innovation centre INESC-ID (Barateiro *et al*., 2010). This establishes a taxonomy of sixteen vulnerabilities and threats that may affect an organisation's ability to meet a series of essential digital preservation requirements, proposed as reliability, authenticity, provenance, integrity, obsolescence avoidance, scalability, and heterogeneity. Vulnerabilities represent weaknesses in the technological environment such as

service failures, faults, or obsolescence, whereas threats represent adverse events such as natural disasters, organisational failures, or legislative changes. The proposed method of implementation is through an ISO 31000 risk management process, though this is not specifically described in relation to the proposed threat model.[37] The LOCKSS and the INESC-ID frameworks both explore very similar risks, though use different terminology to describe them.

Organisation-focused frameworks

The fourth category of solution addresses the final layer of the classification hierarchy, that of the preserving organisation. Threats, risks, and areas of concern represented in these approaches are typically the most comprehensive of the four categories and represent the full suite of format, object, technological, and organisational challenges to digital longevity. The 1996 Task Force Report on Preserving Digital Information played no small role in establishing this type of approach as a response to the challenge of digital longevity, concluding that a process of certification for digital archives was needed 'to create an overall climate of trust about the prospects of preserving digital information' (Garrett and Waters, 1996, p. 24). Seven examples are explored in this section, representing a range of risk assessment, maturity models, and audit/certification approaches.[38] These are: the Trustworthy Repositories Audit and Certification (TRAC) process (CRL, 2007), the Digital Repository Audit Method Based On Risk Assessment (DRAMBORA) approach (DCC and DPE, 2007), ISO 16363 for Audit and Certification of Trustworthy Digital Repositories (ISO, 2012b), the Core Trust Seal (CTS) approach (CoreTrustSeal, 2020), the nestor SEAL for Trustworthy Digital Archives (nestor, 2013), the DiAGRAM tool from the UK National Archives (Barons *et al*., 2021; The National

[37] This work is further explored in the EU co-funded TIMBUS project, scoping an 'enterprise-level' approach that integrates digital preservation risk management into wider organisational risk management processes (Barateiro, Burda and Simon, 2013). Project outputs suggest a 'Holirisk' system was designed for risk assessment that aligned with the ISO 31000 risk management process (Strodl *et al*., 2013; TIMBUS, 2014), though the nature of the underlying risk model is unclear beyond a suggestion that it uses DRAMBORA criteria (see next section). The Holirisk tool itself no longer appears available and this author has unearthed no evidence that it was ever put into practice.

[38] The term 'certification' is straightforward to understand; the term 'audit' on the other hand has different practical associations in different disciplines. For the purposes of disambiguation, within digital preservation it is typically associated with conformance assessments, i.e. to determine whether a given environment or approach conforms to a set of requirements or specification.

Archives, 2023b) and the Rapid Assessment Model (RAM) from the Digital Preservation Coalition (DPC, 2021).[39]

The Trustworthy Repositories Audit and Certification (TRAC) approach (Task Force on Digital Repository Certification, 2007) represents the first substantive assessment framework produced in response to the 1996 Task Force Report. Its influence is evident in many of the subsequent frameworks, particularly its criteria and their organisation into three main sections: organisational infrastructure, digital object management, and technologies/technical infrastructure/security. Each contains several sub-categories, for example financial sustainability and organisational structure/staffing (organisational infrastructure), ingest, preservation planning, and archival storage (digital object management), and systems infrastructure (technologies/technical infrastructure/security). Criteria are expressed as requirements with which an organisation should comply in order to demonstrate it understands the risks and threats within its systems. The checklist can be used independently for self-assessment or as part of a more formal audit process. Overall, it functions as a tool through which to evaluate whether the repository (used in this context to refer to the organisation) understands and is responsive to the threats and risks that may prevent it from achieving its goal of long-term, reliable access.

ISO 16363 (ISO, 2012b) addresses a similar set of criteria (termed metrics) grouped across three very similar categories: organisational infrastructure (covering administration, staffing, legal and financial functions), digital object management (the processes through which content is acquired, ingested, preserved and made available) and, infrastructure and security risk management (specifically as regards the technical infrastructure). It has over one hundred individual metrics, more than in TRAC, expressed as requirements that, once addressed, imply mitigation of anticipated digital preservation risk. Formal certification of ISO 16363 compliance is supported by a third party associated with development of the standard. Self-

[39] Several other assessment frameworks have been produced that are not reviewed here as they have not reached the level of prominence or visibility achieved by those featured. A good summary of these is available in 'Organizational assessment frameworks for digital preservation: A literature review and mapping' (Maemura, Moles and Becker, 2017).

assessment is encouraged prior to certification, though the overhead associated with both is significant, as observed in an iPRES 2019 panel discussion chaired by this author (Giaretta *et al.*, 2019).[40] Certification is a pass or fail process and this author has identified only three organisations that are known to have yet sought or achieved certification despite the standard's availability for almost a decade.[41]

The Core Trust Seal methodology (CoreTrustSeal, 2020) originates from the research data community and replaces the assessment framework previously known as the Data Seal of Approval, from which it has evolved. It assesses an organisation across a set of requirements (or characteristics) relating to broadly the same three areas as TRAC and ISO 16363: organisational infrastructure, digital object management, and technology. With just sixteen main requirements it has significantly fewer than either TRAC or ISO 16363, though each individual requirement is associated with additional criteria that the repository is expected to address in its response. These explore strategic issues associated with mission and scope, through to legal issues and governance, as well as practical issues relating to integrity and security. The assessment process rates responses against a predefined scale of 0 – 4 to indicate varying levels of compliance, a structure noticeably different from TRAC and ISO 16363 which are both more qualitative. Certification typically requires a compliance rating of 4 across all criteria, though a score of 3 may be considered acceptable in some areas. As of September 2023, the CTS website lists almost ninety different organisations as Core Trust Seal certified data repositories.

Two other frameworks in this section also use a scored scale: the nestor SEAL for Trustworthy Digital Archives (nestor, 2013) and the DPC Rapid Assessment Model (RAM) (DPC, 2021). The nestor SEAL was developed in line with the German standard DIN 31644 for Trustworthy Digital Archives. Criteria cover largely the same ground as Core Trust Seal but with the addition of several criteria more clearly

[40] Presentations and notes from this panel session are available from the British Library Shared Research Repository at https://bl.iro.bl.uk/concern/conference_items/6d741c87-214f-4fe2-a718-e47b4bc124c4?locale=en. This observation is also the result of this author's own experience from carrying out an ISO 16363 self-assessment exercise for the British Library in 2015.

[41] These were identified during the panel discussion at iPRES 2019. Details of certified organisations do not appear to be centrally available from the PTAB website.

aligned with ISO 14721 (with which ISO 16363 is also aligned), taking the total number of criteria to thirty. The DPC-RAM has just eleven criteria that represent a mix of organisational capabilities (including provision of technology services) and service capabilities (primarily focused on object-level processes). It is widely promoted by the DPC, with member organisations encouraged to use the framework in order to inform both their own and the DPC's support activities. RAM aims at self-assessment rather than certification, whilst the nestor framework can be used for both self-assessment and certification. Ultimately however, CTS, RAM and nestor all cover very similar ground. [42]

All of these frameworks are structured around a set of requirements, metrics, or criteria, which support a high-level objective around ensuring organisational capabilities support preservation or demonstrate trustworthiness. Risk management may be implied or required by one or more metrics, though none focus primarily on risk management. The remaining frameworks in this section take a more explicit risk-oriented approach and specifically aim to support risk management through their implementation.

DRAMBORA (Digital Repository Audit Method Based On Risk Assessment) is comprised of seventy-eight potential risks and possible mitigations, grouped across eight categories: organisational management; staffing; financial management; technical infrastructure and security; acquisition and ingest; preservation and storage; metadata management, and; access/dissemination (DCC and DPE, 2007). Each risk is individually described though in somewhat inconsistent and indirect terms, for example 'finances are insufficient to adequately resource each of the business's integral activities' (p. 159), 'repository is legally accountable for either failing to fulfil responsibilities or acting beyond the scope of what is permissible, as detailed in legislative instruments' (p. 151), and 'individuals with roles, responsibilities or aptitudes vital to the achievement of business objectives part

[42] The DPC guidance note on bitstream preservation (Wheatley, 2022) takes a different approach from RAM and specifically advocates a risk driven approach, listing nine risk/threats that it considers relevant to storage including bit rot, storage failure, and natural disaster.

company with the repository, rendering the achievement of those objectives less straightforward' (p. 155).

DRAMBORA risks are further explored in the PORRO ontology for digital preservation risk management (McHugh, 2016), developed as a result of comparing completed pilot DRAMBORA audits against high level conceptual areas of TRAC and the Open Archival Information System (OAIS). PORRO has over six hundred related elements that collectively represent institutional requirements, object-centric characteristics, related elements and associated risk cause or effect factors (McHugh, 2016, p. 105). Despite the extent and ambition of this work, the completed ontology is overly complex with over one hundred individual preservation goals and insufficient clarity on how to map and manage their various relationships. The DRAMBORA process meanwhile is somewhat clearer, aligned to ISO risk management standard 31000. It identifies context and objectives, before assessing and evaluating risks using a time-bound probability/impact matrix prior to subsequent and ongoing management in a risk register.[43] The suggestion however that all risks should be assessed (pp. 75 – 76) requires a significant investment of time and effort.

The Digital Archiving Graphical Risk Assessment Model (DiAGRAM) methodology from the UK National Archives and University of Warwick (Underdown, 2019; Merwood, 2020; Barons *et al*., 2021; The National Archives, 2023) takes a wholly different and statistical approach to risk assessment. DiAGRAM is a quantitative framework with an underlying Bayesian statistical model that determines probabilities from expert discussions following the IDEA (Investigate, Discuss, Estimate, Aggregate) protocol. The model is comprised of a series of 'nodes' that each represent a risk and may affect an organisation's ability to successfully preserve digital content, specifically in terms of its renderability and interpretation of intellectual content. Nodes include integrity, content metadata, technical metadata, obsolescence, tools to render, and storage life, though the model and project literature appears not to define these risks in more qualitative terms. The

[43] The time-bound nature of risk is measured in the probability axis of the matrix, with risk horizons ranging from occurrence every month to once every hundred years or more.

tool is largely focused on determining whether an organisation complies with expected good practice, guiding users through a set of questions and using their answers to determine risk levels for different aspects of their preservation capability. The project has received positive feedback in dissemination activities, though the project evaluation report indicates some reservations from interviewees about the practical usefulness of the tool for managing digital preservation risk (Mitcham, Currie and Kilbride, 2021, p. 38).

The assessment frameworks in this section share many similarities though with some exceptions. Many require a significant investment in time and effort to explore the entire organisational domain. If seeking formal certification, this can require collation of relevant documentation as evidence, again a time-consuming activity.[44] They cover broadly similar concerns, though in different ways and to varying degrees of granularity. Collectively, they represent a widely accepted and relatively popular approach to assessment, though whether they function well as risk assessment frameworks is another matter, particularly given that many of them focus on requirements to achieve an organisation's goals and mitigate risks rather than the risks themselves.

Analysis

The frameworks explored in this chapter, regardless of whether they focus on formats, objects, systems, or organisations, represent a broad and varied mix of different types of assessment techniques. Some originate from research projects or commissioned reports, others are practitioner-led. They may be qualitative, quantitative, or use mixed methods to assess risks. Some use weighted factors or binary yes/no judgements with no weighting. Others involve solicitation of expert judgement, for example using Delphi panels or Bayesian statistical analysis. Some result in scores, others result in 'risk levels' or reports. Some use open-ended questions that require submission of evidence to expert panels, others use a self-auditing approach. Often they incorporate different combinations of techniques.

[44] David Rosenthal estimated that production and collection of documentation for the LOCKSS TRAC audit 'consumed between two and three person-years of senior staff time' (Rosenthal, 2014). This is a significant overhead, especially for small institutions.

Collectively they represent the interdisciplinary and exploratory nature of digital preservation and its stakeholders. They are all different, though they all identify criteria that are meaningful for digital preservation and in some way relate to a risk, threat, requirement or vulnerability - although they often neither define these terms nor differentiate between them.

No one technique is necessarily better than another and each can have value in a given context. From the same perspective however, each can also have its drawbacks and there is a degree of contention within the community about certain aspects of some assessments. Dissension stems mainly from two areas, particularly insofar as format assessments are concerned: uncertainty about format obsolescence, and concerns about the suitability of weighted scoring. As criteria relating to formats and format obsolescence prevail in all types of assessment framework, and scoring approaches are similarly widely used, discussion on these points is relevant many of the other frameworks as well.

Risk factors in the format frameworks do not always clearly relate to obsolescence, though obsolescence is often cited as a primary driver for format-level risk management. The sheer number of factors represented in different format assessment frameworks varies wildly and this variation indicates a degree of disagreement over whether many of the factors are truly meaningful indicators of risk and obsolescence. Ryan's academic study into perceptions of 'endangerment' (2014) concludes that just three measures exceed the 'emergent threshold level', namely the availability of rendering software, format specifications, and community or third party support (p. iv). Format assessment frameworks from practitioner institutions however consider significantly greater numbers of criteria (Pennock, Wheatley and May, 2014) (Johnston, 2018) (Skødt, 2020), both before and after publication of Ryan's study.

The relatively loose definition of obsolescence used by many in the community does not help matters. In general, the term obsolescence implies that something is out of date and no longer commonly used or supported. Pearson and Webb (2008) define format obsolescence as 'the state of becoming obsolete' and suggest that a file format becomes obsolete 'when access is no longer possible' (pp. 93 – 94).

Rosenthal (2010) suggests that obsolescence occurs when ‘a new version […of an application…] could not read files written by the old version’ (p. 197). Yet what is obsolete for one institution may not be obsolete for another, as each may have access to different resources, infrastructures, skills and so forth. This leads to a suggestion that it is more meaningful to speak in terms of ‘institutional obsolescence’ than obsolescence in general (Graf and Gordea, 2013; Pennock, Wheatley and May, 2014; Digital Preservation Coalition, 2015b), meaning that the technology in question is no longer in use or easily accessed by a particular institution. This contextualisation of obsolescence, which is often otherwise absent from discussions, suggests that measures should always be assessed in context and that the transferability of assessment results may be limited. This is a matter for individual assessors to consider, though the contention around obsolescence is more than just semantic.

Some experts contend that format obsolescence is simply not such a significant problem as first envisaged (Rusbridge, 2006; Rosenthal, 2010; Jackson, 2012). The reasons for this relate largely to market maturity since the late 1990’s and the rise of the Internet, both of which have led to greater and more acceptable degrees of interoperability between formats and format versions. A large-scale analysis of data held at the UK Web Archive led Jackson (2012) to conclude that ‘most formats last much longer than five years, that network effects appear to stabilise formats, and that new formats appear at a manageable rate’ (p. 158). Both Rosenthal and Jackson observe, correctly, that the bond between format and software is often not a 1:1 relationship and that in a modern computing environment, many contemporary formats are supported by multiple different software applications. The question remains however, whether those different applications open and render the file accurately.

Different software applications can render the same file in different ways, affecting the perceived authenticity of the object as well as the reusability of its intellectual content. There is relatively little published research to illustrate this, though one of the best works is the 2012 ‘Rendering Matters’ report from Archives New Zealand

(Cochrane, 2012). The practical tests undertaken with over one hundred different office files led the author to conclude that:

> *'The choice of rendering environment (software) used to open or "render" an office file invariably has an impact on the information presented through that rendering. When files are rendered in environments that differ from the original then they will often present altered information to the user. In some cases the information presented can differ from the original in ways that may be considered significant.'* (Cochrane, 2012, p. 4)[45]

Cochrane's investigation indicates that accurate rendering of content can be an issue regardless of whether a format is termed obsolete or not. To focus therefore on obsolescence, particularly when the evidence is unclear and there is no agreed community definition of the term in place, runs the risk of overplaying the relative importance of format obsolescence risk. As we noted at the British Library in 2014, 'the relatively 'fuzzy' nature of a file format requires [...] a nuanced understanding of preservation risk that does not solely lie with 'all-or-nothing' format obsolescence' (Pennock, Wheatley and May, 2014, p. 141).

The second - and related - controversial element is scoring. The report by Lawrence *et al*. (2000) cautions against generating single scores for decision making on the basis that not only is the probability of risk hard to quantify, but risk-measurement scales, like risk definitions, are highly contextual (p. 24, citing Williams, Walker, and Dorofee, 1997). Moreover, van der Knijft (2013) observes that the criteria used in most assessment frameworks are based largely on theoretical considerations rather than empirical data on their relevance, as are the scores and weighting often applied to them. This leads them to ask:

> *'What exactly is the point of classifying or ranking formats according to perceived preservation risks, if those risks are largely based on theoretical*

[45] This is consistent with many of the unpublished findings of the Dutch Digital Preservation Testbed project from the early 2000s, in which this author had the role of Experiment Operator. The project carried out experiments to assess how different preservation approaches performed with different types of files, undertaking manual comparisons of source rendering and target rendering on screen. Like the Cochrane 'Rendering Matters' investigation (2012), it found that different software applications rendered files in different and not always predictable ways.

> *considerations, and are so general that they say next to nothing about individual file format instances? Isn't this all a bit like Searching for Bigfoot?'* (van der Knijft, 2013).

He provides a rare example of how a scoring method led his institution to select a file format that later resulted in several practical problems. None of these was apparent from the format assessment, nor could he see how they easily could have been.

The British Library has first-hand experience of other problematic criteria, in particular those that relate to format validation. The practice of format validation stems from a suggestion that instances of file formats are most likely to render properly in the future if they are highly conformant with the file format specification.[46] The availability of validation tools or the ability to 'validate' a file against its published specification is therefore a common criterion in frameworks. The Library observes however that:

> *'there are a number of documented (and anecdotally many more undocumented) examples of PDF migration implemented to ensure JHOVE provided a "valid and well formed" validation result for each preserved file, where there was little or no evidence of the need to take action given the tolerance of PDF viewers to many of the issues JHOVE identifies.'* (Pennock, Wheatley and May, 2014, pp. 143-144).

The German National Library makes a similar observation, observing a high degree of viewer tolerance in practical tests and leading them to conclude that 'this validity problem is negligible at present' (Hein and Schmitt, 2013, p. 317). A requirement for valid files can therefore introduce new and immediate problems when well-intentioned preservation action is undertaken to produce a valid representation of an otherwise invalid file and inadvertently changes the bitstream in a way that

[46] Although the origins of this suggestion are poorly documented, the author was party to discussions in the mid – late 2000s when the practice of format validation was proposed as part of the EU co-funded PLANETS project. It subsequently became a standard element of the object characterisation process: identify, validate, and extract metadata. For more details on this process see Brown (2007).

affects the intellectual content when rendered on screen.[47] One might argue that use of the correct software with valid files would negate this problem and address Cochrane's concerns about accurate and authentic rendering. The Library's experience however is that even valid files may display inaccurately when using rendering software otherwise believed to be correct (Pennock and Day, 2018). The value of comprehensive format validation is at present somewhat uncertain (Wheatley, 2018), casting doubt on the actual risk associated with preservation of 'invalid' files and any related scoring. It is, overall, a complex picture, highly affected by different contexts and resource availabilities. In general however, it suggests that format-based risks are not yet sufficiently understood and that a degree of caution is needed when scoring, particularly if generating scored assessments for comparative decision-making.

This concern over scoring and the relevance of different criteria is representative of a wider uncertainty about *what exactly constitutes a digital preservation risk*. The frameworks above are all, without exception, heavily reliant on digital preservation theory rather than an evidential need, though practical experiences at the British Library, the National Library of the Netherlands, and the German National Library suggest that theories about format-based risk do not necessarily translate well into a practical environment (van der Knijft, 2013; Hein and Schmitt, 2013; Pennock, Wheatley and May, 2014). There is of course a time horizon within which most risks inevitably eventually manifest, but this is often challenging to reasonably or usefully estimate. Altman and Landau (2020) observe that 'while a number of good practices are recognised for digital preservation, many of these practices are heuristic' (p. 2). Furthermore, and in relation specifically to organisational risk approaches, they observe that:

> *'There is little specific guidance or empirically-based information on selecting specific preservation strategies that fit a curating institution's risk-tolerance, threat profile, and budget'* (Altman and Landau, 2020, p. 2).

[47] This has been observed and noted in unpublished internal British Library reports into the impact of 'fixing' validation errors.

Aside from their predictive modelling data on bitstream degradation that was generated from running hundreds of thousands of simulations, and indicative format obsolescence data from Jackson generated through analysis of the UK Web Archive (2012), there are few other evidentiary sources from within the community that corroborate the best or good practices the community typically advocates. In the absence of hard evidence, expert analysis techniques have been deployed to support quantitative or mixed-method approaches (Ryan, 2014; Barons *et al*., 2021). The underlying issue remains nonetheless that these rely on experts with relatively little data and limited practical 'long term' experience to justify their theoretical understandings of risk, despite great knowledge about their subject. Composition of a group of experts and their biases inevitably affects results (Paté-Cornell, 1996, p. 105) and even the use of a structured elicitation protocol does not always ensure accuracy of outcomes (Hemming *et al*., 2018, p. 172). The 'best practice' mitigation approaches, particularly in organisational assessment methodologies, are often unsupported by evidence (Maemura, Moles and Becker, 2017, p. 1631). In the absence of evidence to justify requirements, theories and concerns about the risk of loss or damage to digital content remain just that unless an organisation has experienced such loss or damage and is willing to share their story. This however rarely happens, and the dominant position is typically risk averse rather than risk measured. The German National Library for example notes that despite their uncertainties about the importance of validation, invalid files remain a 'significant risk factor' to be dealt with by 'suitable corrective measures' (Hein and Schmitt, 2013, p. 317). As a result, the overall picture remains one of uncertainty, precautionary approaches, and best or good practice responses rather than measured and contextualised understandings of risk.

This lack of clarity or agreement on what constitutes a meaningful digital preservation risk is further evident in the terminology employed within and across the literature surveyed. Despite the prevalence of certain topics across different assessment frameworks, risks and threats are often described in generally vague terms that are difficult to measure. Even within a single model, threats or risks are often described in somewhat inconsistent or unclear terms. This can make it

challenging to identify precisely the nature of individual risks or threats, and what precisely should be done about them. This is particularly the case for goal- or capability-oriented frameworks where the focus is often on identifying current practices rather than uncovering specific risks. As a result, the risks themselves may be unclear, making it difficult to determine whether the assessment will actually help to manage risk. This concern was made explicit in interviews conducted by Frank (2022) for their study on risk in trustworthy digital audit and certification, where practitioners were sceptical about whether the documentation required for TRAC was sufficient evidence that risks were being appropriately managed (p. 45). On the other hand, solutions in which all unmet requirements are perceived as risks can result in a risk register so broad as to be almost meaningless, particularly when requirements are largely based on theory rather than evidence. This, combined with the prevalence of generic risk references within the literature, may have the unfortunate side-effect of inferring to the non-specialist that digital preservation is somewhat of a paranoid discipline in which the spectre of risk hides around every corner. This is neither a helpful perspective of digital preservation, nor a practical approach to managing digital preservation risk.

The sheer proliferation of frameworks indicates a degree of consensus within the community about their value and role, whilst the recurrence of core criteria across many of the frameworks indicates further broad consensus on what the community currently considers indicative of risk (regardless of whether that is the case for all, or not).[48] Despite this, and particularly in the case of format assessment frameworks, they often receive a limited amount of re-use outside of their originating organisations. This gives rise to a sense of 're-inventing the wheel' that is inferred from repeated attempts to draft new (yet similar) assessment frameworks. Uncertainty about the rigour of frameworks can, in part, be satisfied by greater transparency about development methodologies (Maemura, Moles and

[48] As the current toolset uses a broad range of different terminology (and associated definitions), specific spaces of overlap for example in relation to certain types of risks are open to debate. This chapter therefore addresses that nuance with a narrative rather than modelled exploration of the overlap. Chapter four uses this analysis to establish characteristics of the problem area and the implications of these characteristics, leading to specification of requirements for a new solution. Evaluation of the solution against each of these requirements can be found in chapter seven.

Becker, 2017, p. 1631), though there are many other lessons to be learned from this analysis that can also inform a better approach to risk assessment going forwards. This is an opportunity to learn from experiences and improve good practice rather than repeat mistakes.

Conclusions

It is clear from both the literature and this author's experience in the community over the past two decades that whilst current frameworks each have value in exploring certain aspects of potential digital preservation risks and risk responses, there is a lack of clarity concerning the risk landscape in its entirety. The frameworks above all imply a risk -oriented approach, but many are focused more on defined aspirational states rather than clear risk specification and mitigation. Others, particularly those oriented around formats, focus on assessing theoretical risk sources, often without demonstrating a clear understanding of the relevance of a timeframe within which an underlying risk may manifest. Moreover, the underlying methodology for many of the frameworks is often unclear. This leads to uncertainty on their use or value, their coverage, and the relevance of their criteria.

The way in which risk is understood and described clearly has significant implications for how it is assessed and managed. However, even the basic concept of risk is inconsistently and weakly defined in many of these frameworks, alongside other concepts such as threat, vulnerability, risk factor, weakness, and even obsolescence. These limitations mean that whilst the current knowledge base demonstrates the complexity of digital preservation risk, it does not demonstrate a coherent or nuanced understanding of it. This knowledge is vital for establishing and justifying a solid foundation from which to construct a consistent and comprehensive picture of risk. Whilst a small number of approaches utilise particular risk analysis techniques or concepts (Frank, 2018; Barons *et al.*, 2021), there is little to indicate knowledge of the wider field of risk science and limited evidence of a wider understanding or application of risk analysis concepts and techniques beyond the standard ISO 31000 approach. There is a clear gap in the knowledge base when it comes to the conceptual nature of digital preservation risk,

which needs to be addressed in order to establish a more thorough and consistent representation of its complexity.

Experiences with digital preservation risks at the British Library have led to a deep understanding of the range of risks that can affect the longevity of content. Our experiences of managing digital preservation risk at a corporate level go beyond format obsolescence and storage media to also consider those risks that manifest on a more practical level, including staffing reductions, loss of key skills, governance issues, and policy implementation. Some may manifest in the short term, others over the medium to longer term. A flexible approach is needed that can represent this nuance and be implemented in a targeted and re-usable way, for a range of both object-focused and system-focused assessment scenarios, to support decision making without predicting or prescribing solutions up front. This forms the basis of the core requirements for a new solution for the problem space going forwards.

Chapter Four: Requirements for a Solution

Introduction

Analysis of the frameworks in the previous chapter reveals many similarities in the criteria used by each to explore digital preservation risk, particularly between similarly-focused assessment frameworks. There are nonetheless inconsistencies and uncertainties about what truly constitutes a risk, with the concept itself often poorly defined. This impacts on how it is understood and evaluated. A significant number of frameworks focus on mitigations or requirements rather than risks themselves, whilst those frameworks with a clearer risk-focus tend to take a precautionary rather than informed stance, with a heavy reliance on theory rather than evidence. There is widespread consensus that digital preservation is a 'risky business', yet the ongoing proliferation of new frameworks, tools and techniques indicates that standardised or integrated digital preservation risk management practices remain somewhat elusive. This section summarises the lessons learned from the review in the previous chapter in order to establish requirements for an improved solution going forwards that better represents the overall nature and complexity of digital preservation risk.

Lessons learned

The sheer proliferation of different frameworks developed within the community indicates a degree of consensus on their value for identifying, managing, and mitigating digital preservation risk. The review identifies four different types of assessment framework: format focused, object focused, system focused, and organisation focused. This alone reveals a desire for risk solutions that operate at different levels. Some organisations wish to focus efforts on risk management at a format level, whilst others prefer to integrate management of format-related risks into a wider, object- or collection-focused approach. Other types of frameworks indicate a degree of ambition to address risks or threats from a more holistic perspective, focusing either on the technical system used to manage collections, or the whole organisation. There is clearly room for different types of approach, though organisational frameworks appear to receive greater levels of re-use than

format-focused ones, where there is a clear preference for re-invention over re-use.[49] More significantly for the purposes of this thesis though, analysis of the works reveals more fundamental problems.

Of the assessment types explored in the review, the first - format assessments - remains a controversial topic. Contention stems mainly from uncertainty about the criteria used and the suitability of scoring, especially for comparative purposes. In terms of criteria, there appears to be disagreement over which are appropriate metrics against which to understand and measure format-related risks. Ryan for example (2014), argues for a 'top three' of format obsolescence indicators, whilst the Danish Royal Library considers twenty (Skødt, 2022) and US National Archives and Records Administration considers almost forty (Johnston, 2018). This indicates that institutions require flexibility in the criteria they choose to incorporate into an approach. Regardless of how many metrics are included, both the British Library (Pennock, Wheatley and May, 2014) and the Dutch Royal Library (van der Knijft, 2013) observe that many of the criteria are theoretical and advocate caution in the use of theoretical criteria for decision-making. This suggests that a flexible method for implementation would be appropriate for any new tool, allowing institutions to choose those factors they deem most relevant given their institutional context.

Whilst a small number of organisational frameworks have seen reasonable levels of use – particularly CTS and RAM - the lack of evidence about wider community usage of most frameworks indicates a level of uncertainty about either the methodologies used to develop them, their (re-)usability, or their value.[50] A clear design methodology is needed to demonstrate the rigour that underpins any such models and the methods by which they can be used (Maemura, Moles and Becker, 2017,

[49] Informal conversions at the International Federation of Library Association (IFLA) World Library and Information Congress (WLIC) in August 2023 indicate that this continues to be the case, with suggestions of another risk assessment framework recently devised for local use at the University of Nevada. Details of this have not yet been published though information subsequently shared over email indicates that it is not truly a risk assessment framework but a ranked list of collection types considered to be most at risk.

[50] The CTS website identifies over eighty organisations with certified CTS data repositories (August 2023), whilst the DPC encourages all members to undertake RAM assessments to identify areas which require their support. The PTAB website identifies only one organisation that has received their ISO 16363 certification. The number of business users for DiAGRAM and DRAMBORA is unknown.

pp. 1630 - 1631). Organisational assessments requiring documentation have significant overheads in terms of time and resource, as their scope is simply so vast (Rosenthal, 2014). The DPC Rapid Assessment Model offers an antidote to this, though the exceptionally lightweight approach provides significantly less insight and detail about the specifics of organisational capabilities for each criterion. If specificity and contextualisation are the order of the day, this arguably falls somewhat short.

Analysis also revealed a degree of conceptual and semantic inconsistency within and across these frameworks. Across all four categories, frameworks often fail to include a comprehensive glossary. The terms 'risk', 'risk source', 'threat' and 'risk factor' are used in various different ways even within single documents. The concept of risk is often presented in vague and generic terms without clear exploration of what this means for digital preservation, and contributory concepts such as risk sources and risk factors are poorly or inconsistently defined. This is evident even in the more structured assessment frameworks such as DiAGRAM and DRAMBORA. Yet if a risk is not clearly and accurately defined then it can neither be meaningfully managed, convincingly communicated, nor demonstrably mitigated.

Moreover, whilst all of the approaches reviewed in the preceding chapter are of value in their own right, they do not enable a nuanced understanding of the overall landscape of digital preservation risk. This is particularly apparent in organisational frameworks that focus not on risk but on ascertaining compliance with expected standards. The question of 'what is a digital preservation risk' remains under debate. This is arguably acceptable if one considers that a risk for one organisation is not necessarily a risk to another with access to a different set of institutional resources. That justification however is not typically made clear in the literature. Frameworks that require a certain best practice implementation may not be flexible enough to acknowledge this perspective, failing to allow for contextual relevance and presupposing a certain practice is best where there is insufficient evidence to justify it. Flexibility of response is particularly important when a single mitigation action can affect multiple different risks. Ultimately, mitigations are an institutional decision that should be determined independently rather than pre-defined.

This exploration of the domain and analysis of the literature suggests a number of potential reasons as to why many of the digital preservation risk management frameworks developed to date have not seen widespread take up. These are summarised in the table below alongside their subsequent implications for a new solution. Given the broad range of lessons learned, the term 'solution' is here used to refer not just to a risk assessment tool or framework, but also the answer to the research question of how the nature and complexity of digital preservation risk can be more thoroughly and consistently represented so as to form the foundations for a more flexible yet comprehensive preservation planning approach. It prescribes neither form nor format, but reflects instead the knowledge and the capability to meaningfully address and satisfy the problem space.

#	Characteristic of Problem area	Implication
1	Uncertainty about what even constitutes a risk	Solution should *clearly define what is meant by a 'digital preservation risk'*, *as well as related terms* used in the framework (e.g. threat, vulnerability)
2	Poorly defined risks that are difficult to meaningfully measure	Solution should support *clear definition and elucidation of specific risks in a measurable and manageable manner*
3	Risk is everywhere, making it overwhelming	Solution should *be flexible and allow for focus on specific scenarios* rather than trying to do everything all at once
4	Specific risks are often unclear in assessment frameworks that focus on mitigation or solutions	Solution should *avoid confusing or blending risks with mitigations*
5	Uncertainty about methodologies used, their usability or value	Solution should be *usable, configurable, transparent, and clearly able to demonstrate its value/benefits*
6	Significant overheads for some types of approach in terms of time and resource	Solution should *scale* so that it can be used on discrete scenarios with the same valuable output but less resource requirements

7	Uncertainty about whether format obsolescence is such a big issue after all	Solution should support *identification and assessment of risks other than format obsolescence* alongside those that are format related
8	Uncertainty about weighted scoring	Solution should *not require scoring*; if one is suggested then it must be *transparent and configurable without predetermining or precluding weighting*
9	Unclear compatibility with enterprise risk assessment and management frameworks	Solution should *easily align with other organisational risk assessment or management methodologies*

Table 3: Implications for the Solution

Similar themes are evident in several of these implications, particularly in relation to clarity, transparency, scope, and definitions. These are condensed into a series of more precisely defined requirements for a solution, representing the second stage of the design science research methodology.

This author's learned experience from over two decades of working in digital preservation research and development initiatives is that over-stipulation of requirements frequently leads to an unnecessarily complex endeavour with increased chances of failure, not just in terms of function but also form, which affects usability. This is particularly the case when resource availability is constrained. Requirements are therefore carefully phrased but purposefully minimal, in line with the drive for utility over truth. Relationships between requirements and the implications defined above are indicated in numbered brackets at the end of each statement.

Requirements for the Solution

1. The Solution must provide clarity into the relationship between the concept of risk and the practical manifestation of risk, so that the difference between the two is understood in intellectual terms (#1)
2. The Solution must define all key terms used, so that it is a comprehensive representation of the requisite vocabulary (#2)

3. The Solution must use precise language in its definitions of terms, so that the potential for ambiguity and misinterpretation is minimised (#1, #2)
4. The Solution must clearly identify the main elements of a digital preservation risk ecosystem, so that it can be used to explore risks associated with different organisational and technological aspects of digital preservation (#3, #7)
5. The Solution must not conflate or relate risks with mitigations (#4)
6. The Solution must be demonstrated by one or more methods but remain flexible and not predetermine a particular approach (#3) (#5) (#9)
7. The Solution must not assign importance or severity levels to objects, as this is contextual and down to individual users to assign (#6, #8)

Overall, this review and analysis suggests that the research question can be answered with a holistic model of digital preservation risk that can subsequently be contextualised and used by an institution in a manner most appropriate to their needs. It should be independent of mitigation measures, fully documented, and with suggested guidance on implementation rather than prescriptive methods. It should clearly establish what it meant by the phrase ‘digital preservation risk’ at both a conceptual and practical level, and be flexible enough to support different types of risk assessments in various different ways. The following chapters present the solution developed for this research that meets these needs.

Chapter Five: Deconstructing Digital Preservation Risk

Introduction

The way in which a risk is described invariably influences how it is assessed, understood, and mitigated (Aven, 2016, p. 4). Whilst the underlying premise of digital preservation risk is generally that of uncertainty and loss, the previous chapters have demonstrated significant inconsistency in how the community describes and understands risk in any detail. Some focus on digital preservation as obsolescence avoidance, leading to obsolescence-dominated risk assessment frameworks despite reservations in the community about obsolescence and the methodologies used. Others take a broader perspective, assessing organisational and technological capabilities against a set of benchmarks but often without explicitly relating these to underlying risks. The underlying risk model in most frameworks is unclear, particularly with regards to conceptual relationships, whilst methods for implementation are either poorly defined or inflexible and can lead to uncertainty about whether they will actually help to manage risk. How then can the nature and complexity of digital preservation risk be more rigorously established and consistently represented, so as to address these limitations and support a more nuanced understanding of both risk and risk assessment solutions going forwards?

This chapter explores the nascent theory that risk science offers a more rigorous way to answer that question and establish a more thorough understanding of digital preservation risk, one that moves beyond the generalist ISO 31000 definition and towards a contextualised definition of risk for a digital preservation setting. It uses an approach from risk science as a base from which to explore the detail of the nature and complexity of digital preservation risk, differentiating between the concept of risk and practical manifestations of risk. It explores different perspectives on digital preservation risk through a series of cascading models that identify and map relationships across and between different types of risk-based entities, both conceptual and logical. This structured and reasoned approach to modelling culminates in a comprehensive digital preservation risk source model that functions as a universally recognisable abstract representation of the digital

preservation risk domain, embodying a newly formed and thorough understanding of the complexity and nature of digital preservation risk.

Conceptual understandings of risk

In general parlance, the term 'risk' typically has a negative connotation. Despite this, the global standard on risk management (ISO, 2009) defines risk simply as the 'effect of uncertainty on objectives' (p. 1), whereby an effect is a deviation from what is otherwise expected, either positive or negative. The ISO definition of risk is sufficiently abstract and vague that it operates as a conceptual reference description of risk, applicable to any scenario. Whilst this ISO definition is satisfactory for the purposes of a generic and conceptual description of risk, Hansson (2012) acknowledges the problems with a vague definition of risk for informed decision-making and argues the value of precise terminology for improving an overall understanding of risk (p.30).[51] Such precision has particular value when attempting to communicate risks to different audiences with varying degrees of knowledge and capabilities about the scenario at hand.

From a risk science perspective, Ylönen and Aven argue for a distinction between conceptual and characterised risk, observing that 'much of the confusion observed in practice concerning risk can be tracked back to the concept of risk being mixed with its measurement or characterisation' (2023, p. 592). The concept of risk represents one or more abstract principles, whilst the characterisation of risk relates to descriptions and measurements against which judgements are made – in other words, in relation to the practical manifestation of risk. Characterisation requires contextualisation, and applying this also to the concept of risk can establish a clear link between statements about risk at both conceptual and practical levels. This can be used to establish conceptual reference points against which subsequent characterisations of risk can be consistently described and

[51] As previously noted, a more precise framing or description of risk can also go some way to help counteract distorted risk perception, whereby risks are interpreted in a way that is inconsistent with reality. This distortion is clearly evident in much of the literature on digital preservation risk, as indicated by the review of literature and solutions in chapter three.

measured. A contextualised definition of the concept of digital preservation risk is thus a helpful and practical place to begin.

Risk science offers a structure for contextualising the concept of risk, specifically in relation to planned objectives or values for specific disciplines and practices. Aven (2016, p. 5) cites the example developed by Heckmann *et al.*, (2015), whereby risk is conceptualised as the potential loss in a given context in terms of its target values, evoked by uncertain developments and triggering events. This approach explicitly associates risk with negative outcomes and identifies three essential components of a risk definition that moves it from representing a wholly abstract concept to a contextualised concept: firstly an undesirable outcome (the potential loss in a given context), secondly the target values against which the outcome is measured (contextual stated target values), and thirdly the potential causes of that outcome (uncertain developments and triggering events). It demonstrates the value of contextualisation as a mechanism to develop a better understanding of risk in different settings, and outlines the wider framing or scope within which the risks associated with a given event or situation can be explored. This structure forms the basis of the approach developed within this thesis to develop an improved understanding of digital preservation risk in terms of both concept and practical manifestations.

Defining Digital Preservation

A contextualised understanding of digital preservation risk requires a clear view on the function and purpose of digital preservation – in other words, a clear definition of digital preservation itself. The term digital preservation means different things to different people, even within the established community of practitioners and academics. The ISO 14721 reference model for the digital preservation community functions as a widely accepted and shared vocabulary within the field, though it does not define the term 'digital preservation', the nearest being instead 'long term preservation' - described as an act of 'maintaining information, independently understandable by a designated community, and with evidence supporting its authenticity, over the long term' (ISO, 2012, pp. 1-13). Several different definitions are explored in chapter one, though no single source has yet produced an

authoritative definition embraced by the field. A number of recurring concepts and themes are nonetheless evident across many of the different definitions used. These include the importance of access (RLG-OCLC, 2002; Brown, 2013; British Library, 2017; Owens, 2018) - preservation is a means to this end, not an end in and of itself. Several definitions also refer to authenticity (American Library Association, 2008; Brown, 2013; British Library, 2017), acknowledging that items must be that which they purport to be and unchanged in any meaningful sense so that they can be confidently and reliably re-used.

A further clear theme is that preservation is not a solitary act or action but incorporates multiple different types of activities and processes (RLG-OCLC, 2002; Digital Preservation Coalition, 2015a; UNESCO, 2023). Related is the implication that preservation is an ongoing activity (Lavoie and Dempsey, 2004; University of Manchester Library, 2020; State Library of New South Wales, 2022): this reflects the so-called 'moving target' of digital preservation (Hofman, 1999) and the unavoidable fact that the environments in which most digital resources are managed and preserved is in a near constant state of change. This is driven by ongoing external technological advancements as well as ongoing organisational and socio-economic pressures.

Digital preservation approaches must address and support these themes or concepts whilst remaining responsive to changes in their wider environment, both technological and organisational.[52] Considering these themes and concepts, and in particular the previous definition used at the British Library where this work is undertaken, this thesis therefore defines digital preservation thusly:

> ***Digital Preservation*** *is the series of coordinated organisational and technological activities undertaken in an organisation throughout the lifecycle to ensure its digital content is retrievable, authentic, has integrity, and is accessible over time for current and future users.*

[52] These themes and concepts are explored extensively in numerous other frameworks and models (Task Force on Digital Repository Certification, 2007; McGovern, 2007; ISO, 2012a; nestor, 2013; CoreTrustSeal, 2020), though to different degrees of granularity and from different perspectives.

This definition incorporates the main conceptual and contextual themes identified previously, including access, authenticity, the ongoing nature of digital preservation, and the importance of different operational and technological activities. It represents the multi-faceted nature of digital preservation and, as such, can be interpreted from different perspectives. It represents digital preservation not only as a series of processes, but also as a content-focused endeavour, and even as an outcome. From a process perspective it makes clear that digital preservation is an ongoing and coordinated series of activities throughout the lifecycle and over time, both technological and organisational. From a focus perspective it makes clear that digital content is at the heart of the endeavour – it is the core around which the framework of activities is built. Finally, from an outcome perspective, it makes clear the expectations, purpose, and goals of the endeavour through stipulation of high-level capabilities and properties for both content and its environment: retrievability, authenticity, accessibility and integrity, and an expectation that these are maintained over time.

The definition specifies the scope and parameters within which digital preservation occurs. As a definition, it is institutionally agnostic and applicable to any setting that aims to preserve digital content. It does not assume any particular implementation, organisational structure, or content type. It clearly identifies the goals and target values of a digital preservation endeavour as well as the managed environment in which it takes place.[53] Conceptually, it represents three main elements: a managed operational environment representing organisational and technological activities, the digital content itself, and the target values associated with the content. The relationships between each of these concepts can be illustrated in a simple model (figure 4).

[53] Some preservation programmes may have other targets outside of those explicitly relating to preservation of content, for example in terms of affordability, or organisational acceptance of change. These are specific to an institution and are therefore not included in this definition, though adaptation is possible to support these additional target values if necessary.

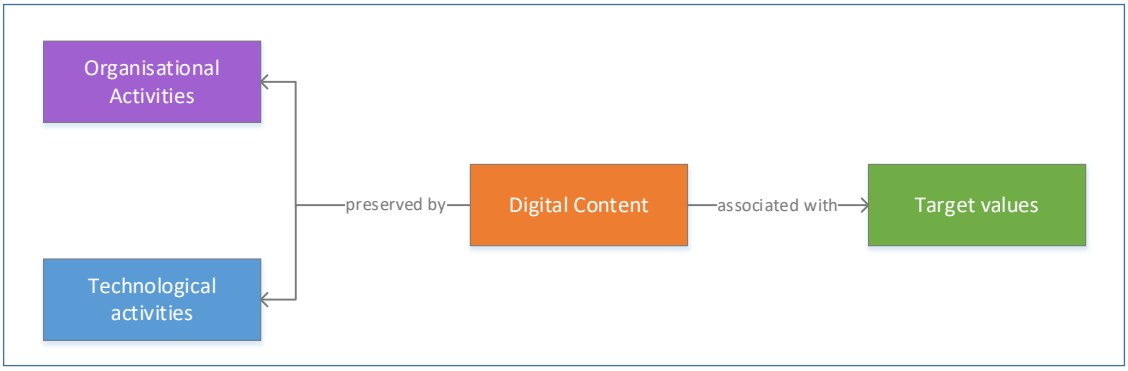

Figure 4: Conceptual relationships in a definition of digital preservation

Stipulation of target values in the definition of digital preservation provides greater clarity on the preservation objective, including the important properties or capabilities that must be maintained over time and against which outcomes are assessed. This subsequently provides our initial frame of reference from which to develop a contextualised understanding of the concept of digital preservation risk.

Defining Digital Preservation Risk

A contextualised definition of digital preservation risk moves beyond a representation of the simple abstract concept of risk in ISO 31000 to a more useful and principled concept of digital preservation risk. The structural framework for a definition of risk designed by Heckmann *et al.*, (2015), is applied to achieve this, in which risk is defined as the potential loss in a given context in terms of the contextual stated target values, evoked by uncertain developments and triggering events. The terms and concepts established in the definition of digital preservation above form the basis of this approach, so that the concept of digital preservation risk can be understood in relation to the purpose and function of digital preservation. This leads to the following:

> **Digital Preservation Risk** is *the potential for complete or partial loss of digital collection content in terms of its target values of retrievability, authenticity, integrity, accessibility, and longevity, arising from sub-optimised risk sources within the managed organisational and technological environment in which the content should otherwise be preserved.*

This is a contextualised definition of risk specifically for a digital preservation setting. It provides sufficient detail to move the concept of risk from a wholly

abstract construct to a disciplinarily contextualised construct that precisely defines the overall problem area. It maps to all three components of the risk structure suggested by Heckmann *et al*.: a) the undesirable outcome <*complete or partial loss of digital collection content*>, b) the target values against which that is assessed <*retrievability, authenticity, integrity, accessibility and longevity*>, and c) the potential causes of a negative outcome <*sub-optimised risk sources in the managed organisational and technological environment, within which the content should otherwise be preserved*>.[54] This level of description illustrates well Hansson's observation that precise terminology can improve the overall understanding of risk, effectively providing greater clarity and direction to experts and non-experts alike. The relationships between each of these concepts is mapped in figure 5, below.

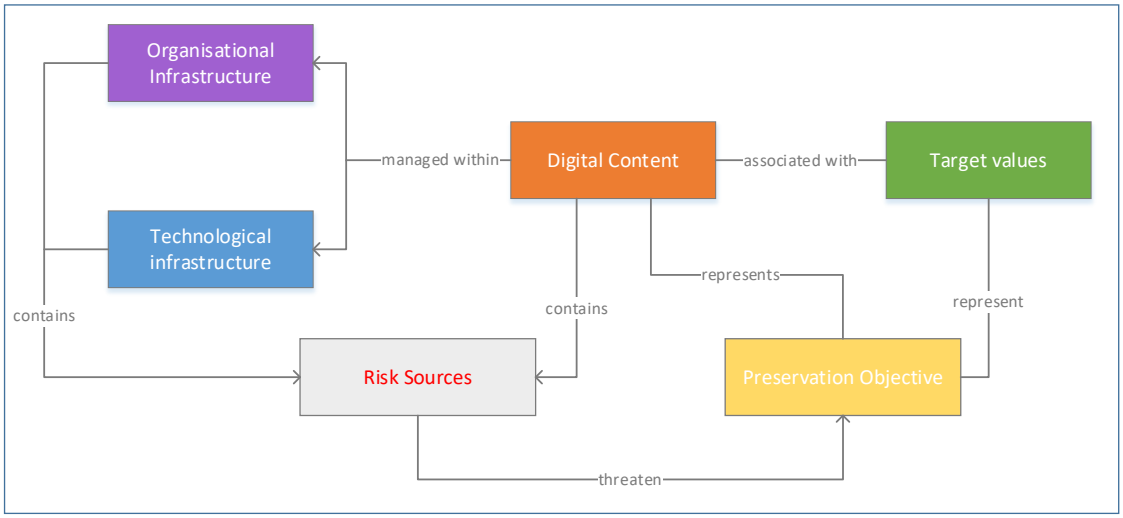

Figure 5: Conceptual relationships in a definition of digital preservation risk

This model extends the initial digital preservation concept model by expanding on and adding additional conceptual entities. It introduces the concept of a risk source, which threatens organisational ability to satisfy preservation objectives and target values. Most importantly, it models the relationship between risk sources and their origins, namely the technological and organisational infrastructure, and the content itself that should be otherwise managed and preserved within that infrastructure. The model thus represents a high level, conceptual understanding and representation of the nature of digital preservation risk.

[54] For a distinction between the terms 'risk' and 'risk source', see the Glossary (Appendix A). The risk source concept is also extensively explored later in this chapter.

The relationship between different target values and different aspects of the operational context can be further illustrated through expansion of this model, as below. This explores in particular the relationship between different target values and the risk originating entities of digital content, technological infrastructure, and organisational infrastructure.

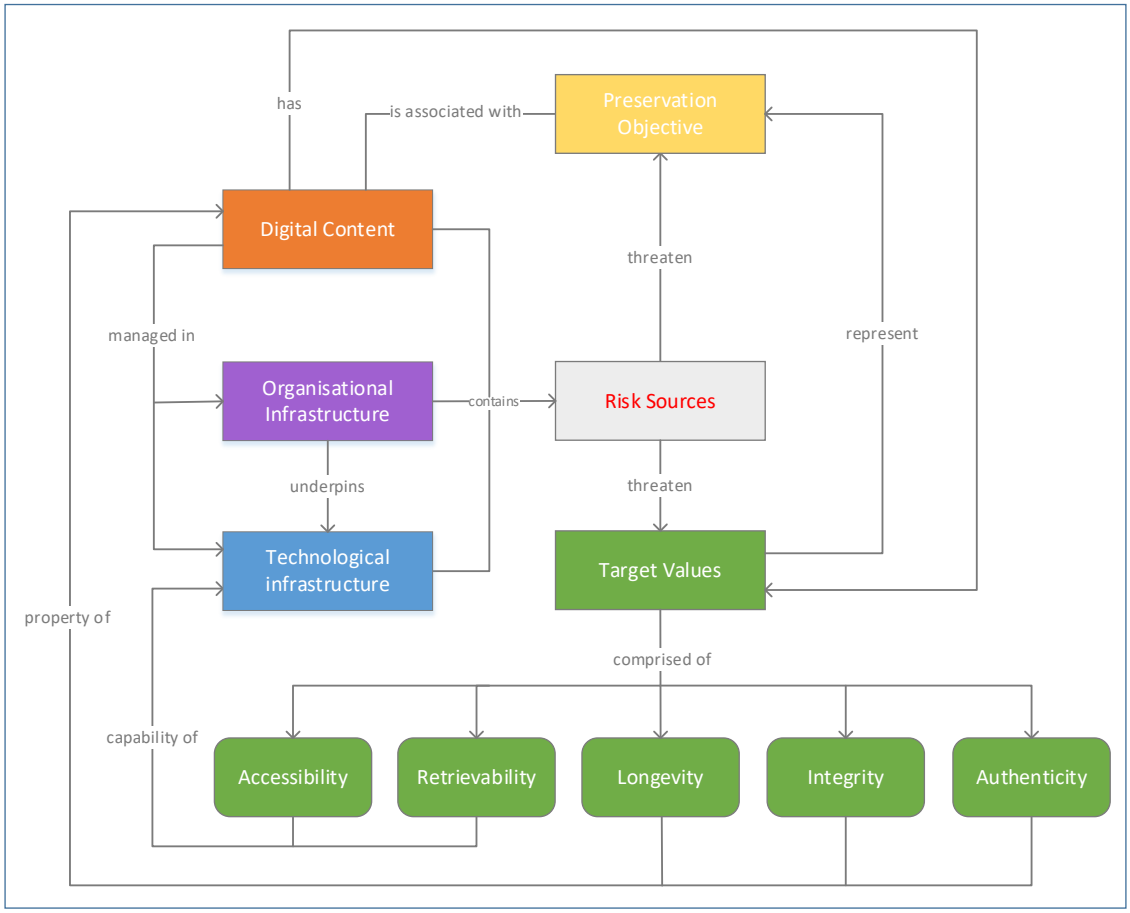

Figure 6: Digital Preservation Risk Context Model

This fuller version of the context model represents the complex relationships that exist between different conceptual entities in the definition. Agnostic of any particular institutional setting and thus broadly relevant across the domain, it establishes a firm foundation from which to explore and develop a deeper disciplinary understanding of the nature and complexity of digital preservation risk.

Many of the entity terms used in the model are defined elsewhere within various community glossaries. However, as with the definition of digital preservation, it is often the case that a single, authoritative definition is not widely agreed. As an established practitioner, the author is aware of these and the concepts they aim to

represent. As comparative analysis of terms is not the purpose of this research, this thesis defines each term afresh as relevant to the context of the research question.

Each of the key entities in the model is defined in more detail below.

Digital Content is at the heart of the preservation endeavour for most memory institutions. It is intangible: as such, it does not exist until meaningfully accessed by an end user. Content can be represented by a single collection object, a group of collection objects, or a whole collection.[55] A series of underlying conceptual entities illustrate this process: *digital file(s), digital object(s),* and *intellectual content*.[56]

Figure 7: Digital Content concepts

The term **Digital Object** refers to *the artefact that is the focus of the preservation effort,* such as the eBook, the album, the poetry anthology, or the mobile app. In libraries, this is often considered as the 'published' output, though not all library artefacts originate from a publishing house - electoral register datasets, archived websites and personal digital archives are all such examples. In non-library organisations, objects might be archival papers, reports, digitised representations of 3D artefacts, or any other form of holdings. All can represent either born digital, or digitised objects. Digitised objects are generated from a physical object, which may or may not also be held by the institution concerned, whilst born digital objects are those which are generated and acquired in digital form. Digital objects contain **Intellectual Content,** which is *the meaningful knowledge or information consumed by humans or machines*. The term **Digital File(s)** refers to *the bitstream(s) that contain all of the encoded information for generating (and optionally also describing) a digital object.* This includes the intellectual content and instructions

[55] The terms 'item' and 'object' are often used interchangeably within the community and even specific institutions. This thesis employs the term 'object' over 'item', as the term 'item' is often suggestive of a single, boundaried intellectual artefact whilst the term 'object' may have fewer mental associations in terms of shape or form. It is therefore potentially less contentious when seeking to apply the terminology across different types of content and different types of heritage collections.

[56] These are individuated in figure 7 rather than figure 6, for reasons of visual clarity.

for rendering it, represented within one or more files. It may also include metadata about the object generated either directly by the software used to create the files, the publisher, or the preservation organisation.[57] Digital Content is the focus of the *preservation objective*, though it contains *risk sources* that unless properly managed, may threaten an organisation's ability to achieve that objective.

Digital Content exists within the operational environment of the holding institution. The operational environment has two aspects in our definition of digital preservation risk: *organisational infrastructure*, and *technological infrastructure*.

The term '**Organisational Infrastructure'** is widely used in business to define the operational framework of an organisation. It is a frequently used term in digital preservation assessment frameworks, first in the Trustworthy Repository Audit and Certification (TRAC) metrics (Task Force on Digital Repository Certification, 2007) and subsequently also in both ISO 16363 (ISO, 2012b) and the Core Trust Seal (CoreTrustSeal, 2020) standards. In a digital preservation context, the term 'organisational infrastructure' is used to represent *the organisational environment in which a managed digital preservation service occurs*. This includes the '3 P's' of People, Policies and Processes as well as areas such as governance, finance, strategic planning, mandates, and legal affairs. Digital content has value within an organisational infrastructure, which also provides the mandate and justification for the continuance of preservation activities.

The term **'Technological Infrastructure'** is also widely used within business. Variations of the term appear in different organisational assessment frameworks, though they often represent the same or similar types of activities and functions. In a digital preservation context, the term 'technological infrastructure' refers to *the technological environment in which digital content is acquired, managed, preserved, and made accessible from*. It refers to an organisation's technology estate, including the hardware, software, systems, servers, facilities and processes that provide the

[57] Files may be represented in an OAIS-type conceptual Information Package, as described in ISO 14721. Exploration of the relationship between the concept of digital content and an information package is however out of scope for this thesis as it does not advance a practical understanding of digital preservation risk.

backbone of its information technology capability. This includes areas such as hardware and software environments used by business functions (both in-house and outsourced), cyber security, systems support and admin, networks, and managed services. Digital files and digital objects are managed within the technological infrastructure, providing and maintaining the systems, networks and processes required to support their ongoing preservation and enable access to their intellectual content.

Whilst the model represents Technological and Organisational Infrastructures as separate entities, in practice they are co-dependent. The people element of an organisational infrastructure is, for example, essential to the successful running of a technological infrastructure. Similarly, the finance element of an organisational infrastructure pays for the technical systems managed within the technological infrastructure. They are nonetheless depicted as separate entities in the model as they are associated with different types of risk source. This separation later enables a more precise exploration into the nature of the different types of risk sources as well as their associated *risk factors*.

Digital Content is associated with a number of **Target Values**. These *target values represent the goals of a preservation endeavour*, defined in our definition of digital preservation as content that has *integrity, longevity,* is *authentic, accessible,* and *retrievable*. Target values are either properties of digital content, or capabilities of the infrastructure. These are illustrated in figure 8, below.[58]

[58] Solid lines in figure 8 depict direct relationships between concepts, whilst dotted lies depict associations between different target values.

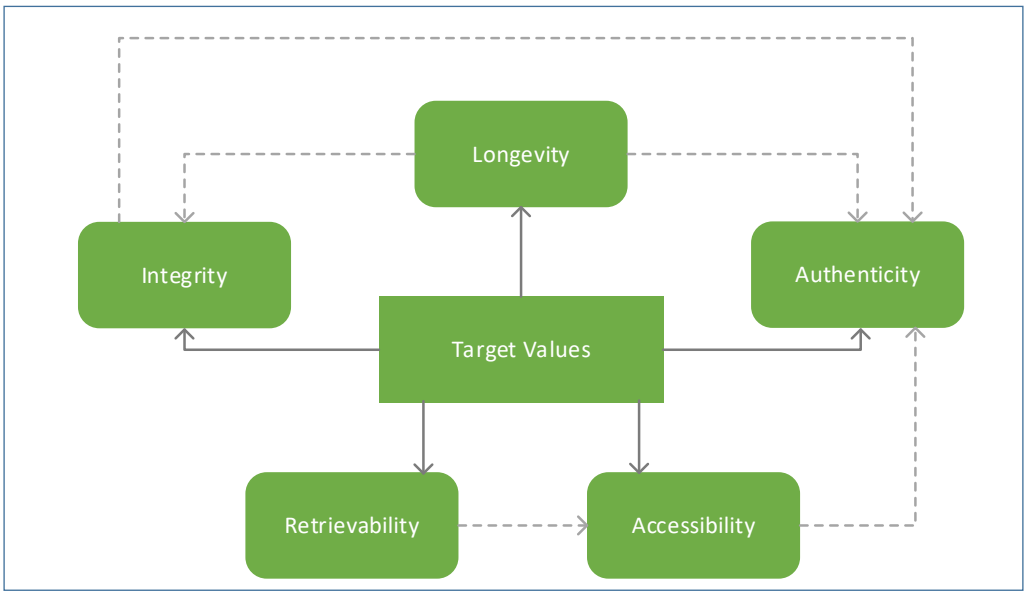

Figure 8: Target value concepts and relationships

Retrievability enables something to be recovered from somewhere. In computing scenarios, retrievability is often associated with ease, i.e. the straightforward manner with which something is recovered or retrieved from somewhere. Speed of retrievability matters to end users, though in preservation terms this value is also an enabler for accessibility – if content cannot be retrieved, then it cannot be accessed. In digital preservation, retrievability is therefore defined as *being able to recover specific digital content files from their storage location with relative ease*.[59] Retrievability is a capability provided by the technological infrastructure.

Accessibility has an inherent dependency on retrievability – if content cannot be retrieved, it cannot be accessed. Contemporary definitions of accessibility often pertain specifically to equitable access for alternatively abled communities and/or people. This is relevant for memory institutions like any others, though our definition is concerned primarily with access challenges caused by the inherently intangible nature of digital content. In the context of digital preservation therefore, the concept of accessibility simply relates to *being able to access the digital content*

[59] The 'Findable' property referenced in the 'FAIR' acronym (Findable, Accessible, Interoperable, and Usable) is similar though specifically requires the use of unique and persistent identifiers, rich metadata, and indexing in a searchable resource (Wilkinson *et al*., 2016). FAIR was designed around scientific data management and stewardship. It does not explicitly support preservation (Sierman, 2019) and thus is not explored further in this thesis.

that is the meaningful focus of the preservation effort. Accessibility is also a capability provided by the technological infrastructure.

Archival definitions of **Authenticity** are closely linked to the concept of provenance, providing assurance that an object is that which it purports to be - namely a 'true' artefact produced from a specific source and/or on a specific date or time, unchanged in any significant manner that may impact on its intellectual meaning and significance. In a digital environment, objects may look or behave differently from device to device depending on the access environment and any configurable settings. These technological dependencies bring an additional dimension to the concept of authenticity, as they can make it a challenge to establish the 'authentic form' of an object as first published or used. Uncertainties over the rendered form of an object can inhibit reliable re-use and the value perception. Authenticity is therefore defined here primarily in relation to the manner in which an object is rendered, represented by *the reliable and accurate rendering of a digital object that maintains its value to the holding institution*.[60] Authenticity is thus considered a property of the digital object.

The property of authenticity is closely bound to that of integrity**. Integrity** is a property of the underlying digital file(s), i.e. the bitstream(s) in which the intellectual content is encoded, in the structure of a given file format. Bitstreams are represented through a binary sequence of zeros and ones (or 'bits'). A file has demonstrable integrity when the binary sequence of zeros and ones remains unchanged from when last checked.[61] Integrity is therefore defined here as *the binary sequence of a digital file remaining whole and unchanged since it was last confirmed*. It is an enabling property for authenticity, as changes to the file integrity

[60] For an early but nonetheless relevant and in-depth discussion of digital authenticity, see 'Authenticity in a Digital Environment' (CLIR, 2000). The concept of significant properties is considered by some as a mechanism for exploring and validating authenticity, though the application of this approach has its limitations – see for example 'Metaphors We Work By: Reframing Digital Objects, Significant Properties, and the Design of Digital Preservation Systems' (Becker, 2018), 'On the significant properties of spreadsheets' (van der Knijft, 2021), and the comments from this author in chapter 2 regarding their experience with significant properties.

[61] Checksum software to capture and demonstrate file integrity is widely used in digital preservation, though modern storage solutions often provide integrated support for file integrity management. This is typically accompanied by automated recovery from replicates but occurs at a storage block level rather than an individual file level.

can affect the way in which content is rendered even when its rendering dependencies are otherwise supported.[62]

Longevity is a property of the digital object that is dependent upon the capabilities of the operational environment and enabled by aspects of both organisational and technological infrastructures. The longevity of an object represents its *endurance over time and throughout the lifecycle so it remains available for current and future users*. The property of longevity is independent of authenticity and integrity but closely linked, as the full value of an object is only truly realised when its authenticity and integrity are also maintained. If the authenticity and integrity of objects and files is uncertain or lost, then the reusability and accuracy of the rendered object is questionable. As a result of this damage or uncertainty, the value of continuing to maintain that object over time – i.e. ensuring its longevity – is reduced. Longevity might thus be considered a companion property to authenticity and integrity in order to ensure that it remains valuable as a target property in its own right.

All of these properties and capabilities are target values for digital preservation. An organisation's **Preservation Objective** is satisfied when these target values are demonstrable with the organisation's digital content. The preservation objective is *the goal of a digital preservation endeavour*. This is demonstrated through the values specified in our definition of digital preservation – namely ensuring that acquired digital content remains retrievable, authentic, has integrity and longevity, and is accessible for current and future users.

Risk Sources threaten an organisation's ability to maintain the target values associated with its digital content and thus achieve its preservation objective.[63] A risk source is a potential cause of a negative outcome. It represents uncertainty, though also opportunity to avoid a negative outcome through optimisation. It is thus defined as a *changeable element in the digital preservation environment that*

[62] For an excellent visual representation of the results of 'bit flip', see Cochrane (2012).
[63] The term 'risk source' is often used interchangeably with 'root cause'. Aven and Thekdi observe that a 'root cause' may require interaction from several other variables before the negative outcome ensues and therefore caution against use of the term, preferring 'risk source' (2022, p. 342).

alone or in combination with others has the intrinsic potential to give rise to a negative outcome. As change over time is inevitable, most elements of the environment or content can be considered to be risk sources - this is perhaps the underlying reason for the perception that in digital preservation, everything is a risk.

Risk Source is a foundation yet complex concept in digital preservation risk. A risk source alone does not necessarily result in a risk - it must be affected in some way so as to produce a negative outcome. Explication of the risk source concept reveals a structure around which to centre an exploration of characteristics that can affect risk sources and result in negative outcomes. Application of this structure across the digital preservation risk domain generates a full overview of digital preservation risk sources and related components, establishing a clear, thorough, and consistently structured representation of the complex landscape of digital preservation risk. This transcends our understanding of digital preservation risk from a contextualised concept into a series of consistently presented, practical opportunities for risk management.

The next sections present this overview in the form of a Digital Preservation Risk Source Reference Model. The model is agnostic of any particular setting, describing sources of risk in sufficiently precise detail to be usable and recognisable whilst generic enough to be relevant across different types of organisations regardless of differences in their nomenclatures, collections, or infrastructures. As such, it represents a significant step forwards in our disciplinary understanding and knowledge of the nature and complexity of digital preservation risk.

Risk Source Model

The Digital Preservation Risk Source Model is structured around a series of conceptual entities derived from a deconstruction of the risk source concept. These entities are *Risk Originating Entities*, *Risk Source Classes*, *Risk Source Instances* and *Instance Types*, and *Risk Factors*. Collectively, these conceptual entities represent a re-usable structure through which to consistently and comprehensively express

different aspects of the risk source concept. The relationship between each of these is illustrated in figure 9, below.

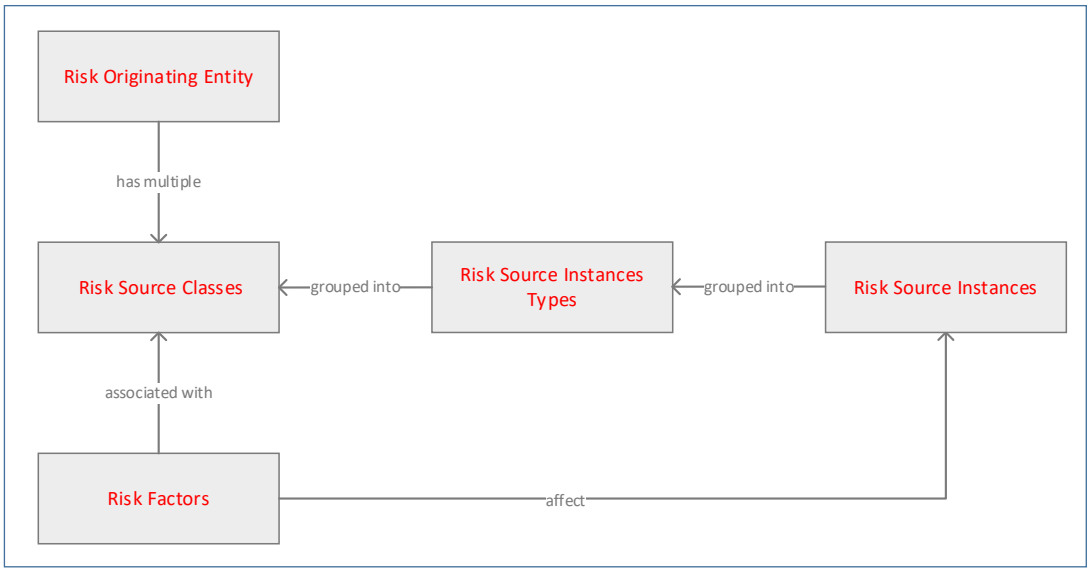

Figure 9: Risk Source Concept Model

A **Risk Originating Entity** is *an aspect of the operational digital preservation environment that contains risk sources*. The Risk Source Model identifies three risk originating entities in the digital preservation domain, in line with the definition of digital preservation risk above - Technological infrastructure, Organisational Infrastructure, and Digital Content.

Each Risk Originating Entity contains several **Risk Source Classes.** A Risk Source Class is *a conceptual grouping of similar risk sources instances*, whilst an **Instance** is a *specific risk source*. The model does not represent individual risk source instances as they can manifest in many different ways, determined to a large degree by operational context. They are represented instead as abstract **Risk Source Instance Types**. Each Instance Type represents a *different type of an instance of a risk source within a given class,* at a lower degree of abstraction than the risk class entity. Instance types in the model are indicative rather than comprehensively definitive, representing a sufficiently recognisable overview of the likely types of instances of risk sources in a given class to guide exploration by a risk assessor in situ.

All instance types and instances in a given class share the same set of **Risk Factors.**[64] A Risk Factor is a more precise indicator of the area of uncertainty associated with an instance of a risk source than the instance alone. A Risk Factor is thus understood as *a variable property of a risk source that can be optimised to reduce uncertainty and the likelihood or impact of a negative outcome*. Factors are neutral representations of a changeable aspect – as such, they may be either formative (causal factors) or reflective (consequential factors). This neutrality is purposeful and supports universal application of the model, as a risk factor may be causal in one context yet reflective in another, depending on the associated contextual variables.[65]

Factors are not simply unmet requirements, as might be perceived from some other frameworks and models (e.g. Dappert, 2013; McHugh, 2016). Requirements may represent risk mitigations, but risk assessments should focus on reviewing potential sources of risk to inform potential mitigations rather than determine whether specific requirements are met.[66] This separation of risk source and mitigation is necessary to retain the universal relevance of the model, as mitigations and interpretations of 'good' or 'best' practice may vary between different institutions.

The Risk Source Model translates these various conceptual entities into a series of logical entities that depict combinations of data from which a specific, individuated and contextualised risk source can be defined. It represents a universal overview of the digital preservation risk source domain. Conceptual Risk Originating Entities are individuated and associated with a series of logical entities, each comprised of a Risk Source Class that represents a grouping of similar types of risk source, one or

[64] The term 'risk factor' appears in several different digital preservation assessment frameworks (Stanescu, 2004; Ryan, 2014; Johnston, 2018; Becker, Faria and Duretec, 2015; Barons *et al*., 2021) though is used somewhat inconsistently and interchangeably with the terms metrics, criteria, and even risk itself.

[65] Ryan (2014) explores the relationship between formative and reflective indicators in her thesis (pp. 59 – 60) and uses a Delphi panel to help establish the formative indicators of format endangerment.

[66] For example, a common requirement is that multiple copies of content are held in different geographic locations. This can mitigate the risk of loss due to failure of a storage location, caused by external attack or a natural disaster. Describing a failure to store multiple copies of content in different geographical locations as a risk thus blends the risk with risk mitigation, rather than allowing mitigation to be determined independently.

more generic risk source Instance Types, and a set of Risk Factors shared by each risk source instance and instance type associated with the class. This holistic, domain-level overview enables a thorough and comprehensive investigation of risk sources in any given scenario, whilst the structure enables consistency in the way situational risks are subsequently described and characterised. It represents a deeper application of the risk science approach represented in the definition of digital preservation risk, supporting a structured exploration of where and why risks might manifest in a digital preservation environment, so that they can be described and subsequently managed in a more consistent and precise manner. This systematic and holistic approach to identifying risk in a digital preservation context represents a significant new risk science-based contribution to digital preservation knowledge.

For purposes of clarity, mapping is focused primarily on the relationships between risk originating entities and their associated risk source classes, factors, and instance types. A degree of redundancy exists within the model in that it may on occasion be perceived to represent similar risks and consequences across different entities. This is purposeful, for as a neutral and universal reference model it does not assume any one perspective or implementation. Different paths can lead to the same or similar risks, though not all will be relevant across all different scenarios. The model nonetheless has some limitations. For example, it does not attempt to directly map relationships between different classes of risk sources or between risk factors themselves, as these are typically less abstract in nature and more situational. This would again imply a certain implementation, which is inconsistent with the reference model approach. Simplicity and utility are both key. In the same manner, the model does not attempt to map the range of consequences that can be associated with risk entities. These are also situational and variable, depending on the wider context in which a risk manifests and the framing of the factor as either causal or reflective. Consequences are associated instead with a later model for characterising specific risks that is explored further in the methods outlined in chapter six. Textual descriptions of each class provided in this document

nonetheless provide some indication of the different types of consequence that may ensue if risk sources are not optimised.

The full risk source model is too detailed for a single page so is represented across a series of Level 1 and Level 2 models. The Level 1 model identifies and maps relationships between all risk originating entities, risk source classes, and risk factors, whilst Level 2 models each focus on the risk source classes, factors and instance types associated with a single risk originating entity. Discussion is focused on the Level 2 models as these provide the detail through which the logical entities are fully represented and identify the combinations of data from which a specific, individuated and contextualised source of risk is subsequently defined.

These models are presented in UML, the Unified Modelling Language. UML is a standard form of notation from the field of computer science for visual object modelling. The model uses only simple UML notation, as appropriate to the relationships featured, the focus on utility, and the expected audience.

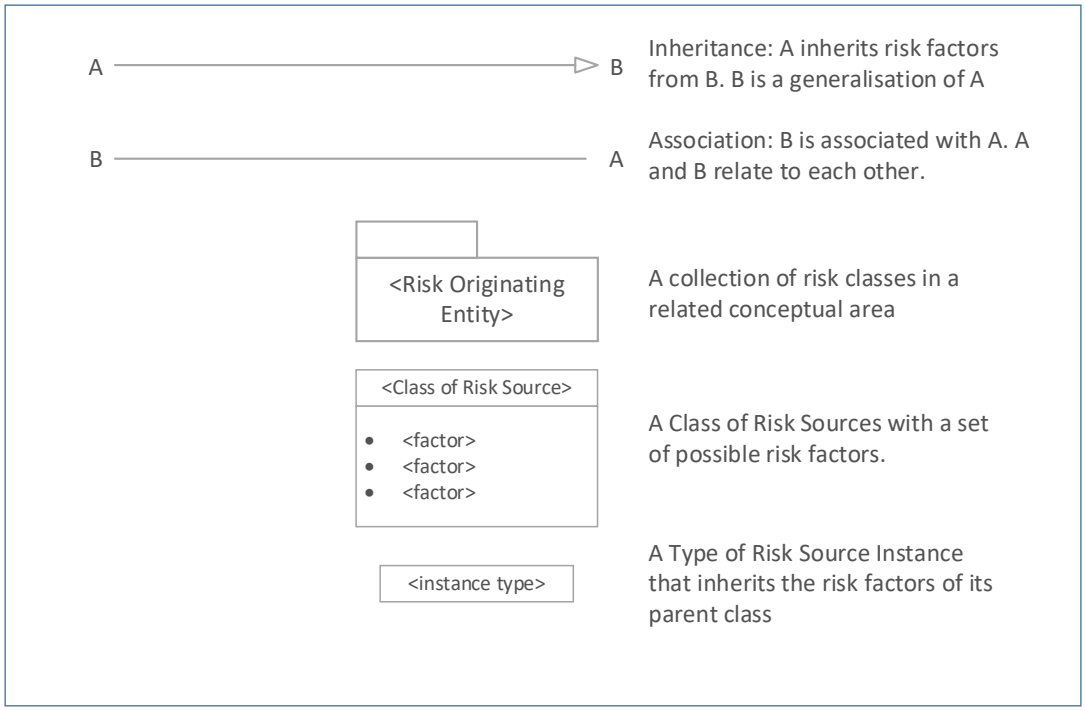

Figure 10: Notation key for the Digital Preservation Risk Source Model

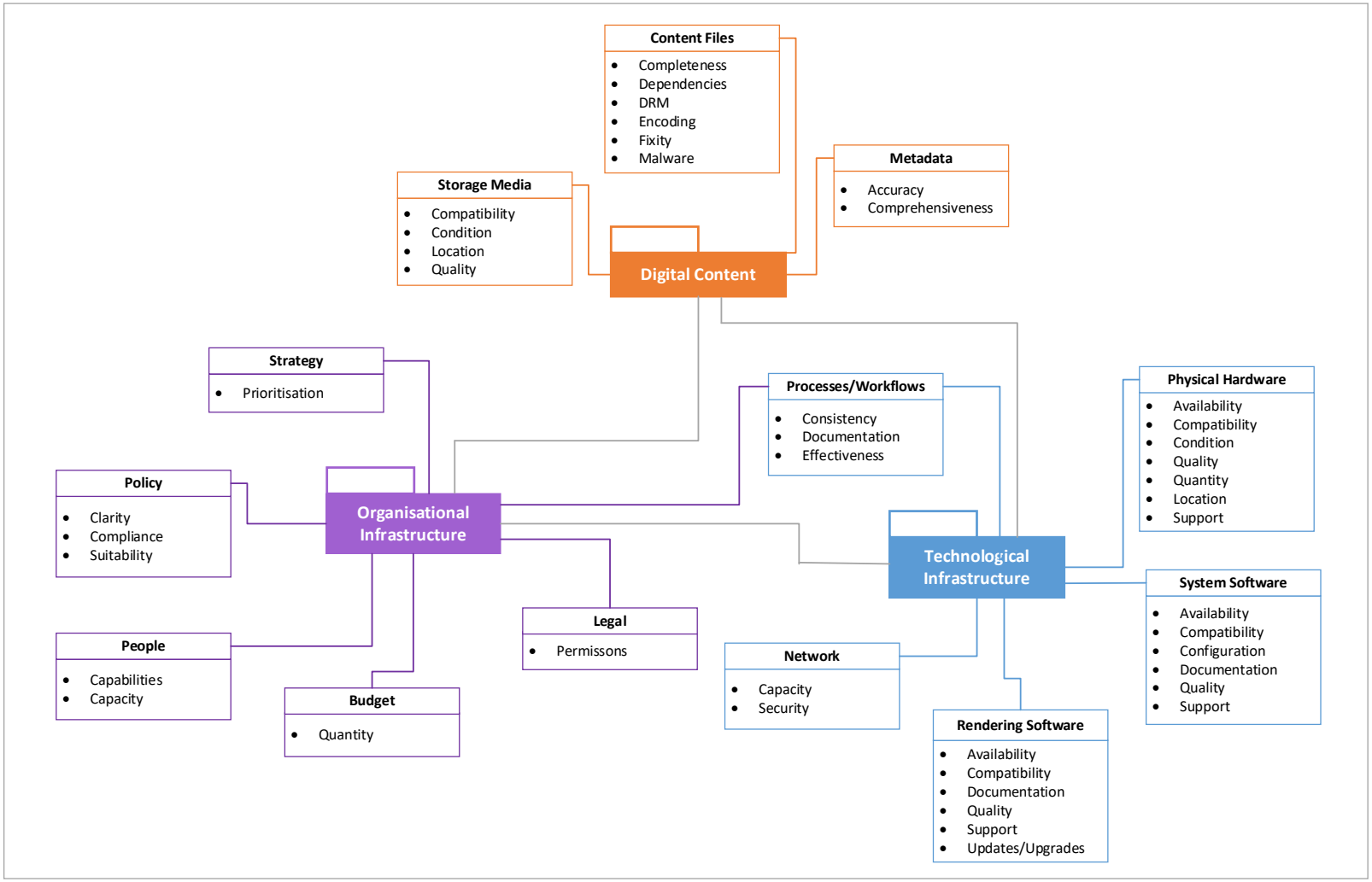

Figure 11: L1 Digital Preservation Risk Source Model

Each Risk Originating Entity and their Risk Source Classes, Instance Types and Risk Source Factors is discussed and explored in more detail below. Descriptions of individual classes, instance types and factors draw upon the author's kernel knowledge, practical experiences, and the solution/literature review presented in chapter three. Terms have been tested and refined through the construction of over 150 individuated risk statements, as well as through the practical use of the risk source model to develop a series of scenario-based risk assessments at the British Library for a number of different technical systems and collections. A description of those is provided in the case study section of chapter six, whilst concise descriptions of key terms are located in the Glossary. Factors and Instance Types however are not individually defined, as precise definitions were found to limit the flexibility of their relevance and application.

Organisational Infrastructure

The **Organisational Infrastructure Entity** contains six classes of risk source: Strategy; Legal; Policy; People; Budget, and; Processes/Workflows. Each is associated with one or more risk factors that, unless optimised, are a potential source of a negative outcome.

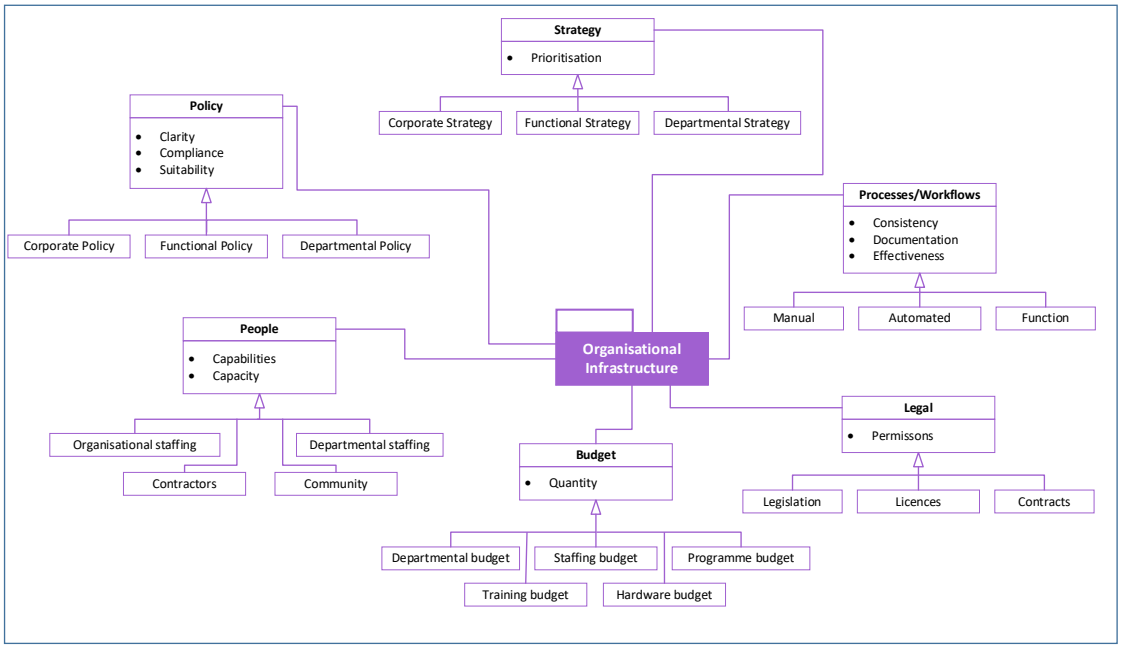

Figure 12: L2 Organisational Infrastructure Risk Area

The **Strategy** class represents the importance of strategic direction and organisational support for digital preservation. Organisations may have multiple

strategies that relate to or otherwise affect digital preservation activities. If the direction or priorities of a given strategy do not explicitly support digital preservation, then this may represent limitations in organisational support and buy-in for digital preservation activities. This can make it difficult to ensure digital preservation is properly resourced and to implement digital preservation principles in business activities.

The Risk Factor for this class of risk source is therefore *Prioritisation.*

Risk Instance Types for this class of risk source include *Corporate Strategies, Functional Strategies, and Departmental Strategies.*

The **Policy** class represents the policy ecosystem required to establish and ensure that digital preservation activities draw upon a series of appropriate and consistent principles. In some organisations this may be a single digital preservation policy; in more complex organisations it may represent a series of policies relating to different collections, departments, and/or functions. Irrespective of the composition of the policy ecosystem, it is important that the policy framework suitably and clearly covers all necessary digital preservation activities, and that a mechanism is in place for monitoring and ensuring compliance with policy.

The Risk Factors for this class of risk source are summarised as *Clarity, Compliance,* and *Suitability.*

Risk Instance Types for this class of risk source include *Corporate Policies, Functional Policies, and Departmental Policies.*

The **People** class represents the staffing or personnel resource associated with digital preservation. Whilst technology is an essential component of a digital preservation service, it needs people to implement the service and ensure a preservation objective can be achieved. For this to occur successfully, there must be enough staff working in suitable roles, with appropriate skills to do their work.

The Risk Factors for this class of risk source are summarised as *Capacity* and *Capabilities.*

Risk Instance Types for this class of risk source include *Organisational Staffing, Departmental Staffing, Contractors,* and *Community.*

The **Budget** class represents the sum of the financial envelope that supports an organisational digital preservation endeavour. It has implicit links to many other risk sources and factors, with funding required for staff posts and technological infrastructure as well as keeping staff skills up to date, to pay licence fees, to support tendering processes, and to update infrastructure when required. As a risk source in its own right, sufficient budget must be available for the organisation to achieve its objectives. This has several associated perspectives: it must be distributed appropriately and with a multi-year perspective in order to support the ongoing nature of long-term digital preservation activities.

The Risk Factor for this class of risk source is summarised as *Quantity*.

Risk Instance Types for this class of risk source include *Departmental Budget, Staffing Budget, Programme Budget, Training Budget,* and *Hardware Budget.*

The **Legal** class represents the legislative and contractual framework that surrounds digital preservation in an organisational setting. This might include, for example, legislation around copyright or electronic legal deposit, restrictions on what can be done with content, and legally binding contracts relating to organisational use of software, systems, or third party services. It is important that legal and contractual frameworks in a digital preservation environment do not prohibit activities that are needed for digital preservation, and preferable that they expressly allow such activities. This might include, for example, freely licensing the use of legacy software for emulation or in emulated environments, or depositing software in escrow and granting a licence for perpetual re-use should an important third-party service provider go into administration.

The Risk Factor for this class of risk source is summarised as *Permissions.*

Risk Instance Types for this class of risk source include *Legislation, Licences, and Contracts.*

The last risk source class in this section is **Processes and Workflows,** also associated with the Technological Infrastructure entity. It is associated with both risk originating entities to acknowledge that whilst processes and workflows are implemented in a predominantly technological environment, they often require manual input and are not all computer based or automated. Moreover, the manner in which organisations implement processes and workflows varies from organisation to organisation, influenced particularly by the level of maturity in a given organisation, the amount of funding available, the number and type of staff associated with digital preservation, and so forth.

Processes and workflows should effectively achieve their objectives. They should not introduce unnecessary complications or complexities as this may make them more difficult to manage over time, especially if operating at scale. Documentation of processes and workflows enables accurate implementation and also facilitates an ongoing understanding over time despite inevitable staff changes. Moreover, processes and workflows should be consistent with one other. This might be achieved through implementation of a shared set of principles, use of the same core set of tools for a given function across different types of content, or consistently scheduling a function (such as validation) at the same point in a workflow for different types of content. This makes the outputs and processes easier to predict and to manage, especially at scale.

The Risk Factors for this class of risk source are summarised as *Consistency, Documentation,* and *Effectiveness.*

Risk Instance Types for this class of risk source include *Manual, Automated, and Function-based processes*.

Technological Infrastructure

The **Technological Infrastructure Entity** contains five classes of risk source: Rendering Software; System Software; Physical Hardware; Network, and; Processes/Workflows (shared with Organisational Infrastructure). Each is associated with one or more risk factors that, unless optimised, are a potential source of a negative outcome. In practice, each class of risk source is inherently connected to several others in a manner more explicit than in the organisational infrastructure. This is due to the complex arrangement of dependencies that exists across a technological infrastructure – in essence, the network underpins and connects the physical hardware, upon which runs system software and application software, using processes and workflows designed to meet the needs and objectives of an organisation. Collectively, this is simply how the technological infrastructure operates. The model however presents them as distinct entities in order to support precise specification of risk source relationships at an instance type level.

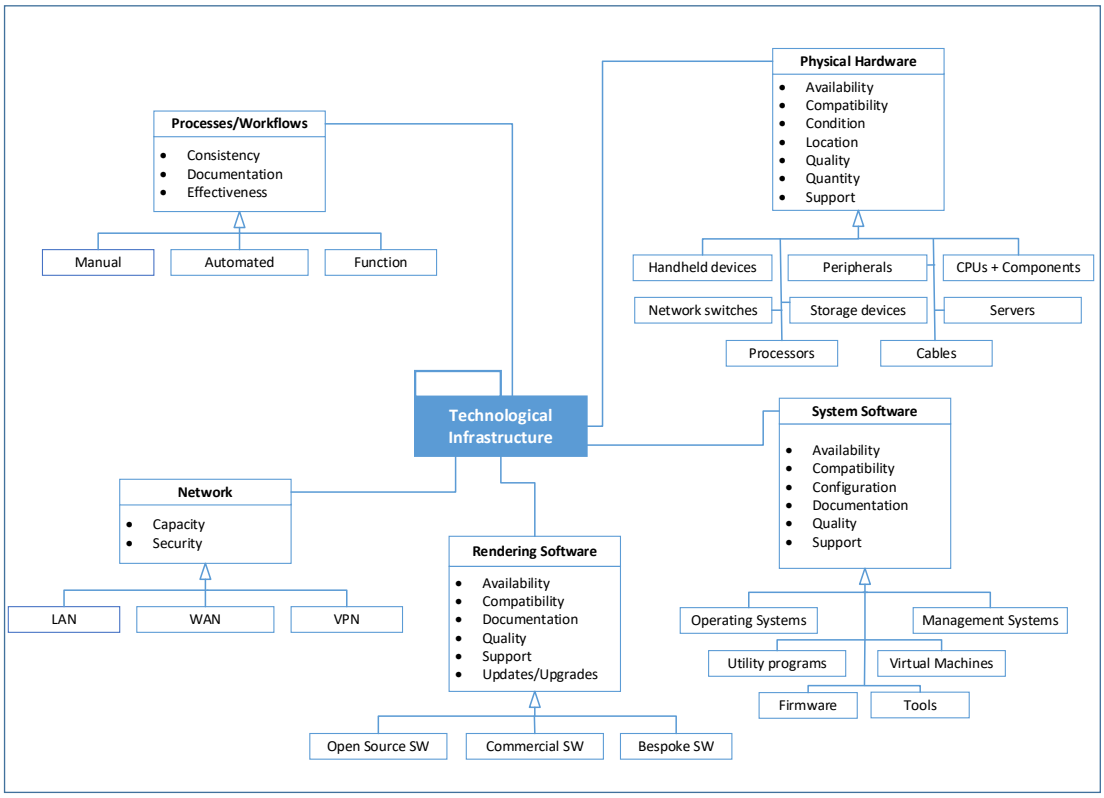

Figure 13: L2 Technological Infrastructure Risk Area

The model represents different types of software as two distinct classes of risk source. The first is Rendering Software, which is the software used specifically to

process digital object files so that their intellectual content becomes accessible. The second is System Software, which broadly consists of a range of different management programmes such as operating systems, tools, or digital library software. Both are software and are associated with similar sets of risk factors. However, they present as two different classes of risk source as they represent different types of instances that in turn represent different types of contextualised risk sources. They subsequently have potential to result in different types of negative outcomes, particularly in terms of target values. They may also be interdependent, whereby software from one class (e.g. rendering software) interacts directly with the file whilst software in another class (e.g. operating system software) enables that interaction. This interaction is common in all contexts though it takes on particular significance in a digital preservation setting where the interaction between the two may have consequences for the authentic rendering of the object. They are therefore classed in the model as two separate classes of risk source.

The **System Software** class represents a range of different computer programmes that facilitate management of a technological infrastructure and the digital objects held within. System software provides an environment within which rendering software is deployed and used, as well as tools for managing that environment and the files within it. This includes operating systems, utility programmes, and firmware, as well as complex digital repository and library management software. Within a digital preservation setting, this also includes preservation tools used in association or independently of a repository system, such as those that support the preservation functions of integrity checking, format identification, validation, and so forth.

Appropriate, good quality system software must be available and documented so that it can be both deployed and used effectively. Software is not an isolated component of the infrastructure so it must be compatible with other components, with which it interacts. Configuration changes can however significantly affect the way software performs so must be carefully managed. For example, changes to antivirus software may result in previously ‘clean’ files identifying as infected and

subject to quarantine - or worse, automatic deletion. Configuration changes to library management systems or tools may result in changes in the processes used by those systems, which can lead to uncertainty about their performance and outputs. Changes to permissions controls associated with different types of instances can allow inappropriate interaction with systems and content, benign or malicious, by internal or external users. These can affect various target values depending on the nature of the change and any subsequent infraction. Finally, as software becomes older there is a greater likelihood of obsolescence due to withdrawal of vendor support (inevitable in a predominantly commercial market driven by software and hardware updates supporting new and improved performance models). This may also result in a product becoming out of sync with cyber security requirements, leading to revocation of permission to deploy the software on the network. It must be carefully managed so as not to adversely affect an organisation's ability to manage, find, or provide access to content.

The Risk Factors for this class of risk source are summarised as *Availability, Compatibility, Configuration, Documentation, Quality,* and *Support.*

Risk Instance Types for this class of risk source include *Operating Systems, Management Systems, Utility programmes, Virtual Machines, Preservation Tools,* and *Firmware.*

The **Rendering Software** class represents computer programmes that process digital files in order to access and render a digital object and its intellectual content. Rendering Software deciphers encoded files according to a given format specification or standard and interprets them so that they can be accessed and understood by users. Depending on the nature of content and software, it may also allow users to interact directly with the objects, for example by running queries, calculating formulae, searching or exploring the contents in a non-linear fashion, generating audio, or even overwriting original content.

Appropriate rendering software must – like system software - be available and documented so that it can be deployed and used effectively. Software must be

compatible with other aspects of the infrastructure upon which it relies, such as operating systems, and be of good quality with minimal avoidable bugs or glitches that affect performance and rendering. Maintenance updates or enhancements to the source code can result in new versions of the software application (major or minor). These need to be carefully managed, as whilst they may be essential, they can change how the software interacts with files and how content is made available to users. Moreover, as rendering software becomes older there is a greater likelihood of obsolescence due to eventual withdrawal of vendor support (inevitable in a predominantly commercial market), with the potential result that the product drops out of sync with the wider technological infrastructure and becomes incompatible with cyber security requirements so can no longer easily be deployed. This too must be managed before it becomes an issue.

The Risk Factors for this class of risk source are summarised as *Availability, Compatibility, Documentation, Quality, Support, and Updates/upgrades.*

Risk Instance Types for this class of risk source include *Open Source Software, Commercial Software, and Bespoke Software.* [67]

The **Physical Hardware** class represents all of the tangible machines, wiring, and other physical components needed to support a technological infrastructure. This includes everything from servers, processors, storage devices and network switches to components such as motherboards, graphics cards and sound cards, as well as wiring, cables, and handheld portable devices such as tablets and laptops.

As with the software classes, this class is associated with multiple different risk factors. Not all are within the power of an organisation to immediately prevent. Physical deterioration of hardware over time, for example, cannot be wholly avoided. Deterioration can affect the integrity of files stored on hardware, and on the ability of the hardware to support processes relating to target values of

[67] These types were selected for representation in the model over more specific types of software (e.g. those that associate with different types of content) for reasons of visual brevity and broad relevance, as all may be relevant to an institution regardless of the type of content they hold.

accessibility and retrievability. Environmental conditions and management protocols should therefore be optimised for hardware longevity and to delay deterioration for as long as viable, given other associated risk factors. For example, as hardware becomes older there is a greater likelihood of obsolescence due to withdrawal of vendor support, for the same reasons as with software. When support for hardware is no longer available, it becomes more difficult to source replacement parts when original parts break, to the point where maintenance becomes prohibitively expensive. Support lifespans should thus be carefully monitored alongside environmental conditions as loss of support may alter a risk tolerance profile for changes in environmental conditions. Obsolescence can also mean hardware is no longer capable of supporting modern software that is designed to run on faster and more powerful or simply different types of machines. As a result, technical dependencies may no longer be so easily supported and access to content becomes an issue.

Re-configuration of a hardware estate – inevitable over time – should be carefully managed as it can introduce uncertainty about whether component parts will still work together as expected to support the requisite processes. The quality, quantity, and type of hardware used also inevitably changes over time, as may the location of hardware, which can affect the viability of security arrangements and the likelihood that a location may be affected by natural disaster such as flood or fire. All require careful consideration. Last but not least, compatibility across the estate is a potential issue: hardware must be appropriate to the needs of the software that it must support – from server stacks to individual PCs and handheld devices. Failure to address this can result in failure to demonstrably achieve target values across the board.

The Risk Factors for this class of risk source are summarised as *Availability, Compatibility, Condition, Location, Quality, Quantity, and Support.*

Risk Instance Types for this class of risk source include *handheld devices, peripherals, CPUs and components, network switches, storage devices, servers, processors, and cables.*

The **Network** class represents the communication protocols in place to enable the exchange and transfer of data and resources across much of the physical hardware utilised in a technological infrastructure. It is thus closely related to the physical hardware class due to its inherent dependency on hardware for data transfer, as well as the system software class for management interfaces. Networks enable the transfer of content and metadata through different processes, for example deposit or acquisition via FTP, processing and cataloguing on a local network, transfer to storage and replication on different sites, and retrieval for access by end users.

This class has relatively few risk factors but they are nonetheless important. The network must have sufficient bandwidth (capacity) to support the expected functions and data transfers, otherwise transfer protocols and policies may not perform as expected and content can be lost. Good network security protocols are essential, as they are often the first line of defence against cyber-security attacks, which can result in the loss of access to systems and contents with potentially disastrous results. Protection requires deployment and frequent maintenance of reliable system software, including firewalls and anti-virus programs as well as appropriate access controls such as user password management solutions and multi-factor authentication. Configuration management tools support both security needs and the appropriate flow of data across a network, as well as monitoring the network for any issues. Loss of connectivity from failure to maintain an appropriate network can lead to problems and uncertainties associated with several different target values, as well as related risk sources in other parts of the infrastructure.

The Risk Factors for this class of risk source are summarised as *Capacity and Security.*

Risk Instance Types for this class of risk source include *Local Area Networks (LAN), Wide Area Networks (WAN) and Virtual Private Networks (VPN).*

The **Processes and Workflows** class is described previously in association with the Organisational Infrastructure entity. From a technological perspective, it remains important that processes and workflows should effectively achieve their objectives and should not have gaps or unnecessary complexities that may make them more difficult to manage. Documentation of workflows is essential so they can be appropriately implemented and understood over time, particularly in relation to any direct action upon files as this may affect authenticity and integrity. Such documentation requires updating as workflows change. Workflows should also, wherever possible, be consistent across different types of content – if not then they can introduce uncertainty and unnecessary technological complexities into overall collection and management processes. The risk factors can thus manifest in a slightly different way from when associated with an organisational infrastructure and thus may require consideration of a wider range of source/factor relationships.

The Risk Factors for this class of risk source are summarised as *Consistency, Documentation, and Effectiveness.*

Risk Instance Types for this class of risk source include *Manual, Automated, and Function-based workflows.*

Digital Content

The **Digital Content Entity** contains three classes of risk source: Content File(s), Metadata, and Storage Media. Each is associated with one or more risk factors that, unless optimised, are a potential source of a negative outcome. In practice, the Digital Content entity always has an explicit and direct dependency on both other risk originating entities, as these represent the context in which preservation is implemented. The full risk source model, available as a separate file within the practical submission that accompanies this thesis, reflects this connection.

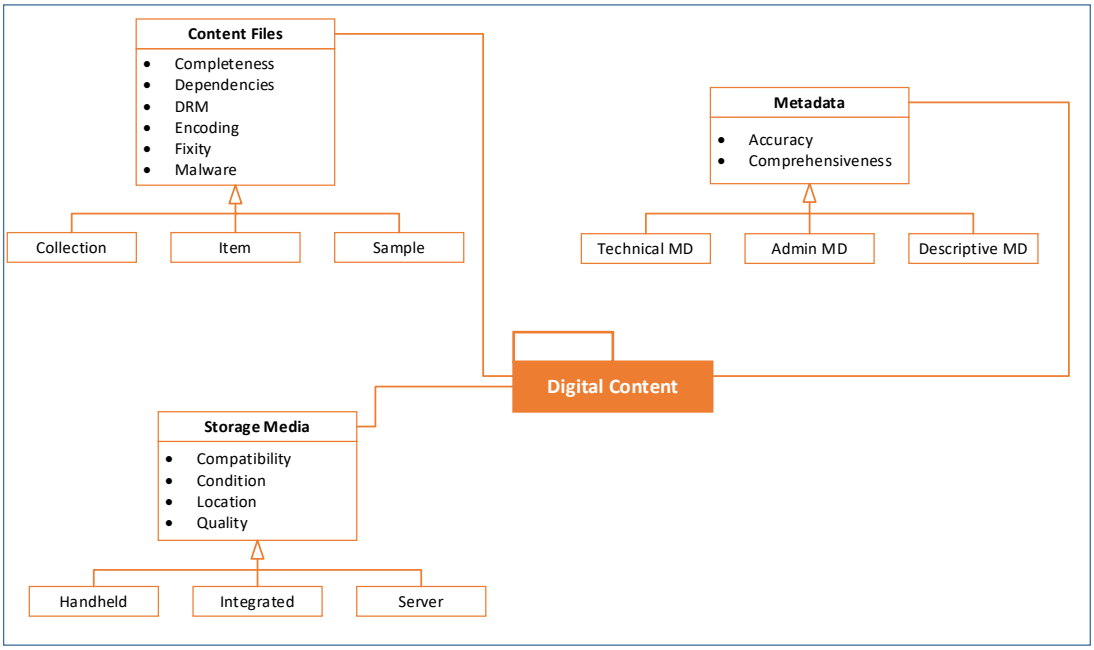

Figure 14: L2 Digital Content Risk Area

The **Content File(s)** class represents the digital file(s) in which an object and its intellectual content is encoded, typically in the structure of a given file format. This might be, for example, a PDF file representing a born-digital eBook, an ePub package containing a set of XHTML and PNG files, JPEG files that represent a digitised book, audio files from an album in MP3 format, a collection of TIFF files that each represent pages of a digitised newspaper, or a collection of GML files representing geospatial mapping data. Some types of content files may also contain executable programmes to access and render the intellectual content, thus blurring the line between content files and rendering software. Mobile apps in the iOS Apple container format are one such example, where content and application programme are all packaged together in the same container. Such cases therefore require cross-referencing with factors with the rendering software class.

Risk factors for this class are predominantly technical though the practical matter of completeness must also be addressed. This ensures all content files associated with a digital object or collection are available and that they collectively represent all required content. This may seem obvious but it is not uncommon for some types of objects (such as web archives) to utilise third party hosted content, or to require specific fonts (such as for mathematical formulae) not commonly packaged with

rendering software libraries. Aside from this, file fixity must be confirmed to demonstrate that a bit stream remains unchanged since it was last checked and its integrity is intact. Files should be accessible when needed without obstruction from integrated digital rights management protocols, such as active password protection or licence keys. Files must also be checked for malware as this can affect file longevity and other target values if left untreated.[68] Encoding must be known, at both file format level (e.g. *.doc or *.jpg), and character level (e.g. ASCII, UTF-8 or UNICODE). Data compression is another possible form of encoding that, if present, needs to be identified to ensure its suitability and so that files can be uncompressed when required, using appropriate tools. These encoding issues represent a particular type of technical dependency between files and software, though other technical dependencies - such as upon certain types of hardware, peripherals, or operating system software - may also exist that can affect an ability to satisfy target values.

The Risk Factors for this class of risk source are summarised as *Completeness, Digital Rights Management (DRM), Dependencies, Encoding, Fixity,* and *Malware*

Risk Instance Types for this class of risk source might include *Collection-level Files, Item-level Files, or Samples of Content Files.*

The **Storage Media** class represents the physical devices upon which files are stored. Storage media is a type of physical hardware and a type of instance associated with that class, but it is also represented in the model as a risk source class in its own right when used for storage of digital content. This acknowledges the essential role of storage media as tangible carrier of otherwise intangible digital objects, and supports scenarios in which content exists independently of integration into a given technological infrastructure (e.g. immediately after deposit or in institutions with emergent preservation programmes and limited

[68] Malware does not necessarily need to be removed from objects – whilst its presence may pose a risk to longevity, automatic removal may introduce secondary risks around authenticity and integrity. Mitigation should therefore be an institutional decision based on the specific collecting context, the technological infrastructure, and type of malware found.

technological capabilities). Storage media types range from handheld media such as DVDs, CDs, floppy disks and USB drives, to integrated storage media such as solid state drives (SSDs) or hard disk drives (HDDs) and purpose-built large scale storage devices such as RAID disc storage systems or server storage.[69]

Regardless of type, it is essential that the storage media supports preservation and retrieval of the digital files it holds, for as long as needed, and that it is maintained in good condition. Physical deterioration of media over time is inevitable, though lifetimes can be maximised through good environmental controls, appropriate handling procedures, and use of high quality artefacts. Use of alternative storage media such as DNA or quartz glass can significantly increase the storage lifetimes associated with more commonly used storage media, though they come with their own challenges (and by association, risks) around costs and retrievability. With age comes also an increased likelihood of technological obsolescence, manifesting as a decline in manufacturing the media and a lack of compatible drives or programmes with which to read the media and extract files. Compatible hardware and software must therefore be located and maintained or supported for as long as the storage media is in active use. The location in which storage media are held must also be carefully considered, with sufficient security in place to prevent accidental or malicious damage, protection against environmental/pest damage, and unlikely to suffer natural disasters. Failure to address these issues can affect retrievability, accessibility, and the integrity of files stored on the storage media.

The Risk Factors for this class of risk source are summarised as *Condition, Compatibility, Location,* and *Quality.*

Risk Instance Types for this class of risk source include *Handheld Storage, Integrated Storage,* and *Server Storage.*

[69] Cloud storage may also be considered here, though it might alternatively be considered a third party service managed under contract. This example illustrates the flexibility of the model to support different interpretations of infrastructure components and the nature of risk sources, as well as their associated risk factors.

The **Metadata** class represents information about digital content that supports its preservation and access. Metadata can take different forms and be stored in different locations, including individual metadata files, embedded within content files, a metadata database, a catalogue, or another type of management system. Regardless of where or how it is stored, appropriate identifiers link metadata to the content that it describes. In a library or archival setting, metadata typically conforms to one or more metadata standards and supports additional functions beyond preservation, such as cataloguing, rights management, and description. It may be described as administrative, technical, descriptive, structural, or preservation metadata, though in practice there is generally some overlap between these categories. From a strictly preservation perspective, its focus is often either technical or administrative, describing for example the technical composition of content files, the relationships between files, the history of the files, or the technical environment appropriate for accessing the content files.

Regardless of the structure, nature or location of the metadata, it is important that it accurately and correctly represents the content files it describes. Accurate fixity and format information, for example, can be used to demonstrate the ongoing integrity of content files and enable identification of compatible rendering and access solutions for subsequent authentic rendering of the content. It must also accurately identify relevant content files and their location, so they are retrievable with relative ease, as well as relationships between component files for an object, or files that represent different versions of objects (for example in cases where publishers have submitted corrected versions of an article, or preservation action has been undertaken to produce a new representation or derivation of content). Metadata must be comprehensive and fully comply with organisational specifications and requirements so that it meets the needs of the holding institution and supports its overall objectives.

The Risk Factors for this class of risk source are therefore summarised as *Accuracy* and *Completeness*

Risk Instance Types for this class of risk source include *Technical Metadata, Administrative Metadata, and Descriptive Metadata.*

Conclusions

This chapter has taken a risk science approach to disentangling digital preservation risk, moving beyond abstract concepts of uncertainty and loss to establish a more insightful and useful conceptual foundation from which to explore digital preservation risk. It establishes a definition of digital preservation that integrates key themes and concepts from across the literature, and from this develops a contextualised definition of the concept that frames uncertainty in relation to specific target values, undesirable negative outcomes, and high level sources of risk. The risk source concept is further explored to identify a series of related entities that collectively represent the significant structural components of a risk source. These are subsequently used as a framework with which to unpick the complexity of digital preservation risk, and through which to represent it more thoroughly and consistently than seen elsewhere to date.

This application of a risk science approach to disentangle and represent the nature and complexity of digital preservation risk enables a more structured, reasoned, transparent, and holistic representation of risk than previously seen within the field. The digital preservation risk source model in particular represents transferable knowledge that can inform risk assessments in practical settings. The clear and consistent representation of risk entities facilitates a consistent understanding of different types of digital preservation risks, regardless of their origins. This in turn provides the structure for a consistent representation of practical risks, which can form the foundation of a flexible and comprehensive preservation planning risk response. Methods for achieving this are explored in the next chapter: demonstrating the solution.

Chapter Six: Demonstrating the Solution

Introduction

This chapter demonstrates how the digital preservation risk source model outlined in the previous chapter can be used in a practical setting to develop situational knowledge of practical, real world risks, thus moving from an abstract, conceptual overview of risk sources to a characterisation of likely risks in relation to a specific context and scenario. Demonstration is stage four of the design science research methodology devised by Peffers *et al.*, (2006), establishing the efficacy and utility of the artefacts designed so far to solve the problem at hand. It represents the application of the solution to the problem space through the introduction and use of a third artefact type: the method.

This chapter presents three different methods for using the risk source model, two of which focus on risk identification whilst the third extends to a full risk assessment process. The first method is a simple, conversational approach that uses the model as the basis of an exploratory conversation about risk in a given scenario. The second is a question-based framework that guides an assessor through a series of questions about each class of risk source and risk factor in the model for a given scenario, translating each of the combinations into single questions with a yes/no/uncertain answer. The output of either of these methods can support the third method, in the form of a full risk assessment. This produces characterised risk statements for a given scenario, with precise and contextualised specifications of risk sources. It provides another mechanism through which to consistently and thoroughly represent digital preservation risk in support of a comprehensive preservation planning risk response, though at a more practical level than represented in the risk source model. The chapter concludes with a summary of how some of the methods have been used at the British Library to generate a series of risk assessments for a range of different scenarios. This demonstrates not only the utility and efficacy of the model but also the methods themselves.

Demonstrating utility: An artefactual approach.

This thesis explores the nature and complexity of digital preservation risk through the framework of a design science methodology. Design science research is a knowledge generating endeavour characterised by creativity, innovation, scientific rigour, and practicality, that solves problems through the production of artefacts (Iivari, 2007; Hevner and Chatterjee, 2010; Baskerville *et al.*, 2019; vom Brocke, Hevner and Maedche, 2020a, p. 3). Design science research artefacts typically take one of four related forms: construct, model, method, and instantiation (March and Smith, 1995; Hevner *et al.*, 2004; Gregor and Hevner, 2013), though a fifth artefact type is also increasingly recognised in the form of theory (Gregor, 2006; Gregor and Hevner, 2013; Baskerville *et al.*, 2018; Iivari, 2020). As such, this research produces several types of artefacts to support the exploration and establish a comprehensive understanding of the nature and complexity of digital preservation risk.

The first of these, the *theory*, is that the application of risk science principles enables a more rigorous, justified, transparent and reasoned understanding of the concept of digital preservation risk than previously seen within the field. This theory provides the basis from which a series of meaningful, logical, and variously granular *models* are developed that represent the digital preservation risk domain at an abstract level, specifically in terms of risk sources. Key terms relating to the models are defined in the corresponding *construct* or glossary of terms. The construct and the models collectively represent the overall nature and complexity of digital preservation risk at a domain level and are represented in a single 'Reference Model' documentation artefact that forms the core of the practical submission for this PhD. However, without a *method* for using the model, it is difficult to effectively demonstrate how it can function to form the foundations for a flexible and comprehensive planning risk response.

Methods are a critical component through which to encourage and enable consistent re-use of digital preservation models and frameworks by the wider community (Maemura, Moles and Becker, 2017, p. 1630). This chapter presents a number of potential methods for using the risk source model in order to demonstrate its utility, which is a principle of design science research (March and

Smith, 1995, p. 253). As mechanisms for demonstrating utility, these also represent the demonstration stage (stage four) of the design science research methodology used to progress this research and support a subsequent evaluation of the model in stage five.

Whilst a method can be a valid design science artefact in its own right, not all methods are valid design science artefacts. As an innovative and creative endeavour, the status of the method in design science research is determined by its nature and the development process – if it is not a novel method, then it does not necessarily represent a significant research contribution (March and Smith, 1995, p. 261). This perhaps goes some way to indicating why methods are relatively infrequently explored in design science research literature compared to other artefact types (Winter, 2008; Thuan, Dreschler and Antunes, 2019).[70] The methods explored in this chapter are presented on the basis of their practicality and utility rather than explicit design science research artefacts. The evaluation in the following chapter nonetheless considers them from a design science perspective in order to provide some insight into their status.

A Method is simply a set of steps through which to apply a solution to a problem. In design science research, methods can represent various different processes or techniques to solve a problem, from mathematical algorithms (Hevner *et al*., 2004, p. 79), to graphical representations (Dresch, Lacerda and Antunes, 2015, p. 109), conversational processes to explicate desirable knowledge (Glassey-Previdoli, Bonazzi and Viscusi, 2021) and data analysis (Gnewuch and Maedche, 2022). March and Smith observe that methods can use parts of a model for input, or even translate a model from one representation to another in the course of problem solving (March and Smith, 1995, p. 258).

The problem area addressed by this research is risk, including the inconsistency within the community on the vocabulary used to define and describe risk, which has inhibited significant advances towards enterprise solutions for digital

[70] Analysis of over one hundred DSR publications submitted to the DESRIST conference series (Thuan, Dreschler and Antunes, 2019) found that only ten percent of the outcome artefacts represented methods.

preservation risk management. The work described in this thesis to develop a deeper conceptual and abstract understanding of digital preservation risk provides much-needed clarity into its nature and complexity. There are however many established risk assessment methods through which this conceptual understanding of risk can support the foundations for a more flexible yet comprehensive preservation planning risk response, originating from outside of the digital preservation field. ISO 31010 on Risk Assessment Techniques, for example, lists no less than forty-two different methods that can support all or part of the overall risk assessment process (ISO, 2019). These range from interviews, checklist reviews, and event tree analysis, to consequence/likelihood risk matrices or probability models such as Bayes theorem and Monte Carlo simulations, using empirical data or expert judgement.

Selection of an appropriate risk assessment technique is heavily influenced by the context of the risk assessment, considering many variables from the overall purpose of the assessment to the data available, the skills and knowledge of the risk assessors, and disciplinary or organisational risk assessment norms. The end outcome may remain open to interpretation regardless of the method chosen, as a completely objective risk assessment is a challenging endeavour (Aven, 2013 p.466). Klinke and Renn explore this within the context of the philosophical risk management debate between constructivism and realism, whereby realists believe that 'technical estimates of risk constitute true representations of observable hazards [with] calculated results', whilst constructivists perceive risk assessments as 'mental constructions that can be checked at best against standards of consistency, cohesion, and internal conventions of logical deduction' (2002, pp. 1072 - 1073). This distinction can also apply to quantitative and qualitative assessments, with quantitative proponents preferring statistical reasoning whilst qualitative assessors prefer a more nuanced, constructivist stance. In line with the underlying philosophy of this thesis, a middle ground is perhaps the most pragmatic response whereby either position is valid if accompanied by adequate explanation and contextualisation. Rigour is possible for either type of approach, to lead to a solid,

justified set of outputs. Transparency of process, contextualisation, and clear definitions of terms all have an important role to play in this regard.

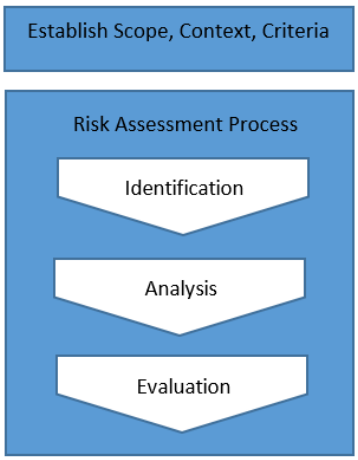

Figure 15: ISO 31000 risk management stages supported by the methods (ISO, 2018)

All methods in this section support one or more stages of the ISO 31000 risk management process (ISO, 2018), particularly those relating to establishing scope and context, and the three main risk assessment stages of identification, analysis, and evaluation. *Scoping* relates to expectations around objectives, particularly the objective of a given risk management process and its alignment with organisational objectives – in this case the preservation objective. *Contextualisation* focuses on the internal and external context of the process, namely the specific environment of the process and the stakeholders affected by it. *Identification* describes risks that may affect objectives, whilst *Analysis* explores the level of risk involved and *Evaluation* determines whether action is advised (though falls short of prescribing any specific form).

A meaningful risk assessment takes time to produce and a degree of adaptation or application of a risk model to a given scenario is typically required, especially in qualitative approaches. The more thorough the model, the more thought, reflection, and consideration is required. There are no quick or easy answers - a thorough review is essential to move away from a 'motherhood and apple pie' approach to risk and generate meaningful, useful insight. The assessment process itself requires in-depth knowledge of the area under investigation, with input from a team of stakeholders who have relevant knowledge of the designated scope and context (Ostrom and Wilhelmsen, 2019, p. 27; Aven and Thekdi, 2022, p. 262). This is particularly important for qualitative approaches, where such knowledge is often the primary source of data around which the assessment is constructed. Each of the methods presented here is therefore designed for leadership or coordination from a knowledgeable digital preservation practitioner, ideally one with experience in

risk assessments and with input elicited from wider relevant stakeholders as appropriate to the designated scope.

The methods in this chapter each represent a different way in which the digital preservation risk source model can be used to explore digital preservation risks, demonstrating its utility and potential applications. Methods One and Two are exploratory and support primarily the Identification part of the risk assessment process, whilst Method Three is a qualitative approach that builds on these and extends to the full risk assessment process including Analysis and Evaluation. This qualitative approach is consistent with the established approach to risk management already used at the British Library, where this research will be implemented. A quantitative method could, in theory, also be developed should a third party have access to reliable data to underpin such an approach.[71]

Method One: A conversational, stepwise process

This first method represents a communication-based process, using the risk source model as its focal point. It is a lightweight method designed to engage stakeholders in an exploratory discussion of potential sources of risk in a given scenario. As such, it supports the Identification stage of the Risk Assessment process.

The goal of this method is to identify broad areas of concern in a given scenario, within the context of the wider potential risk landscape. This method is particularly well suited to those who already have a good understanding of the risk source model and are able to independently guide a discussion on risk without requiring an explanation of each risk source and factor. It can also function as a standalone method to support risk identification in organisations where a full ISO 31000 risk management and assessment process is not appropriate or required, for example when other structured processes are already in place. In such cases, the outputs of

[71] The current lack of reliable data on the risk sources and factors featured in the model suggests such an approach should be treated with caution, as the accuracy and thus usability of the outputs is dependent on the fidelity of the underlying dataset. Methods such as Delphi panels and expert opinion elicitation can go some way to mitigate this but they too must be treated with caution, as they inevitably represent subjective disciplinary biases especially in an emergent field with limited practical experience of the suggested risks. As noted earlier however, Aven (2013 p.407) observes that completely objective risk assessments do not exist and that all are subjective, regardless of method.

this method can inform those processes. Alternatively, it can function as an initiation and identification step, prior to use of method three.

This method has four main steps: define scope and context; identify stakeholders; discuss model and context to identify potential risks; collate outputs. It is a loosely defined and flexible method that should be tailored to the needs of the individual assessor or scenario.

Step #	Description	Function	Suggested output
1	Define scope and context	Initiate process	Document
2	Identify and contact knowledgeable stakeholders	Establish input sources	Scheduled meetings
3	Discuss model and context with stakeholders	Identify areas of concern	Notes
4	Collate outputs from stage 3	Overview	Collated notes

Table 4: Stages of the Conversational Stepwise Method

Step One consists simply of defining the scope and context for the discussion, noting the reason (or trigger) why a risk investigation is required, the senior sponsor for the investigation, and any governance or reporting lines in place. Details on the sponsor and governance arrangements help generate credibility on the process for stakeholders. A tightly defined scope can support a more precise and focused discussion than if it is only described in vague or open-ended terms. This step initiates the process and should result in some form of documentation to share with stakeholders. This might be, for example, an email, a wiki page, or a document on a shared drive. The preservation objective is stated in this documentation as part of the scoping exercise.

Step Two involves identifying and contacting appropriate stakeholders. This requires knowledge of both the scoped area and the institutional context. The scoped area may represent a collection or a system, whilst the institutional context helps to identify stakeholders in relevant roles. Stakeholder meetings can be scheduled, recommended at between 30 – 60 minutes each. Individual meetings with different stakeholders elicit different perspectives and reduce the likelihood of overall influence by a single or dominant participant's ideology.

Discussions occur in *Step Three*, repeated as many times as necessary with individual stakeholders until the coordinating expert is confident that all relevant risk areas are addressed. A printed version of the risk source model is useful during these meetings as a visual prompt and aid to talk through different potential risks and risk sources. Whilst the whole model is in scope for discussion, it is not expected that each class of risk source will be relevant to every scenario. Assessors should therefore exercise reasoned judgement, based on their knowledge and experience, in the depth to which each source and instance type is discussed. Notes can be taken to record the main areas of concern identified through a discussion and capture any additional information.

Step Four consists of review and consolidation of discussion notes with different stakeholders into a single, shareable document, circulated to relevant stakeholders for comment. Feedback is recorded and the assessor must balance or otherwise adequately represent dissenting or conflicting opinions. The final output is then shared with the senior sponsor and the next steps agreed.

In some cases, this method may identify very little cause for concern and no further action is to be taken.[72] This is a pragmatic response given the overhead required for a full risk assessment process. Documentation should nonetheless be retained for a reasonable period in line with organisational retention schedules.

Method Two: A question-based framework

The second method represents a more structured exploration of the risk source model than the first. It translates the model into a series of structured questions about each class of risk source and risk factor, and guides an assessor or assessment team through the questions to consider whether any risks are associated with each class in a given scenario. It functions as a translated form and alternative representation of the model, production of which was also a helpful process for assessing the internal consistency of the model and contributing to formative evaluations during the final stages of the model's refinement.

[72] This outcome is most likely when the risk review occurs on a scheduled basis rather than being triggered by a particular event or cause for concern.

As with method one, the question-based framework primarily supports the Identification stage of the risk assessment process. Question-based approaches are standard research tools through which to gather data. They are also used in other digital preservation risk assessment frameworks (e.g. Lawrence *et al*., 2000; U.S National Archives and Records Administration, 2023; The National Archives, 2023b), so represent a familiar method to many. This method is well suited to inexperienced assessors with little prior knowledge of the model, though it still requires domain-level digital preservation knowledge in order to understand and apply the questions to the given scenario.

The framework has four main sections: Scope and Context, Digital Content, Organisational Infrastructure, and Technological Infrastructure. The first section (1) addresses the initial contextualisation and scoping stage of the ISO 31000 risk management process, whilst the latter three sections (2 – 4) each explore a different risk originating entity from the risk source model and represent the identification stage of the risk assessment process.

The Scope and Context section encourages assessors to frame their assessment prior to embarking on the questions. The Type field indicates whether the focus is on a collection, a system, or something else. This is further explored in the Scope field. The more tightly the scenario is scoped, the more precisely the risks can be identified.

Field	**Description**
Assessment Name:	Logical descriptive name for the assessment
Assessment Type:	Collection level, system level, or other
Assessment Scope:	Description of scope
Assessment Trigger:	Reason for undertaking assessment
Date of Assessment:	Date
Names of Assessors	Assessor names
Risk Appetite Level:	The Risk Appetite for the scoped items, if relevant

Table 5: Scope and Context fields in the Method Two framework

Organisations with established risk appetites may choose to specify them here. The Risk Appetite concept establishes the threshold for acceptable levels of risk in a given context and in relation to planned objectives (Society for Risk Analysis, 2018;

Martens and Rittenberg, 2020; UK Government, 2023). Aven puts this more succinctly as an 'appetite for risky activities in pursuit of values' (2013, p. 465). It is a term frequently and interchangeably used with 'risk tolerance' though has particular prevalence in enterprise risk management contexts (Aven, 2013, p. 462) and thus is preferred here for purposes of alignment with enterprise risk management endeavours.[73] Risk Appetites are defined at levels appropriate to the needs of an organisation. For example, an organisation may allocate a moderate level appetite for corporate risk in general (representing an appetite for a 'reasonable' amount of risk in pursuit of corporate goals) but varying risk appetites for specific types of risks (e.g. financial, legal, or reputational) that may manifest within the organisation.

Well-considered appetite levels reflect organisational priorities and help ensure that risks across the organisation are assessed against a set of clear and consistent benchmarks. [74] They are used mainly to support decision-making and monitoring (Martens and Rittenberg, 2020, p. 10), so are thus most commonly applied in the evaluation stage of the Risk Assessment Process. Specification of risk appetites at the Risk Identification stage has nonetheless been found useful during testing of these methods, to provide a degree of guidance for the assessor on what may or may not be considered a significant risk. When everything is a potential risk, vast amounts of time can be spent on describing and evaluating risks that have little to no consequence, draining staff time and resources. Specification of risk appetite levels at this point helps screen out minor risks, focus attention on those risks that are likely to fall around or above appetite, and thus reduce the amount of time and effort spent on describing and assessing risks that are unlikely to pose a problem. It is a pragmatic approach to finding direction amongst a plethora of potential risks. It

[73] For a more detailed review of the risk appetite concept, its value, and related terms, see Aven (2013) and Purdy (2011).

[74] Berlinger and Váradi (2015, pp. 56 - 58) propose five methods through which to assess and determine risk appetite: choice dilemma, utility theory, heuristic judgements, objective measures, and subject assessment. These are based on Grable and Lytton's 1999 work, 'Financial Risk Tolerance Revisited: The Development of a Risk Assessment Instrument', in Financial Services Review 8. pp. 163–181.

is not however within the scope of this thesis, nor necessitated by the requirements, to explore and establish risk appetite levels as part of this method.

Sections 2 – 4 explore each class of risk source and its corresponding risk factors, as associated with a risk originating entity. Each class and factor is translated into a question, accompanied by an explanation to help clarify the purpose of the question and its relevance to target values. Questions are closed-ended and designed to elicit yes/no/unknown responses. They are consistently framed in such a way that 'yes' is an indication of a low risk response, whilst 'no' or 'unsure' indicates a likelihood of higher risk that needs investigation. For example:

Class	Factor	Question
Digital Content	Completeness	Do the files contain all of the intellectual content to which you expect to provide access?
Justification: *Rendered objects can sometimes display or use information held externally, linked from the content files. If this additional information is important but not available then the authenticity of the rendered objects can be affected.*		

Table 6: Example question and explanation from the Method Two framework

The template has space for answers and notes (not included in the above representation) in which to record meaningful data about both relevant instances of the class and the status of the risk factor.

There are forty-four questions in total. The amount of time taken to complete an assessment varies according to the scope of the assessment, the knowledge of the assessment team, and the ease with which reliable information is available to answer each question. The output from the process is a completed document with sufficient information to inform the next steps. Should the likely risks indicate the need for a full assessment process, method three may be initiated.

Method Three: A Risk Assessment

The third method builds on the identification process supported through methods one and two, and integrates it into a full assessment process to demonstrate how risks can be consistently described, analysed, and evaluated. It is a qualitative,

spreadsheet-based method that guides the user to first define the scope and context of a risk assessment, then produce, characterise and evaluate individual risk statements relevant to the scenario. The method represents a suggested form of risk assessment based around the entities in both the digital preservation risk source and risk source concept models, using standard risk management concepts and structures familiar to any organisation operating enterprise-level risk management programmes. Users are welcome to reconfigure the template for compatibility with their own organisational templates should they so desire, or use their own – the models are purposefully flexible enough to support this.

Spreadsheets offer a simple and standardised way to capture, present, analyse and evaluate risk statements as part of a risk assessment activity. They are relatively common tools for risk management, are simple to use, and spreadsheet software is widely – and freely - available in most organisations. The spreadsheet template designed for use with the model is purposefully simple with a layout that supports graphical comprehension of evaluations without requiring users to immediately understand all of the detail, thus facilitating sharing and communication of results.

The template has three main tabs, each holding different information. The first – Tab One - contains scoping information consistent with that already generated and described in association with method two, supplemented by additional information on the risk appetite. The second is the main assessment tab, where risks are identified, analysed, and scored. The third includes a copy of the risk matrix to support scoring and evaluation of risks.[75]

Tab Two uses a table presentation to develop consistently structured descriptions of individual risks. Fields in column headings each focus on a different structural aspects of the risks, with each individual risk represented in a single row. The fields are explored here in two sections for purposes of clarity: the first section focuses on

[75] Tables from the template are reproduced in Appendix C though these do not represent the functional elements of the spreadsheets. The functional spreadsheet template is available as a separate file associated with the How-To Guide, which represents one of the practical outputs of this research.

Risk Identification and Description, whilst the second supports Risk Analysis and Evaluation.

Risk Identification and Description

The outputs of Method One or Two should identify instances and factors that are a cause for concern in a given scenario. These can be further described using an abstract and generic approach that associates them with specific uncertainties and consequences. This section explains and models this process-based approach before presenting the spreadsheet designed to capture this information.

Instances and Risk Factors can be combined to identify a **Contextualised Risk Source**. In a Contextualised Risk Source, *the Instance is a specific manifestation of an Instance Type, whilst the Factor represents a variable property that may lead to a negative outcome*.

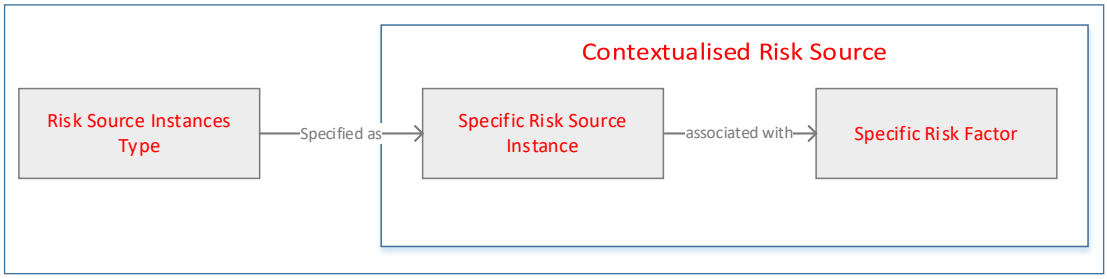

Figure 16: Components of a Contextualised Risk Source

A **Risk Description** is then formed by additionally specifying the **Uncertainty** associated with the Contextualised Risk Source. Uncertainty reflects *a lack of knowledge about a contextualised risk source, specifically in relation to its risk factor*. The highly contextualised nature of uncertainty makes it unsuitable for inclusion in the risk source model, as it can take different forms and is dependent on many variable factors including the context of the risk, the phrasing of the risk, and the interdependencies that an assessor may wish to represent. It is nonetheless important to stipulate the uncertainty within a description in order to more precisely define the nature of each individual risk. This association of a contextualised risk source with an uncertainty to produce a risk description is modelled in figure 17, below.

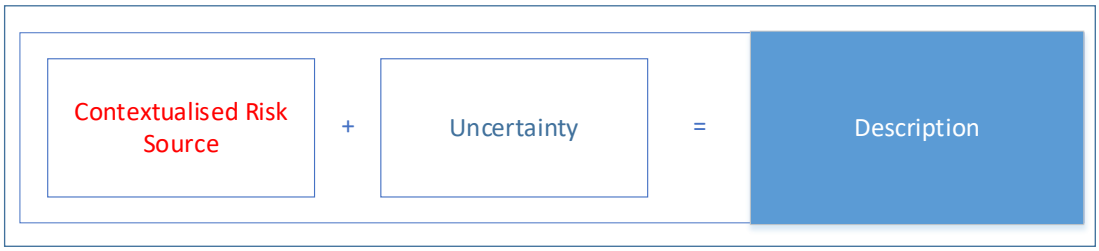

Figure 17: Components of a Risk Description

As risks relate to negative outcomes, the risk description must also be associated with a **Consequence**, reflecting *one or more undesirable potential outcomes*. Like Uncertainties, Consequences are also unsuitable for inclusion in the risk source model as they are dependent on many variable factors, not least the specific uncertainty associated with a contextualised risk source. Both consequences and uncertainties should be identified by the assessment team using their knowledge of both discipline and context.

This generic method for characterisation of individual risks can be represented as a procedural model. The relationships between the different concepts in this process are reflected in figure 18, below.

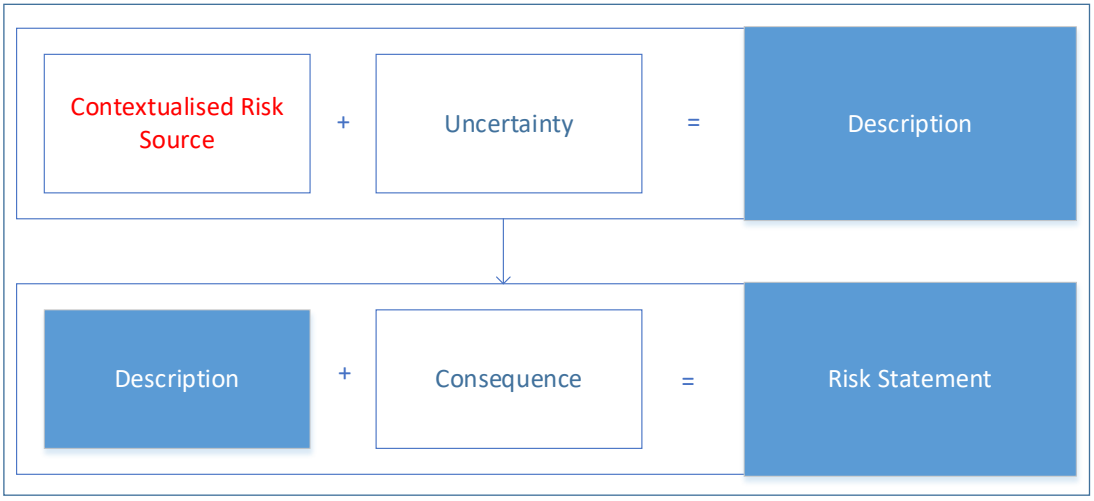

Figure 18: Risk Characterisation Model

The combination of a Risk Description with a Consequence results in a structured **Risk Statement**. This structure can be applied to any potential risk source, to generate a risk statement that identifies the contextualised risk source, the uncertainty associated with that source, and the potential negative consequences or outcomes. Application of this structure facilitates the consistent description and specification of risk at a different level of abstraction and granularity from that in

the risk source model. Moreover, framing the consequences specifically in relation to affected target values ensures that this structure also reflects the contextualised approach proposed by Heckmann *et al*. (2015), and upon which the initial conceptual definition of digital preservation in chapter four is based: an undesirable outcome, target values, and potential causes. Demonstrating such a mapping thus establishes a direct link between concept and characterisation, illustrating the application of risk science principles across each level of model to establish a consistent, cascading, and logical representation of digital preservation risk.

The fields in the spreadsheet are designed to consistently record and represent this structure, maintaining a link between the models by capturing information not just on the contextualised risk statement but also the classes of risk source and instance types to which they relate.

Field	**Description**
Class	Class of Risk Source for this risk
Instance Type	Type of Instance addressed in this risk
Instance	Contextualised Instance of Risk Source addressed in this risk
Factor	Uncertain factor addressed by this risk
Description	Contextualised Risk Source (Instance + Factor) + Uncertainty
Consequence	Potential negative outcome arising from the risk
Affected values	Target values potentially affected

Table 7: Risk Identification and Analysis Fields in Method Three

Class, Instance Type, and Factor fields all draw directly from the risk source model and are generic representations or associations of the risk source, whilst the remaining fields represent the contextualisation and characterisation of an individual, situational risk. Consequence and Affected Values feature in two different fields to encourage users to clearly specify affected target values in association with an undesirable outcome. This also operates as a mechanism through which to check that statements focus on digital preservation outcomes rather than more generic, situational problems.

The following series explores how different types of risk statements might be constructed, both with and without integrated references to target values:

For example, from Class (System Software) and Instance Type (Operating System), Instance (Windows 2008) and Factor (Support) might be associated with Uncertainty (Age).

The age of the Windows 2008 operating system means it may not be supported by vendors (including through extended support) after a specific date.

Consequently, security patches may not be available and may lead to security breaches. Cybersecurity accreditation requirements will not be met. Dependency management (especially between operating system and hardware, and operating system and applications) will become increasingly difficult and potentially impossible.

From Class (Processes and Workflows) and Instance Type (Functional), Instance (Replication) and Factor (Reliability) may be associated with Uncertainty (Completion).

The reliability of replication processes may be inconsistent and they may not always complete as expected.

Consequently, content files may not always be replicated to another location and we may have no way of checking without manual reconciliation reports.

From Class (People) and Instance Type (Departmental), Instance (IT) and Factor (Quantity) may be associated with Uncertainty (Completion of work requests).

The IT department may not have sufficient staff to complete prioritised work requests on top of their business-as-usual activities.

Consequently, business continuity may be affected. Content from new suppliers may not be added to the system and metadata schema upgrades may not be implemented. We may therefore need to maintain backwards compatibility (at an expense) and content may be at increased risk because of maintenance delays.

From Class (Content Files) and Instance Type (Item), Instance (Mobile app) and Factor (encoding) might be associated with uncertainty (validity).

The encoding used by the app might not be valid.

Consequently, it may not be possible to open and deploy the app, leading to loss of access to the object and its intellectual content.

From Class (Metadata) and Instance Type (Technical), Instance (PREMIS file) and Factor (Accuracy) might be associated with Uncertainty (technical characteristics).
The technical metadata in the PREMIS file may not accurately represent the technical composition of an item.
Consequently, we may not know the technical dependencies of the item so are unable to ensure they are appropriately supported in a way that provides access to authentic representations of intellectual content.

There is a degree of variation in the descriptions, for example, the factor or an inversion of the factor (as seen in the statement above for a mobile app) may be sufficient to indicate the nature of the uncertainty, whilst in other descriptions the uncertainty can take a different form. This is a valid approach so long as any subsequent statement is clear. This formula-based, reasoned approach ensures consistency in the structure of statements and therefore a consistent characterisation of individual risks, regardless of the type of risk described or its consequences. This facilitates comparability of risks and also makes it easier to identify dependencies between related risks that may affect subsequent risk treatment options.

It may, in some cases, be appropriate to expand the scope of the risk assessment so that it reflects wider consequences than those primarily associated with preservation of content. For example, degradation in the condition of physical hardware may lead to decreased efficiency, whilst network security problems may also cause reputational issues. This is a valid extension of the statements that manages digital preservation concerns alongside those of the wider environment and represents the wider applicability of the approach. Such an extension is not explored here however, as it is out of scope for this research.

Analysis and Evaluation

Analysis and evaluation of risk statements is undertaken using a qualitative risk impact/likelihood matrix in conjunction with one or more stipulated risk appetites. Risk matrices and appetites are already used at the British Library for assessment and management of risks in other business areas so are well suited for this research, given its underlying motivation. From a wider perspective, they represent widely-used and complimentary but relatively simple approaches for scoring and evaluating risks. Risk matrices provide a framework with which to measure risks, whilst risk appetites provide a benchmark against which to evaluate them.

Risk matrices require assessors to identify a) the impact of a risk manifesting and b) the likelihood that it will occur, using descriptions and values from a predetermined scale. There are different methods through which to generate initial scores for likelihood and consequence, both qualitative and quantitative. Quantitative scores are generally statistically derived, whilst the qualitative scores are manually allocated based on the knowledge and experience of the assessors and the detail provided in the risk statement. The spreadsheet uses a qualitative scoring approach and is straightforward to learn though inevitably subjective. The subjectivity can be countered to a degree by, for example, ensuring assessors are experienced and knowledgeable about their field, integrating a process of peer review, and incorporating evidence-based judgements where available. The single probability/impact scores subsequently map onto a matrix, with their product calculated to represent a single overall ‘rating’ for each characterised risk statement.

The risk matrix approach is widely implemented in many fields and disciplines, from project and programme management to enterprise-level risk management systems. The use of risk matrices is not without its detractors, due mainly to the potential for mis-scoring that arises from, for example, the mathematical or logical limitations of their structure (Cox, 2008; Ball and Watt, 2013), their sheer simplicity with qualitative approaches (Emblemsvåg and Kjølstad, 2006), use of ordinal scales (Krisper, 2021), or just poorly characterised risks (Louis, 2008; Aven and Thekdi, 2022, pp. 46-48). A risk matrix nonetheless remains a valid technique for measuring

risk based on likelihood and consequence (ISO, 2019) so long as its limitations are understood and deemed acceptable – the selection of an appropriate tool is, after all, a contextual decision. Guidance is available to support the solid and thorough construction of risk matrices that addresses some of these issues (Cox, 2008; Baybutt, 2018) and this has been considered in the construction of the matrix used in this method.

The spreadsheet template has an additional five fields to support analysis and evaluation of each risk expressed in the risk statements generated in the previous section:

Field	Description
Status	Brief assessment of risk status
Impact	A score representing the impact of this risk should it manifest
Likelihood	A score representing the likelihood of this risk manifestation
Overall Score	A calculated (Likelihood * Impact) score
Mitigation Constraints	Notes on any constraints in place that may affect mitigation options

Table 8: Risk Evaluation Fields in Method Three

The Status field functions as a short, textual description of the status of the risk as described in the risk statement, prior to generation of a score in the Likelihood field. This can be a simple, summative answer, such as 'out of support', 'unreliable', 'insufficient', or 'known', as relevant to the risk statement. This status field was found to be helpful during formative testing of the spreadsheet method, to indicate to fellow reviewers the reason why a risk was given a certain likelihood scoring. The Likelihood score represents the chance of something happening, whilst the Impact score represents the effect of the consequences on the organisation's objectives, in line with the risk appetite stipulated in the first tab. The total score is then calculated according to the simple formula of (Impact x Likelihood).

The risk matrix in this template uses scales of 1 – 5 and purposefully neutrally worded scales to minimise emotional influence on the qualitative process and encourage rational decision-making (Purdy, 2011; Jensen and Hansen, 2020). For example, it avoids use of the terms 'catastrophic' or 'devastating' for the highest

impact, and uses instead ‘acute’. The wording associated with each level in the corresponding risk appetite scale is similarly neutral, avoiding terms like ‘hungry’ or ‘open’ in favour of more precise and objective terms. The wording for both the matrix and the appetite is thus divergent from those used in standard British Library risk matrix scales and appetite descriptions, though their objective meanings are compatible and the actual scores and scoring ranges are the same.

Impact	Acute (5)	5	10	15	20	25
	Major (4)	4	8	12	16	20
	Moderate (3)	3	6	9	12	15
	Minor (2)	2	4	6	8	10
	Almost None (1)	1	2	3	4	5
		Rare (1)	Unlikely (2)	Possible (3)	Likely (4)	Almost Certain (5)
		Likelihood				

Table 9: Risk Matrix used in the template

The result is interpreted against the following key:

17 to 25	Very High	Action urgent
11 to 16	High	Action required
6 to 10	Moderate	Action preferred
1 to 5	Low	Acceptable

Table 10: Risk Evaluation levels used in the template

The Matrix in the Template represents a low risk appetite, with any risk scored 6 or more indicative of preferred action. It satisfies the risk matrix design principles of betweenness, weak consistency, and consistent colouring (Cox, 2008) and thus reflects good practice in the rigour of its structure and logic. Betweenness requires that at least one intermediate cell passes between a green cell at the lower left of the matrix and a red cell at the upper right. Weak consistency requires that a risk matrix distinguish between the highest and lowest risks. Consistent colouring meanwhile requires that scores are represented by a consistent colour regardless of their location on the matrix – so for example all cells with a value of 10 are yellow, and a cell with a value lower than 10 is not orange or red. These principles are devised primarily for matrices supporting quantitative risk assessment techniques

but nonetheless have value in guiding risk matrix design in general, simply by virtue of their core logic. The appetite represented in the matrix can be modified if needed. This could be done in several ways, for example by reducing the number of overall rating ranges from 4 to 3 and using a scale of 1 – 10 as 'low', 11 – 16 as ''moderate', and 17 or more as 'high'. Scores associated with each axis could also be changed, for example by weighting impact greater than likelihood. Changes such as these however need careful consideration to ensure that the implications are understood, that they are consistent across the whole of the matrix, and that the principles of good matrix design are still applied. It is therefore recommended that if such changes are made, they are thoroughly considered and tested, and the reasoning fully documented.

The final field is an opportunity to capture any information relating to mitigation constraints. For example, in the case of support no longer being available for legacy software, a straightforward migration to a new version of that software may be ruled out by major differences in functionality that affect how it interacts with other software or files. Similarly, mitigation actions to address insufficient numbers of staff with knowledge to deal with legacy systems might be constrained by the limited availability of legacy knowledge outside of the organisation. This field is therefore included to assist in understanding options for subsequent risk treatments.

The template does not include a field for mitigation or treatment actions on a risk-by-risk basis. Our experience at the Library is that many of the more meaningful risks in a scenario-based assessment are related. Mitigation on a risk-by-risk basis therefore leads to unnecessary overlap and uncertainties about responsibilities and interdependencies. A more appropriate approach is a risk response that considers all relevant risks together in a single scenario and proposes a holistic approach to mitigation.

Risk Mitigation and Preservation Planning

In digital preservation, the treatment of risks is often associated with the process of preservation planning. The concept of preservation planning originates from the

ISO 14721 reference model for an Open Archival Information System (OAIS) (ISO, 2012a), where preservation planning is one of six functional entities that collectively represent the management functions of an OAIS. It generates the recommendations and preservation plans to ensure information remains accessible, understandable, and usable, 'even if the original computing environment becomes obsolete' (p. 4-2). To this aim, risk management is considered a 'suitable methodology' for balancing immediate needs against long term imperatives and providing metrics to support decision making (p. 4-15).

The process through which to address risk in ISO 14721 is the production of 'risk analysis reports', an activity that occurs within the 'Develop Preservation Strategies and Standards' function of the preservation planning entity. This activity 'addresses expected risks' and proposes potential mitigations in line with existing or updated policies and procedures (p. 4-15). The resulting risk analysis reports represent a blend of analysis, evaluation, and treatment stages in the ISO 31000 risk management process. Risk management is thus an important preservation planning activity for the OAIS to achieve its goals. The process for initial risk identification however is unclear and the reference model describes predominantly technological risks, in particular obsolescence (pp. 5-1 - 5-15). The broader concept and nature of digital preservation risk is not clearly represented.

ISO 14721 uses the term 'preservation planning' extensively though does not explore the concept – or expected content - of a 'preservation plan' (Pennock, 2020). Production of migration plans is however recommended to avoid loss of access due to technology obsolescence (ISO, 2012a, p. 5-2), informed in part by risk analysis reports. Preservation plan templates developed within the community to date are often migration-focused and, consistent with the ISO 14721 perspective on risk, have a predominant focus on format and obsolescence-based risks (Becker *et al*., 2009; Graf, Gordea and Ryan, 2014; Johnston, 2018; Friedrich, 2019; Skødt, 2022).[76] This is broadly in line with the purpose of the preservation planning

[76] A small number of other institutions are known to use the term preservation plan differently. The World Bank Group for example uses it to represent a document containing comprehensive information about a digital collection including appraisal and selection criteria, metadata profiles, transfer and ingest requirements, storage locations, arrangement and description, and formats

functional entity as obsolescence-avoidance, though it continues to reflect a level of disconnect between perceptions of digital preservation risk when ISO 14721 was first published and contemporary thinking about the broader and more expansive nature of digital preservation risk as an organisational and holistic concern. As of summer 2023, a new version of ISO 14721 is in production that has potential to address these issues though the current draft does not yet do so.

If we accept both that the range of risks in digital preservation is broad, and that the concept of preservation planning includes risk assessment and selection of viable risk treatments, then a preservation planning risk response must take into account this wider set of risks than just the technological. A focus on technological and obsolescence-based risks alone, whilst in keeping with the definition of preservation planning in ISO 14721, is not enough to ensure that information will remain accessible, understandable, and usable over time. The risk assessment model and methods presented in this thesis provide the foundation from which to develop that wider perspective, generating a holistic overview of risks in a given scenario that a preservation planning exercise should subsequently support. The work required to take this forwards is beyond the scope of this thesis but options to do so are discussed briefly in chapter eight.

Implementation: A Case Study

The methods described in this chapter demonstrate the utility of the models for exploring the complex nature of digital preservation risk, culminating in the production of risk assessments that can support the foundations of a comprehensive preservation planning risk response. This section further substantiates that demonstration with a short case study to illustrate how both model and methods have since been used at the British Library to produce a series of risk assessments on real risk scenarios.

(Kramer-Smyth, Gkremo and Thompson, 2023), whilst the Libnova digital preservation software company uses it to describe the parameters of an ingest profile for content being ingested into their preservation repository solution, Libsafe. In general however, most community usage of the term preservation plan, including the derivation ‘preservation action plan’, relates to processes specifically for technological or format-based object interventions.

Twelve risk assessments have so far been produced, representing a mix of system-level and collection-level scenarios. System-level scenarios address collection-holding systems that use different technologies, with different origins, and of different ages. Collection-level scenarios address specific types of digital content or collections, ranging from legacy content on handheld media to large-scale digitised material, born digital personal archives, datasets, and single complex objects. Pre-processing and post-processing scenarios were included for comparative purposes, as were a mix of 'ad hoc' status-ascertaining assessments and trigger-based assessments.

Method One was used to initiate the exploration and identification of potential risks in each scenario. This represented the initial process through which to consider the potential for risks and gather data from relevant parties, with discussions led by the author. This method was appropriate to the assessor's level of knowledge about not only the risk source model and the methods, but also the collections and the organisational/technological context of the Library. The digital preservation risk source model was printed out onto A3 paper to facilitate discussion during face-to-face meetings with other members of assessment teams. Attendees generally found it easier to comprehend the model when they were able to easily visualise it all on a single 'page', and whilst a digital version was shared onscreen during virtual meetings, laptop-sized screens were found to be a limiting factor for the visualisation. All meetings began with a short introduction to the risk source model that explained how it was developed, including the definition of digital preservation risk and target values, as well as the various conceptual entities. Colleagues were then invited to explore any concerns they may have in relation to the stated scenarios and the information in the model. Notes were taken in line with the key stages of the method outlined earlier in this chapter, then consolidated and shared with attendees afterwards. There were no occasions when a potential risk was identified that did not map into the model, and on more than one occasion it was possible to map a concern to more than one area in the model. This helped to demonstrate the relationships between different risks and different

risk sources, and confirmed the value of a conversational, 'team' approach in highlighting these different perspectives on the same risk issues.

Whilst this engagement was not part of a summative evaluation process, colleagues were invited to provide feedback on the model and the method of exploration. This was highly positive, with several interviewees noting the value of a 'big picture' perspective and being able to 'see how all the different parts fit together'. A peer from the conservation department who actively works on conservation risk assessments stated 'I really like this approach, it takes that broader view of everything we might want to consider', whilst a colleague from the Technology area shared that 'this looks really useful - I can see how I might use this approach for my own risk assessments going forwards'. Colleagues in Architecture and Scholarship made similar comments about the usefulness of the model and in particular its visual representation. One noted how much he liked that different 'pathways' in the model could surface the same type of risk but from different perspectives, particularly when thinking about reasons for potential loss of content.[77]

Spreadsheets were prepared for each scenario using method three, representing many different instances of risk sources and factors from across all three risk originating entities. These were produced in the first instance by the author, then circulated with interviewees for discussion and refinement. They clearly characterised individual risks and effectively highlighted areas of concern, providing a mechanism through which to not only identify and describe the risks of a given scenario but also demonstrate the viability of the method for communicating and characterising different types of risks in a single assessment. The visual form of the spreadsheet was found extremely helpful. One colleague stated straightaway 'I understand what this is trying to illustrate – it's a commonality of presentation', whilst another observed that 'to actually articulate all the risks to [… people …] feels like it's difficult. How you actually express and surface those risks so that people

[77] In addition to this direct feedback received during interviews to support case study application, senior colleagues in corporate information management and operations management have also commented positively on both the methodology and the risk science approach that underpins the model.

understand them - this would help with that'. The use of a risk matrix and risk appetite levels was considered to align particularly well with standard corporate risk management practices and was positively received. These experiences and this feedback helps demonstrate the usability of the final outputs, as well as the underlying methodology and models that inform them.

This series of spreadsheets not only demonstrates the utility and efficacy of the artefacts produced during the research but also represents an implementation or instantiation of the research. Instantiations are the fourth artefact type proposed by March and Smith, which demonstrate 'the feasibility and effectiveness of the models and methods' (1995, p. 258). This application of the research therefore functions as the final demonstration of both the utility and the efficacy of the research outputs to answer the research question.

Conclusions

The massive scale and diversity of operations and digital collection content at the Library presents an ideal microcosm in which to demonstrate the utility and efficacy of the risk models produced for this research. This chapter has outlined a series of methods through which the models can be used to explore the nature and complexity of digital preservation risk in a number of different real-world scenarios. These methods, in turn, demonstrate how a risk assessment process can be used to thoroughly and consistently represent different types of risk, and form the foundations for a flexible yet comprehensive preservation planning risk response. This practical application confirms the utility and efficacy of the artefacts for answering the research question whilst also enhancing the British Library's abilities to identify, communicate, and consistently assess different types of digital preservation risk.

This chapter has also reviewed the association between the ISO 14721 'preservation planning' entity (ISO, 2012a) and risk assessment. It has demonstrated how the current approach to risk management within ISO 14721 is focused mainly around obsolescence avoidance and the production of migration plans, limiting the relevance of the preservation planning entity to primarily

technological risks. This reflects a degree of disconnect between legacy thinking about digital preservation risk from when the standard was first devised, and contemporary understandings of digital preservation risk as outlined in this thesis. Expanding the scope of risk analysis in ISO 14721 to support this wider perspective would result in a broader – and arguably more useful - application of the preservation planning function as a risk response. Moreover, alignment of these processes with the ISO 31000 risk management standard (ISO, 2018) enables organisational integration of digital preservation risk management practice with wider enterprise risk management endeavours. Application of the model and the methods to support expansion of the preservation planning concept in this way has the potential to make a significant contribution and advancement in both the state of the art and the state of the practice for managing digital preservation risks.

Chapter Seven: Evaluation

Introduction

The previous chapter demonstrated the efficacy and utility of the models presented in chapter five through application of specially devised methods and a case study of methods in practice. Evaluation takes this demonstration a step further by validating the research outputs and artefacts against specific criteria to determine whether they meet the requirements. Evaluation is one of the two main types of research activities proposed for design science research by March and Smith (1995), and the fifth stage of the design science research methodology devised by Peffers *et al*. (2006), as used in this research. It is a crucial component of design science research that assesses whether the solution works well enough to satisfactorily solve the problem.

Evaluation in design science research occurs during the design and refine process, as well as at the end of the process. These two types of evaluations – also known as formative (or *ex ante*) and summative (or *ex post*) evaluations - represent different functional purposes. The former functions as a process to improve the product, and the latter to assess the final form of the product (Venable, Pries-Heje and Baskerville, 2016, p. 78 - 79). Informal, formative evaluations of the solution occurred as part of the design process and helped shape the utility of the artefacts to address the research question. Structural consistency, practical usability, and visual clarity all emerged as significant and valuable features of the artefacts for this purpose. This chapter mainly represents the additional, summative evaluation undertaken to assess the solution at the end of the research process.

Dreschler and Hevner propose two perspectives to a summative evaluation: Fitness for Use, and Fitness for Evolution (2022, pp. 12 - 13). Fitness for Use focuses on the ability of the artefact to deliver a solution to the problem given the current goals and context, whilst Fitness for Evolution addresses the adaptability of the solution to respond to changes in the problem space over time. This latter perspective is consistent in particular with the criterion for robustness, as described in the

evaluation of the models, whilst the former is focused on utility, which is the overall goal of artefacts in a design science framework and broadly addressed by most of the other criteria in this chapter. Both Fitness for Use and Fitness for Evolution are thus considered to be addressed in the evaluation of the artefacts and the solution as below.

The evaluation approach followed in this chapter is broadly consistent with that suggested in the Framework for Evaluation of Design Science (FEDS) that focuses on evaluation goals, strategies, properties, and processes (Venable, Pries-Heje and Baskerville, 2016). The overall goal of the evaluation is to determine how well the research answers the research question. This chapter takes a three-tiered approach to determining this, considering not only the requirements for the solution established in chapter four, but also the criteria for design science constructs, models, and methods as defined by March and Smith (1995), and an evaluation of methodology itself against the design science research guidelines recommended by Hevner *et al*. (2004). Each addresses a different aspect of the research though there is nonetheless a degree of repetition across the three frameworks. This is minimised where possible or otherwise acknowledged in the text.

Evaluation of the Solution

The Requirements in chapter four use the generic term 'solution' as a way to describe the knowledge and capability required to establish an answer to the research question framed in chapter one, on how the nature and complexity of digital preservation risk can be more thoroughly and consistently represented than in works to date to support the foundations for a more flexible and comprehensive preservation planning risk response. Each of the requirements is reproduced below with an explanation of how the requirement has been met. The foundation design science research requirement for utility of the solution underpins each individual requirement.

The Solution must provide clarity into the relationship between the concept of risk and the practical manifestation of risk, so that the difference between the two is understood in intellectual terms: The solution draws upon risk science to identify an

approach through which the concept of risk and the practical manifestation or characterisation of risk can be distinguished, as advocated by Ylönen and Aven (2023). Using a definition of digital preservation that incorporates many of the standard themes and concepts found in the literature, it establishes a meaningful conceptual definition of digital preservation risk using a risk conceptualisation structure from Heckmann *et al*. (2015). It uses this as the basis from which to generate an abstract model of risk sources, populated for the digital preservation risk domain, which can subsequently be used to identify situational and practical manifestations of risks in a given institutional setting. The solution further develops this with the introduction of a risk characterisation model that functions as a generic method through which to generate contextualised and individual statements of risk. The abstract risk source model operates as an enabling and logical bridge between the two, providing a structured and rational process through which to develop from the concept of digital preservation risk to a characterised, practical risk statement.

The Solution must define all key terms used, so that it is a comprehensive representation of the requisite vocabulary: All key terms are defined in the Glossary, included in this document as Appendix A. The Glossary includes all conceptual entities and high level definitions, though excludes terms that require further contextualisation in order to characterise and describe individual manifestations of risks. Terms are also further explored and discussed in the body of this thesis.

The Solution must use precise language in its definitions of terms, so that the potential for ambiguity and misinterpretation is minimised: All terms have been defined as precisely as possible whilst remaining true to the core objective of utility. It is nonetheless possible that terms may be misinterpreted due to variations in local contexts and underlying knowledge bases. Discussions with British Library staff during production of the case study assessments were positive about the clarity of terminology, though only through wider community release and review can this be thoroughly tested.

The Solution must clearly identify the main elements of a digital preservation risk ecosystem, so that it can be used to explore risks associated with different

organisational and technological aspects of digital preservation: The solution provides this clarity in the digital preservation risk source model, developed from the key conceptual entities defined in the digital preservation risk context model and the risk source concept model. In particular, the risk source concept model establishes the key concepts to consider in an exploration of a digital preservation risk ecosystem, namely risk originating entities, classes of risk sources, risk factors, and instances/instance types. These form the structural basis of the digital preservation risk source model, which functions as a reference model for exploring different types of digital preservation risks. The methods developed in this research effectively illustrate how the digital preservation risk source model can be used to explore risks associated with different organisational and technological aspects of digital preservation, substantiated through a description of use at the British Library in a real world risk assessment exercise.

The Solution must not conflate or relate risks with mitigations: The solution addresses this requirement through its prescriptive conceptual combination of risk source classes and risk factors. These are stipulated in the digital preservation risk source model, which was designed to identify abstract, generic sources of risk independent of any given implementation or requirement. Construction of the Risk Identification Framework functioned as a quality check of this principle through the translation of factors and sources into questions, as the nature of the question often indicated whether it was a true risk or a risk masquerading as an inverted requirement that otherwise functioned as a mitigation for an anticipated risk. Method three attempts to continue this principle by encouraging re-use of the source/factor relationship model to specify contextualised risk sources and a structure for further elaboration in the form of a risk statement. Assessors may nonetheless still choose to represent instances and factors in terms of a contextualised and pre-emptive mitigation-based requirement. This might be considered a feature of an implementation of the solution rather than the solution itself, though further guidance on this could assist assessors in avoiding such an outcome.

The Solution must be demonstrated by one or more methods but remain flexible and not predetermine a particular approach: The solution is represented by abstract models and generic methods that can be implemented in different ways. Between models and methods, the models in this solution are the most transferable. The risk source model in particular, which is core to answering the question of how to consistently and thoroughly represent the nature and complexity of digital preservation risk, is an abstract reference model. As such, it is characterised by its universal relevance to the domain and does not presuppose any particular implementation. The concepts, entities and relationships in the model should therefore be re-usable with different methods to support risk assessment and subsequent preservation planning risk responses. The true test of this however will occur once the model has been released more widely. The methods themselves do not extend to preservation planning, though a review of the relationship between risk assessment and preservation planning nonetheless outlines how they can form the foundations for a more comprehensive preservation planning risk response.

The Solution must not assign importance or severity levels to objects, as this is contextual and down to individual users to assign: The solution does not assign importance or severity levels to entities in the risk source model. A mechanism is provided in method three for assessing importance and severity of characterised risk statements, though this is not pre-weighted in any way so does not inherently assign greater severity to any one risk over another. Method three uses the risk appetite concept to inform overall scoring, though this is applied to a scenario in general rather than individual risks and entities. Moreover, methods are suggested for demonstration purposes rather than prescriptive implementations of the model, so users are free to use the model in association with weighted methods if they so wish.

Evaluation of the Artefacts

As the nature of the artefacts constructed for this thesis is defined primarily in terms of the artefact types proposed by March and Smith (1995), it is also appropriate to evaluate them on that basis. The DSR evaluation criteria suggested by March and Smith are applied here to the practical, usable outputs produced for

this research, namely the construct, model and methods. Reflections on the underlying theory are integrated into this structure as theory is expressed primarily within these outputs rather than as a distinct artefact in its own right. The implementation is also not evaluated as a distinct artefact, as within the context of the research question explored in this thesis it primarily demonstrates the utility of the previous artefacts in relation to problem/solution space rather than the generation of new knowledge.

March and Smith's criteria are oriented around an information technology series of artefacts, though they are not clearly defined in any substantive detail. A wider review of design science literature found that whilst many of the terms are used in other design science research evaluations, they are often either undefined, or inconsistently defined. This is ironic, given the propensity of the field to extol the importance of rigorous evaluation in a design science research process. The loose definitions of each criterion nonetheless allows a degree of interpretation that supports their use with this research. The interpretations below thus reflect the author's understanding of each criterion in relation to the artefacts produced for this research, the research question, and the wider reading in general.

There is some overlap between these criteria and the requirements for the solution discussed in the previous section, as well as between the DSR criteria themselves. This could have been remedied by a thorough and pre-emptive stipulation of all criteria to be used in the evaluation, though this was not undertaken as it is beyond the scope of the research to perform a normative evaluation of the standard DSR evaluation criteria. There is therefore a degree of repetition in this section. This is retained to demonstrate that the criteria have been thoroughly considered, though also represents a lesson learned regarding the use of different sources of evaluation criteria and the reusability of the DSR criteria in a research project with multiple related artefact types.

The Construct

Constructs form the vocabulary used to describe a domain and its problems in relation to the proposed solution. The primary form of the construct produced for

this research is a glossary. March and Smith's criteria for evaluation of the construct are: completeness, simplicity, elegance, ease of use, and understandability.

Completeness of the construct refers to the coverage of the construct in relation to the problem statement. Does it include all of the necessary entries to comprehensively represent the domain in terms of the problem/solution space? Is anything missing that might lead to misunderstandings?

The core terms used to describe the domain and its problems in relation to the proposed solution are found mainly in the models and methods, including foundation definitions of digital preservation and digital preservation risk. These are included in the glossary found in Appendix A, though terms are also discussed extensively in the text. The glossary has been mapped against models and methods to confirm this. Some types of terms are purposefully excluded from the construct in order to facilitate contextualised interpretations of characterised risk. This is appropriate given the abstract nature of the models, agnostic of any particular implementation. Such terms relate primarily to specific risk factors and specific instance types, though general definitions of the term 'risk factor' and 'instance type' are provided whilst the text in the thesis (particularly chapter five) provides guidance on how each instance type and factor in the model might be interpreted.

Simplicity of the construct relates to the language used. Is the language clear and not unnecessarily verbose? Is the terminology reasonably familiar to the community or communities that will use it?

The language used to define terms in the glossary is as simple as could be produced whilst still resulting in clear and concise definitions. The terminology used should be reasonably familiar to the community, though the interdisciplinary nature of the topic means that practitioners of certain backgrounds may be more familiar with some terms than others.

Elegance of the construct is interpreted in relation to its clarity, particularly in relation to its explanatory capacity.[78] Are the definitions sufficiently precise and clear to avoid ambiguity and misinterpretation?

All definitions have been carefully phrased for clarity and to minimise the potential for ambiguity. Terms have been precisely defined and in cases where there is likely to be uncertainty, clearly differentiated. There is nonetheless the potential for misinterpretation by users with different foundation knowledge bases or contextual differences in language usage. This therefore needs to be monitored and reviewed as necessary when the research is published externally.

Ease of use of the construct considers simply how easy it is to use. Usability in this context relates to how easy it is to find definitions of terms used. In short, is it easy to find the terms defined?

The Glossary is presented in alphabetical order so that terms are easy to find, regardless of the entity to which they relate. Where terms can have more than one role or function, this is also acknowledged in the glossary. However, as not all terms are defined, this may pose an issue when users seek definitions that have been purposefully excluded. Most terms are also discussed in the text of the thesis as part of the explanatory descriptions, emboldened or italicised at key points.

Understandability of the construct relates to elegance, clarity, suitability for the audience, and simplicity. As these aspects have each already been covered in relation to the criteria above, they are not repeated here.

The Model(s)

March and Smith's open-ended description of a model represents a set of constructs and their relationships, though the construct is already evaluated above. The five criteria proposed for models are: fidelity with real world phenomena, completeness, level of detail, robustness, and internal consistency.[79] In order to

[78] For more on the concept of elegance as a characteristic of science, see 'Elegant Science' (Casadevall and Fang, 2018).

[79] These criteria were also used by McGovern (2009) to evaluate the model in their thesis on technology responsiveness for digital preservation.

maintain a distinction in this section between models and methods, this section focuses primarily on the abstract models produced in chapter five. The procedural characterisation model - which represents a generic method, as explored in chapter six - is evaluated in the section that follows.

Fidelity of the model with real phenomena represents the faithfulness of the representation in the models to the real world. Do the models accurately and recognisably portray and describe the important elements of the problem/solution space? Such fidelity is essential for the models to serve as a practical and real world solution, particularly in design science where the goal is to solve real world problems.

The faithfulness of the models to the real world is clearly established through the process that underpins their construction and their subsequent utilisation in real world scenarios. The definition of digital preservation provides a starting point to demonstrate this fidelity, representing key conceptual and practical aspects of the domain. Utilisation of these aspects in the definition of digital preservation risk ensures they are sufficiently prominent in the resulting context model, which in turn identifies the main conceptual elements of the risk source model. These are populated using a combination of research and pre-existing knowledge generated through practical experience. The fidelity of the populated risk source model with the real world context is demonstrated through its application to the British Library context, and its wider application and fidelity is substantiated through its use of terms and concepts also widely found elsewhere in the domain literature. The production methodology is thus considered to have generated a justifiable and faithful abstract representation of the real world context.

Completeness of the model relates to the coverage of the models in relation to the problem statement. Do they include all of the necessary entities in order to help answer the research question? Are they comprehensive?

The models were designed using a structured process intended to generate a thorough, logical, and complete yet abstract overview of digital preservation risk. They are based on a clear and reasoned definition of digital preservation, from

which was developed a contextualised definition of digital preservation risk. This structured approach establishes and justifies the conceptual coverage of the models, as the relevance of each part can be traced back through to the definitions provided. Explication of the risk source concept in particular ensures the models include sufficient detail to identify different elements of risk, and that they are consistently represented. The transparency and logic of this process demonstrates how the models can collectively be considered a comprehensive and thorough representation of the complexity of digital preservation risk. In addition to this, the case study unearthed no cases where risks were suggested that could not be inferred from the main risk source model. The model therefore satisfies this criteria in terms of breadth of coverage.

Level of Detail in the model represents an assessment of whether they have an appropriate amount of detail for their intended use. Do they represent the problem area to a sufficient degree of granularity to help answer the research question?

The detail is expressed not just in the risk source model but also in the cascading relationships between models. The amount of detail and depth is appropriate for an abstract reference model, representing different concepts to a level of granularity that, with guidance from the methods, enables production of precise and situational risk statements. It is significantly more detailed than other risk models produced within the field and has sufficient granularity to address, in particular, the first part of the research question around how to thoroughly and consistently represent the nature and complexity of digital preservation risk.

Robustness of the model explores how well they perform across various different scenarios and whether the utility of the models changes in a meaningful way when applied to different underlying data. Robustness should relate to scope, as it is not feasible to expect models to support scenarios that they were not designed for. From this perspective, robustness represents the extensibility and reusability of models in line with their purpose and expected scope.

The robustness and reusability of the models is demonstrated by their relevance to different scenarios and data, so long as both are within the scope the models were

designed to support. The robustness and reusability of the risk source model is evidenced through its case study application at the British Library, a sufficiently large and diverse organisation that it can reasonably be considered to function as a microcosm of digital collecting. Extensibility of this model is also possible should institutions identify value in doing so. Though this might imply a degree of incompleteness, changes to the environment are inevitable over time. This should not therefore indicate uncertainty about the completeness of the model, but acceptance of the observation that we are always 'shooting at a moving target' (Hofman, 1999) and will therefore eventually need to adapt the model in order to maintain its fidelity with real world scenarios. Such adaptation can be achieved by working through the same process through which the model was developed and which is documented in this thesis. That said, it is only through wider testing that the robustness of all models can be fully evaluated.

Internal consistency of the model ensures they have no inherent conflicts or contradictions. Are entities and terms consistently defined and presented in relation to one another?

Related entities are represented with a consistent form and structure within individual models. For example, in the digital preservation risk source model, all Risk Originating Entities contain a series of individuated and consistently structured Risk Source Classes. These are each associated with a series of Instance Types that share the same set of Risk Factors associated with their parent Class. They are presented consistently across the model, using standard UML nomenclature. Where risk source classes share other relationships, these are noted in the model and explained in the text. The logical progression and consistency between definitions and models is similarly explained in the text. Case study application identified no inconsistencies in terms of conceptual descriptions or relationships, either within or across models. Consistency is also applied across different models where possible, for example through consistent colouration.

The Method(s)

Methods are the steps or sets of tasks through which a solution is performed and demonstrated. March and Smith's description of a method is primarily computational, though in this thesis, the methods primarily represent manual steps. In combination with the models, these demonstrate how the research question can be answered. They are evaluated here to consider their utility and how they might be improved going forwards. The four criteria proposed by March and Smith for evaluation of methods are operationality, efficiency, generality, and ease of use.

Operationality of the method relates here to the ability of humans to effectively use it. This interpretation is consistent with March and Smith's example of the term when applied to a non-algorithmic method. It is therefore combined here *with ease of use*.

All methods are designed to be re-usable by external users. How they are used can however vary, which might impact on their perceived effectiveness. Methods one and three have both been effectively used by the author to construct thorough, considered risk assessments. The author's familiarity with the methods and underlying models made them straightforward and easy to use. Method two has not yet been used in practice, though is considered to represent a guided and more structured approach to using the risk source model than method one. As a translated representation of the model it facilitates interpretation and application in a practical setting and thus should make it easier for new users to use the model than without, though this will inevitably vary between different users and has yet to be put to the test.

Efficiency of the method considers the necessity of each step and how streamlined it is. Are there redundant steps in the methods? Could they be improved to reduce the amount of time and effort required to use them?

The purpose of the methods in this research was primarily to demonstrate utility of the solution rather than optimise methods for implementation and efficiency. Efficiency is relative to the amount of effort input to an activity and its resulting

output. The depth of exploration supported by method two indicates it is less efficient than method one, though familiarity with the underlying risk source model is likely to increase efficiency over time. There are no obviously redundant steps in the methods, though they could inevitably be improved with further work and with feedback from the wider community. There is a time-consuming element to the overall process that may be considered to represent inefficiency, although it is not wholly unavoidable given the nature and complexity of the subject area.

Generality of the method relates to its broader relevance to the community of potential users and external validity. Is the method relevant to external users? Is it reusable by different organisations and does it support different digital preservation risk identification and assessment scenarios?

The demonstrable alignment of the methods with international risk management standard processes indicates their generality and broader relevance, particularly to communities that already use such processes. This includes the use of standard and familiar risk assessment techniques, including the risk matrix approach. The methods require no particularly mathematical or computational expertise and thus have a low barrier to usability, which increases the likelihood for re-use by different organisations. The generality of the method is further demonstrated by its application to a number of different collection and system scenarios that feature a range of different types of risks. This indicates a good degree of reusability for different scenarios. The characterisation model, which represents a generic method for characterising risk, is sufficiently abstracted to potentially support general re-use by several different domains. Wider testing is required to substantiate this.

Evaluation of the Methodology

Evaluation is an essential stage of design science research. March and Smith's criteria focus on evaluation of the artefacts, though Hevner *et al*. (2004), take a different approach and propose a framework for evaluating the overall application of the methodology to the research area. They identify seven guidelines for effective design science research, all of which should be addressed in some form before the research may be considered complete. The guidelines are oriented

primarily around the production of design science IT artefacts within an information systems environment. Whilst this explicitly includes underlying models, constructs and methods, as well as hardware and software, the guidelines nonetheless twice make reference to the technological aspect of outputs, specifically the development of technology-based solutions and communication to technology-oriented audiences. These references have been removed so that the guidelines remain usable for the purposes of this research, which produces more widely relevant artefacts that are of use within but also beyond an information systems domain.[80] Each guideline is featured below with a corresponding explanation of how it has been addressed.

'Design as an Artefact: Design Science Research must produce a viable artefact in the form of a construct, model, method or instantiation'. This research has produced several artefacts: a construct to define the vocabulary of the problem and solution space; models to explore and represent the solution space in relation to the problem area; and methods that apply and extend the models to demonstrate an answer to the research question. The viability or feasibility of the artefacts is demonstrated by their applicability to each subsequent artefact in cascading succession and ultimately through the case study implementation of the research at the British Library.

'Relevance of the problem: The objective of design science research is to develop […] solutions to important and relevant business problems'. The importance and relevance of the problem is outlined in the problem statement and research motivation stipulated in chapter one of this thesis, with a particular relevance for the British Library. The solutions review in chapter three found that whilst previous work on digital preservation risk was helpful in certain areas, it did not generate a consistently clear or holistic perspective on the range of potential risks, nor did it often do so in a manner that was demonstrably compatible with corporate risk management processes. This research developed a solution, described in chapters five and six, that significantly improves disciplinary understanding of the nature and

[80] Amended guidelines are identified through the use of square brackets in affected areas of text.

complexity of digital preservation risk so that it can be managed in a more thorough, comprehensive, and consistent fashion than before. The outputs thus address a significant problem within the discipline and solve an important, relevant business need for the British Library.

'Design Evaluation: The utility, quality and efficacy of a design artefact must be rigorously demonstrated via well-executed evaluation methods'.[81] The appropriateness of an evaluation method is dependent on the nature of the artefact and the chosen evaluation metrics. Appropriate design evaluation methods for this guideline range from the observational and analytical to experimental, testing and descriptive (Hevner *et al.*, 2004, p. 86). A qualitative approach is appropriate to the nature of the artefacts in this research, as their conceptual form is not associated with any quantifiable performance metrics that could be assessed using experimental, analytical, or test-based approaches. The evaluation approach used in this research has instead assessed the artefacts against the requirements for the solution, and the criteria suggested by March and Smith for the construct, models and methods. Execution of this evaluation process is thoroughly described in this chapter.

Hevner *et al*. also argue that the design evaluation should include an assessment of style. The style of the models was well received by colleagues and commended for their clarity. The use of colour was found helpful to indicate different component parts of the models, including the application of consistent colouring through both context and risk source models. Reservations from an archival reviewer about whether UML notation was appropriate for the target audience, particularly non-technology colleagues, led to the decision to use minimal notation. It also led to the development of an alternative representation of the main risk source model in the form of a question-based framework, which subsequently took the form of method two.

[81] This criterion uses three related terms that are not defined by Hevner *et al*. Different interpretations of the terms can be applied to different types of artefact and different problem/solution spaces. Each is understood here as follows: Utility relates to overall usability; Quality relates to the degree to which requirements are satisfied; Efficacy relates to effectiveness and having the desired outcome – i.e. whether it satisfies the main objective.

'Research Contributions: Effective DSR must provide clear and verifiable contributions in the areas of the design artefact, design foundations, and/or design methodologies'. The main contribution of this research is the series of knowledge artefacts it has produced: the construct, models, and methods. In order to support and encourage re-use within the field, these are also available as a series of separate documents, available independently of the thesis. The main artefact is the CHARM Reference Model for Conceptualising and Characterising Digital Preservation Risk. This collates all of the models and the glossary produced for this research into a single document. Also available is a 'How-To' Guide with suggestions on how to use CHARM, re-usable templates for methods two and three (i.e. a Risk Identification Framework and a Risk Assessment Spreadsheet), and a separate file representing the whole Digital Preservation Risk Source Model on a single page. Collectively, these represent the practical contributions of this PhD submission.

The contribution and advancement to knowledge achieved through these artefacts is clearly illustrated against the current state of the art as identified in the chapter three solutions review. They extend the existing knowledge base by establishing a rigorous, justified, and meaningful definition of digital preservation risk, differentiating between the concept of risk and the characterisation of risk. Models and methods can both be used by the wider community for risk assessment activities as appropriate to their needs, and inform subsequent preservation planning endeavours. This represents a significant contribution to knowledge and closes a gap between our prior conceptual understanding of risk and the logic of our institutional responses and requirements for managing digital preservation risk.

The DSR Knowledge Contribution Framework (Gregor and Hevner, 2013 p. 345) provides a useful mechanism for classifying the type of knowledge contribution made by this research. The Framework identifies four quadrants: routine design, exaptation, improvement, and invention. Routine design applies known solutions to known problems and thus does not make a major contribution to knowledge, though improvement, invention, and exaptation all represent a research and knowledge generation opportunity. The research presented in this thesis and represented in its artefacts falls clearly into the 'Improvement' quadrant of the

Framework, representing a new solution for a known problem whilst providing a deeper understanding and insight into the problem itself. This deeper insight into the problem itself arguably pushes the solution contribution towards its border with the invention quadrant, which invents new solutions for new problems. It nonetheless remains an improvement rather than an invention, as the problem is not wholly new.

Positioned in light of Kenney and McGovern's 'Five Organisational Stages of Digital Preservation' framework (2003), the solution represents a capability transition from stage two or three of a digital preservation project or programme, which is typically the form in which digital preservation risk initiatives are developed and implemented, to a stage four institutionalisation capability that can incorporate and situate the implementation within a larger institutional environment. Positioning of the contribution against these two frameworks helps establish and demonstrate its significance.

'Research Rigor: Design science research relies upon the application of rigorous methods in both the construction and evaluation of the design artefact'. The solution was developed through the rigorous application of the design science research methodology from Peffers *et al*. (2006). Each stage of the process is thoroughly presented and detailed in this thesis, from problem statement and objectives definition, to design and development, demonstration, and evaluation. The only stage yet to be fully addressed is the communication stage, explored in more detail further below. The rigour of the evaluation is outlined in this chapter, as already described in the paragraphs above. Research rigour is further demonstrated by the logical process that underpins the intellectual development of the models.

'Design as a Search Process: The search for an effective artefact requires utilising available means to meet desired ends whilst satisfying laws in the problem environment'. The laws (i.e. requirements) of the problem environment were clearly stipulated early in the research and helped frame a minimal set of requirements. As a design process, the research followed a pragmatic approach to devise its solution. This was an effective epistemological stance as it allowed the

author to explore different avenues of research without prescribing too much up front, and shape the final form of the research outputs accordingly. The available means were constrained to the nature of this research as a PhD endeavour, namely limitations on the time available that affected the depth of the research, limitation on wider engagement with the community that may have otherwise shaped the research, and the skills of the author to undertake and represent the research. The author's prior experience in the field functioned as a helpful means for 'hitting the ground running' with the research. As a member of staff, the author's access to British Library systems, colleagues, and collections meant that these could be used as exemplars with which to test and refine the solution. These circumstances and opportunities were integrated into the research process as means allowed.

'Communication of Research: DSR must be presented effectively to [relevant] audiences'. Communication is typically the final stage during which the research is presented to other researchers and practising professionals. This is partially addressed already through early feedback from supervisors and colleagues both internal and external, representing different professional roles including technology, management and preservation practitioners. Much of this feedback features in the chapter six case study description. The nature of the research as a PhD thesis and the requirement for originality however necessitates a degree of restriction in wider communications until such time as the thesis is submitted. Communications with the wider scholarly and practitioner communities will be expanded thereafter, including promotion of the openly available reference model documentation. A review period is expected prior to potential updates in response to feedback received.

Conclusions

This chapter has considered how well the solution has met the requirements for the research, whether the artefacts can be considered to comply with high level criteria common to design science research, and whether the design science research methodology has been faithfully and consistently applied to the research problem. In exploring these three perspectives, it represents a thorough and considered evaluation of the research as required by the research methodology. A review of

the research outputs against the requirements that represent the research question confirms that the research has established a good, thorough, and viable solution. A review of the main practical artefacts against the DSR evaluation criteria proposed by March and Smith validates their status as design science research artefacts, confirming that they are well-formed and usable for their expected purpose. A review of the research as an application of a design science methodology further supports this conclusion, justifying the relevance of the research, the rigour of the research methodology, and the importance of the knowledge contribution it represents. Overall, the evaluation suggests that the theory on the utility of risk science for this research was correct and that risk science can enable a more rigorous, justified, transparent and reasoned understanding of the concept and nature of digital preservation risk than previously seen within the field. Risk science represents a valuable interdisciplinary addition to both the theory and practice of digital preservation for disentangling the complexity and uncertainty of digital preservation risk.

Evaluation of the research presented in this thesis against the three-tiered evaluation framework used in this chapter thus validates it from multiple perspectives. This is not to say that the outputs are perfect - the chapter acknowledges that some criteria are more demonstrably met than others, and that only through broader release to the wider community can the research truly be validated. Within the limitations of the research however, this qualitative evaluation ultimately concludes that the research outputs meet the objectives of the research, especially considering the extraction of key knowledge artefacts into a separate series of documents for dissemination. The CHARM Reference Model provides a deeper understanding and thorough representation of the nature and complexity of digital preservation risk, supporting the foundations for a more flexible and comprehensive preservation planning risk response as demonstrated through the methods. This thorough review and analysis confirms the rigour of both the research process and the research outputs to satisfactorily address the problem space outlined at the start of this thesis.

Chapter Eight: Conclusion

Introduction

This thesis has explored in depth the tangled concept of digital preservation risk. It has shown how an interdisciplinary approach can support a more thorough and rigorous exploration of digital preservation risk by combining risk science with design science to develop a deeper understanding of both the concept and potential practical solutions for managing it. This chapter provides a summative review of the research and artefacts presented in this thesis, discussing its implications for the British Library and the wider digital preservation community. It demonstrates how the research question has been answered, frames the significance of the contribution to knowledge made by the research, and makes suggestions for further development and implementation of the research going forwards.

Revisiting the Research Question

Prior to the research presented in this thesis, many different solutions had been devised over the years to explore and address the wicked problem of digital preservation risk. Despite this, it remained a nebulous concept characterised by uncertainty, inconsistency, and repeated reinventions of the wheel. This research directly addressed that problem in order to improve our ability to demonstrate a cohesive, consistent and comprehensive understanding of digital preservation risk going forwards, supporting a more comprehensive preservation planning risk response than typically currently practised within the field.

The motivation to address this problem came from the author's role at the British Library, where they are responsible for managing and preserving an exceptionally large and heterogeneous digital collection that already spans over four decades of technological and organisational change. The author's kernel knowledge, developed through over two decades of practical experience, had made clear that the existing toolsets did not meet the Library's needs. A more nuanced understanding of digital preservation risk was needed to facilitate improved and holistic risk management,

particularly in relation to the concept of preservation planning. This led to the research question: *How can the nature and complexity of digital preservation risk be more thoroughly and consistently represented so as to support the foundations for a more flexible yet comprehensive preservation planning risk response?*

The research followed a logical and clearly structured methodology to answer this question. It first analysed the current solution space to establish the main areas where improvement was required. It then went back to basics to establish a comprehensive definition of digital preservation that represents its overall domain, purpose and function. This formed the conceptual foundation from which to build and develop a meaningful and substantiated definition of digital preservation risk: 'the potential for complete or partial loss of digital collection content in terms of its target values of retrievability, authenticity, integrity, accessibility, and longevity, arising from sub-optimised risk sources within the managed organisational and technological environment in which the content should otherwise be preserved'. This definition utilises a structure adopted from the field of risk science for the purposes of precision and utility, representing the conceptual nature of digital preservation risk as a complex set of interdependencies between entities that, if not appropriately managed, can threaten an organisation's ability to achieve its preservation objective.

Distinguishing between the concept of risk and the characterisation of risk led to a deeper investigation into the complexity of digital preservation risk that developed around the concept of the 'risk source'. Risk Sources represent changeable elements within the digital preservation environment that alone or in combination with others have the intrinsic potential to give rise to a negative outcome. These threaten an organisation's ability to maintain the target values for its digital content and thus achieve its preservation objective. Explication and analysis of this risk source concept revealed a series of logical, structural entities that could be used to describe and define different aspects of risk. This was subsequently applied across the whole domain setting, using core entities from the definition of digital preservation as a set of grounding reference points, to produce a thorough representation of the complexity of digital preservation risk.

The resulting series of models and the underlying vocabulary are presented and discussed in this thesis and its associated practical output, the CHARM Reference Model for Conceptualising and Characterising Digital Preservation Risk. An associated set of methods was also designed to support use of the model for risk assessments, represented in the corresponding CHARM 'How-To' Guide. These outputs represent the primary design science research artefacts produced by this work. In addition, a review and comparison of the ISO 14721 preservation planning function with the ISO 31000 risk management standard established the direct relationship between the two and observed the limitations of preservation plans that are intended to ensure content remains accessible, understandable, and usable but that focus primarily on technological and obsolescence-based risks. The risk model and methods explored in this thesis can thus be used to support the foundations for an improved and holistic preservation planning risk response compared to that currently typically practised within the field.

By differentiating between the concept and characterisation of risk and charting a logical progression from one to the other, this research demonstrates how risk science can support an improved and clearer understanding of digital preservation risk. It has produced a solution that represents a holistic perspective on digital preservation risk that can be used to produce comprehensive, scenario-based risk assessments, in line with internationally recognised standards for risk management. These outputs represent a thorough answer to the research question on how the nature and complexity of digital preservation risk can be more thoroughly and consistently represented, so as to support the foundations for a more flexible yet comprehensive preservation planning risk response.

Research Scope and Contribution

The research presented in this thesis introduces and applies rigorous and substantiated methods to support both deeper analysis of the problem space and construction of a solution. The scope is broad and ambitious in both subject and relevance, pertinent to any institution that needs to preserve for the long term but particularly cultural heritage and memory institutions – of which there are many thousands around the world. The breadth of the research covers the entirety of the

digital preservation risk landscape, whilst the depth of the research represents a thorough and extensive analysis of both the problem and the solution space. The interdisciplinary and novel nature of the research, combining risk science with design science to explore an already interdisciplinary field, brings an extra depth and dimension to this already broad scope and demonstrates how the interdisciplinary application of concepts, methodologies and approaches from these fields can help further digital preservation knowledge in both theory and practice. The solutions review presented in chapter three provides new insight into the disconnect between theory and practice, and the limitations of the current toolsets. The models and the vocabulary developed in chapter five represent a new, creative solution, whilst the methods explored in chapter six provide ways to utilise that solution to achieve end goals and improve risk management practices. The scope and the scale of the research is extensive, grounded in logic and built from a firmly established conceptual basis that justifies the overall approach.

These explorations result in multiple contributions to knowledge. Whilst the digital preservation community has many different ways to identify and measure risk, investigations during the early stages of the research revealed that the underlying concept of digital preservation risk itself is often only weakly defined. Analysis of the solutions developed and used within the field to date compared to the wider disciplinary literature on risk revealed gaps between theory and practice. Serious reservations were found on the suitability of scored approaches to risk assessment based on theoretical measures with relatively little underlying justifying evidence, yet such approaches continued to be developed. New assessment frameworks appear relatively frequently that explore similar ground to those that have gone before them, but without clearly demonstrating how they build on and improve prior work. There is a clear tendency to ‘re-invent the wheel’. The analysis revealed unjustified inconsistency in how the same measures are expressed in different frameworks, uncertainty about the relevance of measures used in different frameworks, and even doubts about the suitability of some widely used assessment frameworks for the purposes of risk management.

From a more positive, developmental perspective, the analysis of existing solutions resulted in a new classification scheme for existing assessment frameworks. This represents a hierarchy of focus in terms of either formats, objects, systems, or organisations, and establishes the need for a risk assessment framework that can function at different levels. Further analysis identified a series of 'lessons learned' that highlighted the importance of precisely defined terms, of clearly defining – and distinguishing between – the concept of risk and the manifestation of risk, and on the importance of a clear methodology. These lessons represent valuable insight and knowledge about some of the problems that have prevented disciplinary advancement in understanding of digital preservation risk to date, and what to avoid going forwards.

The models designed in response to the research question and lessons learned exercise – collectively represented in the CHARM Reference Model document - represent the main knowledge contribution made by this thesis. They show how risk science provides a structure for a more in-depth exploration of digital preservation risk, disentangling the concept of digital preservation risk to highlight its main conceptual and relational entities, in particular the risk source. Explication of the risk source concept to generate a logical data structure for exploring and characterising individual digital preservation risks, represents new insight into the different components of risk and how they relate to one another. Application of this structure across the entire digital preservation risk ecosystem enables risks to be described consistently and thoroughly regardless of their type. Implementation of the structure at an abstract domain level delivers an artefact that functions as a universal reference model, representing prescriptive knowledge that can subsequently be used by practitioners to derive and generate situational knowledge in the form of contextualised risk assessments. Further knowledge about how to achieve this is represented in the suggested methods for implementation, whilst the glossary (or construct) provides the background knowledge to understand the key terms used in the model and enable accurate interpretation.

The decision to develop a reference model was taken to facilitate its relevance to the wider community. The abstract nature of the model supports its use across the

domain for which it has been developed, in this case digital preservation, without prescribing the form or institutional context in which it should be used. Application of the model to a given risk scenario using the suggested methods enables meaningful characterisation of risk in a practical setting. The reference model thus supports a deep yet situationally agnostic understanding of the nature and complexity of digital preservation risk, at both an abstract and practical level. The comprehensive knowledge represented in this research therefore fills what is otherwise a gap in disciplinary knowledge between theoretical expectations of risk and practical solutions for risk management.

Significance of the Contribution

The results and artefacts presented in this thesis are considered to be both empirically and theoretically founded, with the potential to measurably advance the maturity of disciplinary capabilities for applied digital preservation risk management. It thus makes a significant disciplinary contribution to both knowledge and practice. The level of the contribution to practice can be demonstrated against Kenney and McGovern's 'Five Organisational Stages of Digital Preservation' model (2003), which functions as a form of maturity model. Maturity models are widely used in many different settings to assess current capabilities, provide a baseline against which to assess improvements, and identify a roadmap for optimisation. The Kenney and McGovern model identifies five stages of an organisational response to digital preservation:

1. Acknowledge: Understand that digital preservation is a local concern
2. Act: Initiate digital preservation projects
3. Consolidate: Segue from projects to programmes
4. Institutionalise: Incorporate the larger environment
5. Externalise: Embrace inter-institutional collaboration and dependency

The experience of the author and the analysis of other assessment frameworks undertaken during this research indicates that most of the field - particularly with regards to digital preservation risk - is still in the early stages of maturity, most likely in the area of stage two and occasionally stage three. Stage two is characterised by

activities which are project-based, often funded by external monies, not integrated with the wider organisational functions, and of limited duration despite a focus on longer term solutions. One of more of these features is evident in almost all of the assessment frameworks explored in this thesis. Statements made to the Digital Preservation Coalition during the membership application process corroborate this, frequently referring to short term projects, ambitions to integrate digital preservation beyond a single team (and in many cases single person), and an interest in policy development.[82] This relative immaturity is also suggested in a report on the State of the Art in Digital Preservation (Rieger, 2018), which found that in general, risk assessment and management strategies were 'inadequate' and framed digital preservation still as a community, rather than institutional, endeavour. Although that report is now almost five years old, the author's experience, substantiated by the more recent work analysed and reviewed in chapter three, suggests that relatively little has changed.

This research makes a substantive contribution to the field as it marks a capability transition from stages two or three of Kenney and McGovern's model to stages three or four, enabling incorporation of holistic digital preservation risk management practice into the wider institutional environment and alignment with standard risk management practices. Only through a significant contribution to knowledge can such an advance be made. The research advances the interdisciplinary knowledge base of digital preservation by using approaches drawn from risk science and design science to support and advance disciplinary thinking about risk. The outcome of this is a thorough and scientific foundation for the future of digital preservation risk management. By applying a rigorous and transparent methodology to the production of its model and suggesting methods for its use, the research addresses a limitation in the design of previous assessment frameworks highlighted by Maemura, Moles and Becker (2017, p. 1619). The transferable nature of this research is demonstrated by the real world application of the reference model specifically to the British Library context, a microcosm

[82] Access to member application forms is limited to Board members and staff. The author has seen and reviewed these as a Board member of the DPC and has the agreement of the DPC Executive Director to acknowledge them anonymously in this thesis.

representation of digital collecting heritage institutions. In terms of the DSR Knowledge Contribution Framework (Gregor and Hevner, 2013 p. 345), it makes a clear improvement-based and prescriptive contribution to knowledge that not only develops a new solution but also provides a deeper understanding of the problem at hand.

This research provides a new way for thinking about and responding to digital preservation risk. The insight and knowledge it contains will enable the British Library to improve its approach to managing digital preservation risk whilst also contributing a new tool for risk assessment to the wider digital preservation community. The contribution this new knowledge makes to the field is of high conceptual and practical significance, as it has potential to substantially inform the way the discipline understands digital preservation risk and the subsequent practice of digital preservation risk management.

Further Research and Next Steps

There are several avenues that future research may take to build on the knowledge represented in this thesis and the artefacts produced by the research. There are research opportunities to improve the artefacts, explore re-use of the artefacts, incorporate the wider preservation planning concept into the methods, extend the research to address other aspects of design and risk science, and explore opportunities for re-use of the model structure to other settings. The first opportunity however pertains to stage six of the design science research methodology: communication to the wider audience.

Communication of the research to relevant audiences provides an opportunity to gather broad theoretical and practical feedback to improve the research and enable its contribution to the discipline. It highlights not only the rigour, utility, novelty and effectiveness of artefacts but also the relevance of the problem area that they address. Several research papers and communications activities are planned to support this. Conferences including iPRES and PASIG are potential avenues to reach the digital preservation community, whilst the DESRIST conference series and the SRA Journal are opportunities to engage with the design science research and risk

science communities. iPRES poster and panel sessions will be proposed in addition to papers, in order to encourage direct interaction, exploration, and feedback specifically with the digital preservation community. The author is already in discussion with the Alexandria journal about a potential paper, whilst opportunities to publish in JASIST and the Journal for Documentation, as well as other relevant outlets, will also be considered.

The artefacts will also be promoted and feedback invited through non-academic channels. Blog posts, social media, webinars and smaller events can be used for this purpose, as well as outreach to digital preservation membership organisations such as the DPC, the OPF, and the NDSA, alongside IFLA. It may be possible to co-opt the Reference Model into a community standard, for example by working with the aforementioned membership organisations, with an update schedule for revisions. Opportunities for discussions on the feasibility of this will be sought in due course.

This wider community engagement will provide an opportunity to seek feedback for refinement of the models and methods. Whilst the utility of the artefacts has been demonstrated, there are nonetheless improvements that could be made. The evaluation section highlighted some of these, for example in relation to the simplicity and elegance of the vocabulary, and the operationality and efficiency of the methods. Community feedback on the vocabulary will help identify terms that are understood differently by different user groups as well as terms excluded from the glossary that need to be included, whilst wider implementation of the methods will test and highlight potential challenges in translating the abstract concepts in the model to situational, characterised risks. Concepts of instances and instance types are areas where the author feels that wider feedback would be especially helpful, particularly given the indicative rather than comprehensive nature of the instance types as currently suggested. Feedback on method two would also be particularly helpful as this is currently a demonstrator method for the model rather than a tried and tested method. With wider use, accumulations of completed risk assessment spreadsheets will enable analysis for trends and patterns that not only help refine the model but also provide insight into particularly prevalent or

prioritised risks from across the community or particular types of institutional contexts.

There may also be an opportunity for further methods to be developed that reuse the CHARM Reference Model in different ways. It could function as an underlying dataset or risk structure for new methods, for example with different types of risk assessment techniques. More novel applications of the model might translate it into other forms, for example a card game that could support learning about digital preservation risk in educational settings. The methods themselves might also be further abstracted from their representation in this thesis to help further test and refine the underlying concepts. The vocabulary has its own reuse potential for standardising the way the community talks about digital preservation risk independently of the model, thus addressing one of the underlying problems in the community of inconsistently used terms and definitions. Discussion would be required to drive this suggestion forward and find ways to standardise the terminology for wider community use.

Reusability of the structure of the model provides a further opportunity for future research. The structure of a contextualised risk source and a characterised risk statement is disciplinarily agnostic and should in theory therefore be reusable outside of a digital preservation setting, to consistently describe risks in other domains. Can the underlying structure of the risk source model be reused to establish the main risk originating entities, risk source classes, factors and instance types in other domains? If so then can the methods also transfer to those domains, for example to construct risk statements with the same structure as used in method three? This could increase both the impact and significance of the research contribution by extending it from a demonstration of applied risk science, to the production of generic, fundamental risk science, in the form of a disciplinarily agnostic risk source model structure. It would also provide an opportunity to progress the ‘theorise and justify’ research activities of design science by further explicating and analysing the characteristics of the artefacts as well their transferability to different domains. The wider application of risk science to digital

preservation is a further potential research area, particularly from a theoretical perspective. What else can the community learn from this emerging field?

The research could also be extended to explore the concept of preservation planning in more detail. Such exploration might occur within the context of future revisions to the ISO 14721 standard, or produce practical methods and templates for a more comprehensive preservation planning risk response than typically seen in the community at present. The British Library will address this latter suggestion in the form of the IPS project (May, Pennock and Russo, 2019). This system, designed originally to support technical preservation plans, will be reviewed to determine how best to support the wider understanding of digital preservation risks represented in this research. Early plans are to enhance the policy suite of the system to include risk assessments, and to design a more flexible preservation plan template that can capture and respond to different types of risks, prioritised as part of a scenario-based risk assessment.

This research has demonstrated that by extending the interdisciplinary reach of digital preservation beyond established fields of archival science, library and information science, and computer science, to design science and the emergent field of risk science, new opportunities can be identified that extend, enhance, and enrich the existing knowledge base in digital preservation and digital preservation risk. The community has a habit of talking mainly to itself – as a colleague on stage remarked at iPRES 2019, 'where are the representatives of other fields at our conference, why are we only talking amongst ourselves'? One could ask instead, 'where are *we* represented in other conferences, and why are *we* mainly talking amongst ourselves'? Better outreach and engagement with other fields is essential if the field is to truly develop, not just in terms of digital preservation risk but digital preservation in general, beyond the current five levels of the Kenney and McGovern maturity model, to fully embrace not just inter-institutional collaboration but interdisciplinary collaboration. This research demonstrates the potential of that for digital preservation risk, culminating in the production of a novel and creative, yet rigorous and practical solution to the research problem. Opportunities such as

those outlined in this chapter should be explored to build further on this experience and its research. We remain always shooting at a moving target, after all.

Bibliography

Abrams, S. (2015) 'A foundational framework for digital curation: The Sept domain model', *Proceedings of the 12th International Conference on Digital Preservation, iPRES 2015*, Chapel Hill, 2 - 6 November, pp. 30 - 39. Available at: https://phaidra.univie.ac.at/detail/o:429533 (Accessed: 1 September 2023).

Abrams, S. (2021) 'Tacit attitudinal principles for evaluating digital preservation success', *Archival Science,* Volume 21, pp. 295 - 315. Available at: https://doi.org/10.1007/s10502-021-09360-5 (Accessed: 16 September 2023).

Abrams, S. (2023) *A Communicological Critique of Evaluative Norms for Digital Preservation Success.* PhD thesis. Queensland University of Technology. Available at: https://eprints.qut.edu.au/238194/1/Stephen%2BAbrams%2BThesis%283%29.pdf (Accessed: 1 September 2023).

Abrams, S., Cruse, P. and Kunze, J. (2012) *Total Cost of Preservation (TCP) : Cost and Price Modeling for Sustainable Services.* University of California Curation Center working paper. Available at: https://wiki.diglib.org/images/0/0a/TCP-total-cost-of-preservation.pdf (Accessed: 1 September 2023).

Aliseda, A. (2007) 'Abductive Reasoning: Challenges Ahead', *Theoria,* 22(3), pp. 261-270. Available at: https://doi.org/10.1387/theoria.446 (Accessed: 16 September 2023).

Altman, M. and Landau, R. (2020) 'Selecting efficient and reliable preservation strategies: modeling long-term information integrity using large-scale hierarchical discrete event simulation', *International Journal of Digital Curation, IJDC,* 15(1). Available at: https://doi.org/10.2218/ijdc.v15i1.727 (Accessed: 13 September 2023).

American Library Association (2008) *Definitions of Digital Preservation.* ALCTS Preservation and Reformatting Section working paper I.D. 21609b50-bc60-46e4-848e-dc5fdabdb128. Available at: https://www.ala.org/alcts/resources/preserv/defdigpres0408 (Accessed: 1 September 2023).

Anderson, D. R. *et al*. (2009) *An Introduction to Management Science: Quantitative Approaches to Decision Making.* London: Cengage.

Antunes, G. *et al*. (2011) 'Modeling Contextual Concerns in Enterprise Architecture', *2011 IEEE 15th International Enterprise Distributed Object Computing Conference Workshops*, Helsinki, Finland, 29 August - 2 September. pp. 3 - 10. Available at: https://doi.org/10.1109/EDOCW.2011.9 (Accessed: 13 September 2023).

Aven, T. (2013) 'On the Meaning and Use of the Risk Appetite Concept', *Risk Analysis,* 33(3), pp. 462 - 468. Available at: https://doi.org/10.1111/j.1539-6924.2012.01887.x (Accessed: 16 September 2023).

Aven, T. (2016) 'Risk Assessment and risk management: Review of recent advances on their foundation', *European Journal of Operational Research,* 253(1), pp. 1 - 44. Available at: https://doi.org/10.1016/j.ejor.2015.12.023 (Accessed: 16 September 2023).

Aven, T. (2020) 'Risk Science Contributions: Three Illustrating Examples', *Risk Analysis: An International Journal,* 40(10), pp. 1889-1899. Available at: https://doi.org/10.1111/risa.13549 (Accessed: 16 September 2023).

Aven, T. and Krohn, B. S. (2014) 'A new perspective on how to understand, assess and manage risk and the unforeseen', *Reliability Engineering and System Safety,* Volume 121, pp. 1 - 10. Available at: https://doi.org/10.1016/j.ress.2013.07.005 (Accessed: 16 September 2023).

Aven, T. and Thekdi, S. (2022) *Risk Science: An Introduction.* New York: Taylor and Francis.

Bakhshandeh, M. *et al*. (2013) 'Towards a Legal Ontology for the Digital Preservation Domain', *International Conference on ICT LAW 2013,* Porto, Portugal, 8 - 9 November. Available at: https://www.inesc-id.pt/ficheiros/publicacoes/9533.pdf (Accessed: 1 September 2023).

Ball, D. J. and Watt, J. (2013) 'Further Thoughts on the Utility of Risk Matrices', *Risk Analysis,* 33(11), pp 2068-2078. Available at: https://doi.org/10.1111/risa.12057 (Accessed: 16 September 2023).

Barateiro, J. *et al*. (2010) 'Designing Digital Preservation Solutions: A Risk Management-Based Approach', *International Journal of Digital Curation (IJDC),* 5(1), pp. 4 - 17. Available at: https://doi.org/10.2218/ijdc.v5i1.140 (Accessed: 13 September 2023).

Barateiro, J., Antunes, G. and Borbinha, J. (2011) 'Long-term security of digital information: Assessment through Risk Management and Enterprise Architecture', *2011 IEEE EUROCON - International Conference on Computer as a Tool*, Lisbon, Portugal, 2011, pp. 1-4. Available at: https://doi.org/10.1109/EUROCON.2011.5929270 (Accessed 13 September 2023).

Barateiro, J., Burda, D. and Simon, D. (2013) 'Leveraging DP in Commercial Contexts through Enterprise Risk Management', *Proceedings of the 10th International Conference on Preservation of Digital Objects: iPRES 2013,* Lisbon, Portugal, 2 - 6 September. pp. 104 - 109. Available at: https://purl.pt/24107 (Accessed: 13 September 2023)

Barons, M. *et al*. (2021) 'Safeguarding the nation's digital memory: towards a Bayesian model of digital preservation risk', *Archives and Records,* 42(1), pp. 57-78.

Available at: https://doi.org/10.1080/23257962.2021.1873121 (Accessed: 16 September 2023)

Barwick, J. (2012) *Where have all the games gone? An exploratory study of digital game preservation.* PhD thesis. Loughborough University. Available at: https://hdl.handle.net/2134/10222 (Accessed: 16 September 2023).

Baskerville, R. (2008) 'What design science is not', *European Journal of Information Systems,* Volume 17, pp. 441 - 443. Available at: https://doi.org/10.1057/ejis.2008.45 (Accessed: 16 September 2023).

Baskerville, R. *et al*. (2018) 'Design Science Research Contributions: Finding a Balance between Artifact and Theory', *Journal of the Association for Information Systems (JAIS),* 9(5), pp. 358 - 376. Available at: https://aisel.aisnet.org/jais/vol19/iss5/3 (Accessed: 16 September 2023).

Baskerville, R. *et al*. (2019) 'Inducing Creativity in Design Science Research' *Extending the Boundaries of Design Science Theory and Practice. DESRIST 2019. Lecture Notes in Computer Science*, vol 11491. Springer, Cham. pp. 3 - 17. Available at: https://doi.org/10.1007/978-3-030-19504-5_1 (Accessed: 14 September 2023).

Baybutt, P. (2018) 'Guidelines for designing risk matrices', *Process Safety Progress,* Volume 37, pp. 49 - 55. Available at: https://doi.org/10.1002/prs.11905 (Accessed: 16 September 2023).

Beagrie, C. (2006) 'Digital Curation for Science, Digital Libraries, and Individuals', *International Journal of Digital Curation (IJDC),* 1(1), pp. 3 - 16. Available at https://doi.org/10.2218/ijdc.v1i1.2 (Accessed: 14 September 2023).

Bearman, D. (1989) *Archival methods: Archives and Museum Informatics Technical Report #9.* Available at: http://www.archimuse.com/publishing/archival_methods/ (Accessed: 1 September 2023).

Becker, C. (2018) 'Metaphors We Work By: Reframing Digital Objects, Significant Properties, and the Design of Digital Preservation Systems', *Archivaria, The Journal of the Association of Canadian Archivists,* Volume 85, pp 6 - 36. Available at: https://archivaria.ca/index.php/archivaria/article/view/13628 (Accessed: 16 September 2023).

Becker, C. *et al*. (2009) 'Systematic planning for digital preservation: evaluating potential strategies and building preservation plans', *International Journal on Digital Libraries,* Volume 10, pp. 133 - 157. Available at https://doi.org/10.1007/s00799-009-0057-1 (Accessed: 14 September 2023).

Becker, C. and Rauber, A. (2011) 'Decision criteria in digital preservation: What to measure and how', *Journal of the American Society for Information Science and Technology,* 62(6), pp. 1009-1028. Available at: https://doi.org/10.1002/asi.21527 (Accessed: 14 September 2023).

Becker, C. *et al*. (2011) 'Control objectives for DP: Digital preservation as an integrated part of IT governance', *Proceedings of the American Society for Information Science and Technology,* 48(1), pp. 1 - 10. Available at: https://doi.org/10.1002/meet.2011.14504801124 (Accessed: 14 September 2023).

Becker, C., Faria, L. and Duretec, K. (2015) 'Scalable Decision Support for Digital Preservation: An Assessment', *OCLC Systems and Services: International digital library perspectives,* 30(4), pp. 249 - 284. Available at: https://doi.org/10.1108/OCLC-06-2014-0026 (Accessed: 16 September 2023).

Becker, C., Maemura, E. and Moles, N. (2020) 'The Design and Use of Assessment Frameworks in Digital Curation', *Journal of the Association for Information Science and Technology,* 71(1), pp. 55 - 68. Available at https://doi.org/10.1002/asi.24209 (Accessed: 14 September 2023).

Beck, M. (2004) 'Obstacles to the Evolution of Risk Management as a Discipline: Some Tentative Thoughts', *Risk Management: An International Journal,* 6(3), pp. 13 - 21. Available at: https://doi.org/10.1057/palgrave.rm.8240186 (Accessed: 14 September 2023).

Bennett, J. C. (1997) *A Framework of data types and formats, and issues affecting the long term preservation of digital material,* London: British Library Research and Innovation Centre. Available at: https://researchportal.bath.ac.uk/en/publications/a-framework-of-data-types-and-formats-and-issues-affecting-the-lo (Accessed: 14 September 2023).

Berlinger, E. and Váradi, K. (2015) 'Risk Appetite', *Public Finance Quarterly,* 60(1), pp. 49 - 62. Available at: https://unipub.lib.uni-corvinus.hu/2858/ (Accessed: 14 September 2023).

Bermès, E. and Fauduet, L. (2009) 'The Human Face of Digital Preservation: Organizational and Staff Challenges, and Initiatives at the Bibliothèque nationale de France', *Proceedings of the Sixth International Conference on Preservation of Digital Objects: iPRES 2009*, San Francisco, 5 - 6 October. Available at: https://escholarship.org/uc/item/6bt4v3zs pp. 24 - 29 (Accessed: 14 September 2023).

Bézivin, J. (2005) 'On the unification power of models', *Software and Systems Modeling,* Volume 4, pp. 171 - 188. Available at: https://doi.org/10.1007/s10270-005-0079-0 (Accessed: 14 September 2023).

Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2010) *Sustainable Economics for a Digital Planet: Ensuing Long-Term Access to Digital Information (final report).* Available at: https://discovery.ucl.ac.uk/id/eprint/19116/ (Accessed: 1 September 2023).

Blumenthal, K. *et al*. (2020) 'What's Wrong with Digital Stewardship: Evaluating the Organization of Digital Preservation Programs from Practitioners' Perspectives', *The*

Journal of Contemporary Archival Studies, Volume 7, Article 13. Available at: https://elischolar.library.yale.edu/jcas/vol7/iss1/13 (Accessed: 1 September 2023).

Brabazon, T., Hunter, N. and Quinton, J. (2022) 'The Scientist, the Artefact, and the Exegesis: Challenging the Parameters of the PhD', *IJCAS: The International Journal of Creative and Arts Studies,* June, 9(1), pp. 47 - 68. Available at: https://doi.org/10.24821/ijcas.v9i1.6409 (Accessed: 14 September 2023).

Bradley, K. (2007) 'Defining Digital Sustainability', *Library Trends,* 56(1), pp. 148 - 163. Available at: https://doi.org/ 10.1353/lib.2007.0044 (Accessed: 14 September 2023).

British Library (2017) *Sustaining the Value: British Library Digital Preservation Strategy 2017 - 2020.* London: British Library. Available at: https://www.bl.uk/digital-preservation/britishlibrary/~/media/c0068e4bb25c4ac39d1f4c5a020a21c8.ashx (Accessed: 1 September 2023).

British Library (2020) *Enabling Access for Everyone: the British Library's Content Strategy 2020 - 2023.* London: British Library. Available at: https://www.bl.uk/britishlibrary/~/media/bl/global/about%20us/policy%20documents/british-library-content-strategy-summary-final.pdf (Accessed: 1 September 2023).

Brown, A. (2007) 'Developing Practical Approaches to Active Preservation', *The International Journal of Digital Curation (IJDC),* 1(2), pp. 3 - 11. Available at: https://doi.org/10.2218/ijdc.v2i1.10 (Accessed: 14 September 2023).

Brown, A. (2013) *Practical Digital Preservation: A how-to guide for organisations of any size.* London: Facet Publishing.

Bryman, A. (2012) *Social Research Methods.* 4 ed. Oxford: Oxford University Press.

Buchanan, R. (1992) 'Wicked Problems in Design Thinking', *Design Issues,* 8(2), pp. 5-21. Available at: https://doi.org/10.2307/1511637 (Accessed: 14 September 2023).

Cabot, J. and Vallecillo, A. (2022) 'Modeling should be an independent scientific discipline', *Software and Systems Modeling,* Volume 21, pp. 2101 - 2107. Available at: https://doi.org/10.1007/s10270-022-01035-8 (Accessed: 14 September 2023).

Calhoun, K. (2014) *Exploring Digital Libraries: Foundations, Practice, Prospects.* London: Facet.

Canadian Institute of Conservation (2017) *Agents of Deterioration.* Ontario. Available at: https://www.canada.ca/en/conservation-institute/services/agents-deterioration.html (Accessed: 14 September 2023).

Candela, L. *et al*. (2007) *The DELOS Digital Library Reference Model: Foundations for Digital Libraries.* DELOS Network of Excellence on Digital Libraries project report.

Available at: https://www.researchgate.net/publication/200462045_The_DELOS_Digital_Library_Reference_Model_-_Foundations_for_Digital_Libraries (Accessed: 14 September 2023).

Caron, B. *et al*. (2015) 'Experiment, Document and Decide: A Collaborative Approach to Preservation Planning at the BNF'*, Proceedings of the Twelfth International Conference on Preservation of Digital Objects: iPRES 2015*, Chapel Hill, 2 - 6 November, pp. 54 - 58. Available at: https://services.phaidra.univie.ac.at/api/object/o:429538/get (Accessed: 14 September 2023).

Casadevall , A. and Fang, F. C. (2018) 'Elegant Science', *mBio,* 9(1). Available at: https://doi.org/10.1128/mbio.00043-18 (Accessed: 14 September 2023).

CEDARS project team (2001) *CEDARS project report (April 1998 – March 2001) 2001.* CEDARS project report. Available at: https://www.webarchive.org.uk/wayback/archive/20050410120000/http://www.leeds.ac.uk/cedars/pubconf/papers/projectReports/cedarsrepmar01exec.html (Accessed: 1 September 2023).

CEDARS (2002) *Cedars Guide to: Digital Preservation Strategies.* CEDARS project report. Available at: https://web.archive.org/web/20030802095942/http://www.leeds.ac.uk/cedars/guideto/dpstrategies/dpstrategies.html (Accessed: 28 August 2023).

Choudhury, S., Huang, C. and Palmer, C. L. (2020) 'Updating the DCC Curation Lifecycle Model', *International Journal of Digital Curation (IJDC),* 15(1). Available at: https://doi.org/10.2218/ijdc.v15i1.721 (Accessed: 14 September 2023).

CLIR (2000) *Authenticity in a Digital Environment,* Washington DC: CLIR. Available at: https://www.clir.org/pubs/reports/pub92/ (Accessed: 14 September 2023).

Cochrane, E. (2012) *Rendering Matters: report on the results of research into digital object rendering.* Archives New Zealand working paper. Available at: https://web.archive.org/web/20190125001936/http://archives.govt.nz/sites/default/files/rendering_matters.pdf (Accessed: 1 September 2023).

Cochrane, E. *et al*. (2019) 'Towards a Universal Virtual Interactor for Digital Objects', *Proceedings of the 16th International Conference on Digital Preservation: iPRES 2019*, Amsterdam, 16 - 20 September, pp. 191 - 200. Available at: https://ipres2019.org/static/proceedings/iPRES2019.pdf (Accessed: 14 September 2023).

Constantopoulos, P. and Dritsou, V. (2007) *An ontological model for digital preservation.* Online research article. Available at: https://www.researchgate.net/publication/228973204_An_ontological_model_for_digital_preservation (Accessed: 14 September 2023).

Constantopoulos, P. and Dallas, C. (2008) *Aspects of a digital curation agenda for cultural heritage.* Athens: Digital Curation Unit working paper. Available at: http://www.dcu.gr/wp-content/uploads/2016/10/Aspects-of-a-digital-curation-agenda-for-cultural-heritage.pdf (Accessed: 14 September 2023).

Conway, P. (1996) *Preservation in the Digital World,* Washington: Council on Library and Information Resources (CLIR). Available at: https://www.clir.org/pubs/reports/conway2/index/ (Accessed: 14 September 2023).

Conway, P. (2010) 'Preservation in the Age of Google: Digitisation, Digital Preservation, and Dilemmas', *The Library Quarterly: Information, Community, Policy,* 80(1), pp. 61-79. Available at: https://doi.org/10.1086/648463 (Accessed: 14 September 2023).

Cook, T. (1997) What's Past is Prologue: A History of Archival Ideas since 1898 and the Future Paradigm Shift. *Archivaria,* Issue 43, pp. 17 - 63.

CoreTrustSeal (2020) *Trustworthy Data Repositories Requirements.* Available at: https://www.coretrustseal.org/why-certification/requirements/ (Accessed: 1 September 2023).

Corrado, E. M. (2022) 'Digital Preservation Is Not Just a Technology Problem', *Technical Services Quarterly,* 39(2), pp. 143 - 151. Available at: https://doi.org/10.1080/07317131.2022.2045432 (Accessed: 14 September 2023).

Corrado, E. and Moulaison, H. L. (2014) *Digital Preservation in Libraries, Archives and Museums.* Maryland: Rowman and Littlefield.

Covello, V. T. and Mumpower, J. (1985) 'Risk Analysis and Risk Management: An Historical Perspective', *Risk Analysis,* 5(2), pp. 103 - 120. Available at: https://doi.org/10.1111/j.1539-6924.1985.tb00159.x (Accessed: 14 September 2023).

Cox, R. (1992) 'Archival Preservation Interests and Issues: An American Perspective', in B. Buckner Higginbotham and M. E. Jackson, eds. *Advances in Preservation and Access.* Westport: Meckler, pp. 228 - 243.

Cox, T. L. J. (2008) 'What's Wrong with Risk Matrices?', *Risk Analysis,* 28(2), pp. 497 - 512. Available at: https://doi.org/10.1111/j.1539-6924.2008.01030.x (Accessed: 14 September 2023).

Cumming, R. B. (1981) 'Is Risk Assessment a Science?', *Risk Analysis,* 1(1), pp. 1 - 3. Available at: https://doi.org/10.1111/j.1539-6924.1981.tb01347.x (Accessed: 14 September 2023).

Cunningham, A. (2008) 'Digital Curation/Digital Archiving: A View from the National Archives of Australia', *The American Archivist,* 71(2), pp. 530 - 543. Available at: https://www.jstor.org/stable/40294529 (Accessed: 14 September 2023).

Curtis, J. *et al*. (2007) 'AONS—An obsolescence detection and notification service for Web archives and digital repositories', *New Review of Hypermedia and Multimedia,* 13(1), pp. 39 - 53. Available at: https://doi.org/10.1080/13614560701423711 (Accessed: 14 September 2023).

Dallas, C. (2016) 'Digital curation beyond the “wild frontier”: a pragmatic approach', *Archival Science,* Volume 16, pp. 421 - 457. Available at: https://doi.org/10.1007/s10502-015-9252-6 (Accessed: 14 September 2023).

Dappert, A. (2009) *Report on the Conceptual Aspects of Preservation, Based on Policy and Strategy Models for Libraries, Archives and Data Centres.* PLANETS project report. Available at: https://planets-project.eu/docs/reports/Planets_PP2_D3_ReportOnPolicyAndStrategyModelsM36_Ext.pdf (Accessed: 1 September 2023).

Dappert, A. (2013) *DePICT: a conceptual model for digital preservation.* PhD thesis. University of Portsmouth. Available at: https://researchportal.port.ac.uk/files/5899441/Final_Thesis_Angela_Dappert_pdfa.pdf (Accessed 1 September 2023).

Dappert, A. and Farquhar, A. (2009) 'Modelling Organizational Preservation Goals to Guide Digital Preservation', *International Journal of Digital Curation (IJDC),* 4(2), pp. 119 - 134. Available at: https://doi.org/10.2218/ijdc.v4i2.102 (Accessed 14 September 2023).

Dappert, A. and Farquhar, A. (2009) 'Significance is in the Eye of the Stakeholder', *Proceedings of the 13th European Conference on Research and Advanced Technology for Digital Libraries: ECDL09*, Corfu, 27 September - 2 October, pp. 297 - 308. Available at: https://dl.acm.org/doi/10.5555/1812799.1812838 (Accessed 14 September 2023).

Data Seal of Approval (2013) *Data Seal of Approval.* Available at: https://web.archive.org/web/20170518144146/https://assessment.datasealofapproval.org/guidelines_52/html/ (Accessed: 8 September 2023).

Day, M. *et al*. (2014a) 'Identifying Digital Preservation Requirements: Digital Preservation Strategy and Collection Profiling at the British Library', *Proceedings of the 11th International Conference on Digital Preservation: iPRES 2014*, Melbourne, 6 - 10 October, pp. 219 - 227. Available at https://ipres-conference.org/ipres14/sites/default/files/upload/iPres-Proceedings-final.pdf (Accessed: 14 September 2023).

Day, M. *et al*. (2014b) 'Implementing Digital Preservation Strategy: Developing content collection profiles at the British Library', *IEEE/ACM Joint Conference on Digital Libraries: JCDL 2014*, London, 8 - 12 September, pp. 21 - 24. Available at: https://doi.org/10.1109/JCDL.2014.6970145 (Accessed: 14 September 2023).

Day, M. *et al*. (2016) 'The preservation of disk-based content at the British Library: Lessons from the Flashback project', *Alexandria,* 26(3), pp. 216 - 234. Available at https://doi.org/10.1177/0955749016669775 (Accessed: 14 September 2023).

Day, M. *et al*. (2018) 'Preservation Planning for Emerging Formats at the British Library', *Proceedings of the 15th International Conference on Digital Preservation: iPres 2018,* Boston and Cambridge (Massachusetts), 24 - 27 September, session 208.3. Available at: https://doi.org/10.17605/OSF.IO/U5W3Q (Accessed: 14 September 2023).

Day, M. and Pennock, M. (2022) 'Towards a Collections Model for Preservation Planning at the British Library', *Proceedings of the 18th International Conference on Digital Preservation: iPres 2022*, Glasgow, 12 - 16 September, pp. 436 - 437. Available at: http://doi.org/10.7207/ipres2022-proceedings (Accessed: 14 September 2023).

DCC and DPE (2007) *DRAMBORA: Digital Repository Audit Method Based on Risk Assessment.* Glasgow: HATII at University of Glasgow and Digital Curation Centre. Available at: https://www.researchgate.net/publication/31869604_DRAMBORA_The_Digital_Repository_Audit_Method_Based_on_Risk_Assessment (Accessed: 1 September 2023).

Deng, Q. and Ji, S. (2018) 'A Review of Design Science Research in Information Systems: Concept, Process, Outcome, and Evaluation', *Pacific Asia Journal of the Association for Information Systems,* 10(1), pp. 1 - 36. Available at: https://doi.org/10.17705/1pais.10101 (Accessed: 14 September 2023).

di Cosmo, R. and Zacchiroli, S. (2017) 'Software Heritage: Why and How to Preserve Software Source Code', *Proceedings of the 14th International Conference on Digital Preservation: iPRES 2017,* Kyoto, 25 - 29 September. Available at: https://phaidra.univie.ac.at/detail_object/o:340041 (Accessed: 1 September 2023).

Dick, A. L. (1995) 'Library and Information Science as a Social Science: Neutral and Normative Conceptions', *The Library Quarterly: Information, Community, Policy,* 65(2), pp. 216 - 235. Available at: http://www.jstor.org/stable/4309022 (Accessed: 14 September 2023).

Digital Preservation Coalition (2015a) *Digital Preservation Handbook.* Available at: https://www.dpconline.org/handbook/glossary (Accessed: 1 September 2023).

Digital Preservation Coalition (2015b) *Digital Preservation Handbook: Preservation Action.* Available at: https://www.dpconline.org/handbook/organisational-activities/preservation-action (Accessed: 1 September 2023).

Digital Preservation Coalition (2017) *Non-Print Legal Deposit Digital Preservation Review.* London: British Library. Available at: https://www.bl.uk/projects/digital-preservation-capability-assessment (Accessed: 1 September 2023).

Digital Preservation Coalition (2021) *Rapid Assessment Model (RAM).* Available at: http://doi.org/10.7207/dpcram21-02 (Accessed: 1 September 2023).

Dionne, G. (2013) 'Risk Management: History, Definition, and Critique', *Risk Management and Insurance Review,* 16(2), pp. 147 - 166. Available at: https://doi.org/10.1111/rmir.12016 (Accessed: 14 September 2023).

Dobreva, M. and Duff, W. (2015) 'The ever changing face of digital curation: introduction to the special issue on digital curation', *Archival Science,* Volume 15, pp. 97 - 100. Available at: https://doi.org/10.1007/s10502-015-9243-7 (Accessed: 14 September 2023).

Dollar, C. (2016) 'Digital curation beyond the “wild frontier”: a pragmatic approach', *Archival Science,* 16(4), pp. 421 - 457. Available at: https://doi.org/10.1007/s10502-015-9252-6 (Accessed: 14 September 2023).

Dollar, C. M. and Ashley, L. J. (2014) *Assessing Digital Preservation Capability Using a Maturity Model Process Improvement Approach.* Commercial white paper. Available at: https://www.securelyrooted.com/dpcmm (Accessed: 1 September 2023).

Douven, I. (2021) 'Abduction', in: E. N. Zalta, ed. *The Stanford Encyclopedia of Philosophy.* Stanford: Metaphysics Research Lab, Stanford University. Available at: https://plato.stanford.edu/archives/sum2021/entries/abduction/ (Accessed: 14 September 2023).

Dresch, A., Lacerda, D. P. and Antunes Jr, J. A. V. (2015) *Design Science Research: A Method for Science and Technology Advancement.* eBook: Springer. Available at https://doi.org/10.1007/978-3-319-07374-3 (Accessed: 14 September 2023).

Dreschler, A., Gerber, A. and Hevner, A. (2022) *The Transdisciplinary Reach of Design Science Research: Proceedings of the 17th International Conference on Design Science Research in Information Systems and Technology, DESRIST 2022,* St Petersburg, Florida, 1 - 3 June. Springer Cham. Available at: https://doi.org/10.1007/978-3-031-06516-3 (Accessed: 14 September 2023).

Dreschler, A. and Hevner, A. (2022) *Knowledge Paths in Design Science Research.* Delft: now Publishers Inc.

Duit, R. and Glynn, S. (1996) 'Mental Modeling', in: G. Welford, J. Osborne and P. Scott, (eds.) *Research in science education in Europe: current issues and themes.* London: Falmer Press, pp. 166 - 176.

Duranti, L. (1995) 'Reliability and Authenticity: The Concepts and Their Implications', *Archivaria,* Issue 39, pp. 5 - 10. Available at:

https://archivaria.ca/index.php/archivaria/article/view/12063 (Accessed: 14 September 2023).

Ebneyamini, S. (2022) 'Towards Developing a Framework for Conducting Management Studies Using Design Research Methodology', *International Journal of Qualitative Methods,* Volume 21. Available at: https://doi.org/10.1177/16094069221112245 (Accessed: 14 September 2023).

Emblemsvåg, J. and Kjølstad, L. E. (2006) 'Qualitative risk analysis: Some problems and remedies', *Management Decisions,* 44(3), pp. 395-408. Available at: https://doi.org/10.1108/00251740610656278 (Accessed: 14 September 2023).

Erpanet (2003) *Risk Communication Tool.* Erpanet project report. Available at: https://erpanet.org/guidance/docs/ERPANETRiskTool.pdf (Accessed: 1 September 2023).

Fei, D. (2018) 'Abductive Thinking, Conceptualisation, and Design Synthesis', in Ahram, T., Karwowski, W., Taiar, R. (eds) *Human Systems Engineering and Design. IHSED 2018. Advances in Intelligent Systems and Computing*, vol 876. Springer, Cham. Available at: https://doi.org/10.1007/978-3-030-02053-8_16 (Accessed: 14 September 2023).

Ferreira, F., Vieria, R. and Borbinha, J. (2014) 'The Value of Risk Management for Data Management in Science and Engineering', *EEE/ACM Joint Conference on Digital Libraries: JCDL 2014*, London, 8 - 12 September, pp. 439 - 440. Available at: https://doi.org/ 10.1109/JCDL.2014.6970214 (Accessed: 14 September 2023).

Fettke, P. (2009) 'How Conceptual Modeling is Used', *Communications of the Association for Information Systems,* Volume 25, pp. 571 - 593. Available at: https://doi.org/10.17705/1CAIS.02543 (Accessed: 14 September 2023).

Fettke, P., Loos, P. and Zwicker, J. (2005) 'Business Process Reference Models: Survey and Classification', in: Bussler, C.J., Haller, A. (eds) *Business Process Management Workshops: BPM 2005. Lecture Notes in Computer Science,* vol 3812. Springer, Berlin, Heidelberg. Available at: https://doi.org/10.1007/11678564_44 (Accessed: 14 September 2023).

Feyerabend, P. (1970) 'Against Method: Outline of an Anarchistic Theory of Knowledge', in In M. Radner and S. Winokur (eds.), *Analyses of theories and methods of physics and psychology: Minnesota studies in the philosophy of science,* NED-New edition, vol. 4, University of Minnesota Press, 1970, pp. 17–130. Available at: https://www.jstor.org/stable/10.5749/j.ctttsvns.4 (Accessed: 14 September 2023).

Fischer, C. and Gregor, S. (2011) 'Forms of Reasoning in the Design Science Research Process', *Proceedings of the 6th International Conference, DESRIST 2011*, Milwaukee, USA, May 5-6, pp. 17 - 31. Available at: https://doi.org/10.1007/978-3-642-20633-7 (Accessed 15 September 2023).

Frank, R. (2018) *The Social Construction of Risk in Trustworthy Digital Repository Audit and Certification.* PhD thesis. University of Michigan. Available at: https://hdl.handle.net/2027.42/147539 (Accessed 1 September 2023).

Frank, R. D. (2020) 'The Social Construction of Risk in Digital Preservation', *Journal of the Association for Information Science and Technology,* 71(4), pp. 474 - 484. Available at: https://doi.org/10.1002/asi.24247 (Accessed 15 September 2023).

Frank, R. D. (2022) 'Risk in trustworthy digital repository audit and certification', *Archival Science,* Volume 22, pp. 43 - 73. Available at: https://doi.org/10.1007/s10502-021-09366-z (Accessed 15 September 2023).

Frank, R. D. and Rothfritz, L. (2023) 'Designated Community: Uncertainty and Risk', *Journal of Documentation,* 79(4), pp. 880 - 897. Available at: https://doi.org/10.1108/JD-07-2022-0161 (Accessed 15 September 2023).

Frank, U. (1999) 'Conceptual Modelling as the Core of the Information Systems Discipline - Perspectives and Epistemological Challenges', *AMCIS 1999 Proceedings*, Milwaukee, Wisconsin, August 13 - 15. Available at: https://aisel.aisnet.org/amcis1999/240 (Accessed 15 September 2023).

Friedman, L., Friedman, H. and Pollack, S. (2008) 'The Role of Modeling in Scientific Disciplines: A Taxonomy', *Review of Business,* 29(1), pp. 61 - 67.Available at: https://ssrn.com/abstract=2322506 (Accessed 15 September 2023).

Friedrich, M. (2019) 'Preferred, Obsolete Or In-Between? Developing A Criteria Catalogue For AV-Material: Preservation planning at the German National Library of Science and Technology (tib)', *Proceedings of the 16th International Conference on Digital Preservation: iPRES 2019*, Amsterdam, 16 - 20 September, pp. 399 - 400. Available at: https://ipres2019.org/static/proceedings/iPRES2019.pdf (Accessed 15 September 2023).

Garrett, J. and Waters, D. (1996) *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information.* CPA and RLG commissioned report. Available at: https://www.clir.org/wp-content/uploads/sites/6/pub63watersgarrett.pdf (Accessed: 1 September 2023).

Gattuso, J. (2014) 'Converting Wordstar to HTML4', *Proceedings of the 11th International Conference on Digital Preservation: iPRES 2014*, Melbourne, 6 - 10 October, pp. 149 - 159. Available at: https://ipres-conference.org/ipres14/sites/default/files/upload/iPres-Proceedings-final.pdf (Accessed: 15 September 2023).

Gavrel, K. (1990) *Conceptual problems posed by electronic records: A RAMP study,* Paris: UNESCO.

Giaretta, D. *et al*. (2009) 'Significant Properties, Authenticity, Provenance, Representation', *Proceedings of the Sixth International Conference on Preservation*

of Digital Objects: iPRES 2009, San Francisco, 5 - 6 October. pp. 67 - 73. Available at: https://escholarship.org/uc/item/0wf3j9cw (Accessed 15 September 2023).

Giaretta, D. *et al*. (2019) 'Dawn of Digital Repositories Certification under ISO 16363. Exploring the Horizon and beyond', *Proceedings of the 16th International Conference on Digital Preservation: iPRES 2019*, Amsterdam, 16 - 20 September, pp. 463-465. Available at: https://ipres2019.org/static/proceedings/iPRES2019.pdf (Accessed 15 September 2023).

Giere, R. (2001) 'The nature and function of models', *Behavioural and Brain Sciences,* Volume 24, p. 1060. Available at: https://doi.org/10.1017/S0140525X01320125 (Accessed 15 September 2023).

Gilani, W. *et al*. (2013) 'TIMBUS: Digital Preservation for Timeless Business processes and Services'*, eChallenges 2013 Conference Proceedings,* Dublin, 9 - 11 October. Available at: https://www.researchgate.net/publication/259570416_TIMBUS_Digital_Preservation_for_Timeless_Business_Processes_And_Services (Accessed: 1 September 2023).

Glassey-Previdoli, D., Bonazzi, R. and Viscusi, G. (2021) 'Business Model Design and Ecosystem Innovation: A Method for Visualizing Interactions', *The Next Wave of Sociotechnical Design: 16th International Conference on Design Science Research in Information Systems and Technology, DESRIST 2021.* Kristiansand, Norway, August 4 – 6. Lecture Notes in Computer Science, vol 12807. Springer, Cham. pp. 387 - 398. Available at: https://doi.org/10.1007/978-3-030-82405-1 (Accessed: 15 September 2023).

Gnewuch, U. and Maedche, A. (2022) 'Toward a Method for Reviewing Software Artifacts from Practice', *The Transdisciplinary Reach of Design Science Research: 17th International Conference on Design Science Research in Information Systems and Technology, DESRIST 2022*. St Petersburg, FL, USA, June 1 – 3. Lecture Notes in Computer Science, vol 13229. Springer, Cham. pp. 337 - 350. https://doi.org/10.1007/978-3-031-06516-3 (Accessed: 15 September 2023).

Goldkuhl, G. (2012) 'Design Science in Search for a Paradigm: Pragmatism Is the Answer', in Helfert, M., Donnellan, B. (eds) *Practical Aspects of Design Science: EDSS 2011*. Communications in Computer and Information Science, vol 286, pp. 84-95. Springer, Berlin, Heidelberg. Available at: https://doi.org/10.1007/978-3-642-33681-2_8 (Accessed: 15 September 2023).

Goldkuhl, G. (2020) Design Science Epistemology. A pragmatist inquiry. *Scandinavian Journal of Information Systems,* 32(1). Available at: https://aisel.aisnet.org/sjis/vol32/iss1/2/ (Accessed: 15 September 2023).

Gollins, T. (2009) *Parsimonious preservation: preventing pointless processes!* London: The National Archives (UK). Available at: https://cdn.nationalarchives.gov.uk/documents/parsimonious-preservation.pdf (Accessed: 15 September 2023).

Graf, R. and Gordea, S. (2013) 'A Risk Analysis of File Formats for Preservation Planning', *Proceedings of the 10th International Conference on Preservation of Digital Objects: iPRES 2013*, Lisbon, 2 - 6 September, pp. 177 - 186. Available at: https://purl.pt/24107 (Accessed: 15 September 2023).

Graf, R., Gordea, S. and Ryan, H. (2014) 'A Model for Format Endangerment Analysis using Fuzzy Logic', *Proceedings of the 11th international conference on digital preservation: iPRES 2014*, Melbourne, 6 - 10 October, pp. 160-168. Available at: https://ipres-conference.org/ipres14/sites/default/files/upload/iPres-Proceedings-final.pdf (Accessed: 15 September 2023).

Granger, S. (2000) 'Emulation as a Digital Preservation Strategy', *D-Lib Magazine*, 6(10). Available at: https://doi.org/10.1045/october2000-granger (Accessed: 15 September 2023).

Gray, J. and Rumpe, B. (2021) 'Reference models: How can we leverage them?', *Software and Systems Modeling,* Volume 20, pp. 1775–1776. Available at: https://doi.org/10.1007/s10270-021-00948-0 (Accessed: 15 September 2023).

Gregor, S. (2006) 'The Nature of Theory in Information Systems', *MIS Quarterly,* 30(3), pp. 611-642. Available at: https://doi.org/10.2307/25148742 (Accessed: 15 September 2023).

Gregor, S. and Hevner, A. (2013) 'Positioning and Presenting Design Science Research for Maximum Impact', *MIS Quarterly,* 37(2), pp. 337 - 355. Available at: https://www.jstor.org/stable/43825912 (Accessed: 15 September 2023).

Grindley, N. (2013) *The Digital Curation Sustainability Model (DCSM).* 4C project report. Available at: https://www.4cproject.eu/documents/DCSM-V1.01-18Mar2015.pdf (Accessed: 1 September 2023).

Hansson, S. O. (2010) 'Risk: objective or subjective, facts or values', *Journal of Risk Research,* 13(2), pp. 231-238. Available at: https://doi.org/10.1080/13669870903126226 (Accessed: 15 September 2023).

Hansson, S. O. (2012) 'A Panorama of the Philosophy', in S. Roeser, R. Hillerbrand, P. Sandin and M. Peterson (eds). *Handbook of Risk Theory.* Dordrecht: Springer, pp. 28 - 51. Available at: https://doi.org/10.1007/978-94-007-1433-5_2 (Accessed: 15 September 2023).

Hansson, S. O. (2013) 'Defining Pseudoscience and Science', in M. Pigliucci and M. Boudry (eds). *Philosophy of Science: Reconsidering the Demarcation Problem.* Chicago: University of Chicago Press, pp. 61 - 78.

Hansson, S. O. and Aven, T. (2014) 'Is Risk Analysis Scientific?', *Risk Analysis,* 34(7), pp. 1173 - 1183. Available at: https://doi.org/10.1111/risa.12230 (Accessed: 15 September 2023).

Hao, J., Li, J. and Wu, D. (2023) 'Bibliometric analysis of risk science from 1996 to 2021: insights and implications', *Journal of Risk Research,* 26(5), pp. 485 - 501. Available at: https://doi.org/10.1080/13669877.2023.2176914 (Accessed: 15 September 2023).

Hay-Gibson, N. V. (2008) *A River of Risk: A Diagram of The History and Historiography of Risk Management*. Northumbria University Library working paper. Available at: https://core.ac.uk/download/pdf/4147339.pdf (Accessed: 15 September 2023).

Heckmann, I., Comes, T. and Nickel, S. (2015) 'A critical review on supply chain risk – Definition, measure and modeling', *Omega,* Volume 52, pp. 119-132. Available at: https://doi.org/10.1016/j.omega.2014.10.004 (Accessed: 15 September 2023).

Hedstrom, M. (1984) *Archives and Manuscripts: Machine Readable Records,* Chicago: Society of American Archivists.

Hedstrom, M. (1991) 'Understanding Electronic Incunabula: A Framework for Research on Electronic Records', *American Archivist,* 54(3), pp. 334-354.

Hedstrom, M. (1995) 'Electronic Archives: Integrity and Access in the Network Environment', *American Archivist,* Summer, Volume 58, pp. 312 - 324. Available at: https://www.jstor.org/stable/40293923 (Accessed: 15 September 2023).

Hedstrom, M. (1997/98) 'Digital Preservation: A Time Bomb for Digital Libraries', *Computers and the Humanities,* 31(3), pp. 189 - 202. Available at: https://www.jstor.org/stable/30200423 (Accessed: 15 September 2023).

Hedstrom, M. (2002) 'The Digital Preservation Research Agenda', *The State of Digital Preservation: An International Perspective - Conference Proceedings*, Washington D.C., 24 - 25 April, pp. 32 - 37. Available at: https://www.clir.org/wp-content/uploads/sites/6/pub107.pdf (Accessed: 15 September 2023).

Hedstrom, M. (2003) *It's About Time: Research Challenges in Digital Archiving and Long-term Preservation.* National Science Foundation and Library of Congress workshop report. Available at: https://chnm.gmu.edu/digitalhistory/links/pdf/preserving/8_4b.pdf (Accessed: 1 September 2023).

Hedstrom, M. and Lee, C. (2002) 'Significant properties of digital objects: definitions, applications, implications', *Proceedings of the DLM-Forum, 2002,* Barcelona, 6 - 8 May. Available at: https://ils.unc.edu/callee/sigprops_dlm2002.pdf (Accessed: 14 September 2023).

Hein, S. and Schmitt, K. (2013) 'Risk Management for Digital Long-Term Preservation Services', *Proceedings of the 10th International Conference on Preservation of Digital Objects: iPRES 2013*, Lisbon, 2 - 6 September, pp. 314 - 317. Available at: https://purl.pt/24107 (Accessed: 15 September 2023).

Heinze, D. (1982) *Managing Science: Introductory Concepts and Applications.* 2 ed. Cincinnati, Ohio: South Western Publishing.

Hemming, V. *et al*. (2018) 'A practical guide to structured expert elicitation using the IDEA protocol', *Methods in Ecology and Evolution,* Issue 9, pp. 169 - 180. Available at: https://doi.org/10.1111/2041-210X.12857 (Accessed: 15 September 2023).

Heslop, H., Davis, S. and Wilson, A. (2002) *An Approach to the Preservation of Digital Records.* National Archives of Australia working paper. Available at: https://www.naa.gov.au/sites/default/files/2020-01/An-Approach-to-the-Preservation-of-Digital-Records.pdf (Accessed: 15 September 2023).

Hesse, M. (1976) 'Models versus paradigms in the natural sciences', in Collins, Lyndhurst, (ed.) *The Use of Models in the Social Sciences.* London: Westview Press, pp. 1 - 15.

Hevner, A. R. *et al*. (2004) 'Design Science in Information Systems Research', *MIS Quarterly,* 29(1), pp. 75 - 105. Available at: https://doi.org/10.2307/25148625 (Accessed: 15 September 2023).

Hevner, A. and Chatterjee, S. (2010) *Design Research in Information Systems: Theory and Practice.* New York: Springer. Available at: https://doi.org/10.1007/978-1-4419-5653-8 (Accessed: 15 September 2023).

Higgins, S. (2008) 'The DCC Curation Lifecycle Model', *The International Journal of Digital Curation, IJDC,* 3(1), pp.134 - 140. Available at: https://doi.org/10.2218/ijdc.v3i1.48 (Accessed: 15 September 2023).

Higgins, S. (2018) 'Digital Curation: The development of a discipline within information science', *Journal of Documentation,* 74(6), pp. 1318 - 1338. Available at: https://doi.org/10.1108/JD-02-2018-0024 (Accessed: 15 September 2023).

Hockx-Yu, H., Johnson, S. and Pennock, M. (2012) *Twittervane: Crowdsourcing for web archiving.* Washington DC: UNT Digital Library, presentation at IIPC 2012. Available at: https://digital.library.unt.edu/ark:/67531/metadc1457749/ (Accessed: 15 September 2023).

Hofman, J. (1999) Shooting at a moving target'. The development of a repository for the preservation of digital information. *Proceedings of the DLM Forum on Electronic records*, Volume Supplement IV, pp. 80 - 87. Available at: http://www.interpares.org/display_file.cfm?doc=ip1_dissemination_cpr_hofman_dlm_1999.pdf (Accessed: 15 September 2023).

Hole, B. *et al*, (2010) 'LIFE3: A predictive costing tool for digital collections', *Proceedings of the 7th International Conference on the Preservation of Digital Objects: iPRES 2010*, Vienna, 19 - 24 September, pp. 359 - 364. Available at: https://phaidra.univie.ac.at/detail/o:245912 (Accessed: 15 September 2023).

Hunter, J. and Choudhury, S. (2006) 'PANIC: an integrated approach to the preservation of composite digital objects using Semantic Web services', *International Journal on Digital Libraries,* 6(2), pp. 174-183. Available at: https://doi.org/10.1007/s00799-005-0134-z (Accessed: 15 September 2023).

Hutten, E. H. (1954) 'The Role of Models in Physics', *The British Journal for the Philosophy of Science,* 4(16), pp. 284 - 301. Available at: https://www.jstor.org/stable/686078 (Accessed: 15 September 2023).

Iivari, J. (2007) 'A Paradigmatic Analysis of Information Systems As a Design Science', *Scandinavian Journal of Information Systems,* 19(2), pp. 39 - 64. Available at: http://aisel.aisnet.org/sjis/vol19/iss2/5 (Accessed: 15 September 2023).

Iivari, J. (2020) 'Editorial: A Critical Look at Theories in Design Science Research', *Journal of the Association for Information Systems (JAIS),* 21(3), pp. 502 - 519. Available at: https://doi.org/10.17705/1jais.00610 (Accessed: 15 September 2023).

Innocenti, P. (2012) 'Bridging the gap in digital art preservation: interdisciplinary reflections on authenticity, longevity and potential collaboration', in Konstantelos, L., Delve, J., Billeness, C., Baker, D. and Dobreva, M. (eds.) *Preservation of Complex Objects: Volume 2, Software Art*. Series: The preservation of complex objects. JISC pp. 71 - 83. Available at: https://www.cdpa.co.uk/POCOS/books/pocos_vol_2.pdf (Accessed: 15 September 2023).

ISO (2009) *Guide 73: Risk Management - Vocabulary,* Geneva: International Standards Organisation.

ISO (2012a) *ISO 14721: A Reference Model for an Open Archival Information System,* Geneva: International Standards Organisation.

ISO (2012b) *ISO 16363: Space data and information transfer systems — Audit and certification of trustworthy digital repositories,* Geneva: International Standards Organisation.

ISO (2018) *ISO 31000 Risk Management - Guidelines,* Geneva: International Standards Organisation.

ISO (2019) *ISO 31010:2019 Risk management - risk assessment techniques,* Geneva: International Standards Organisation.

Jackson, A. (2012) 'Formats over Time: Exploring UK Web History', *Proceedings of the Ninth International Conference on Preservation of Digital Objects: iPRES 2012*, Toronto, 1 - 5 October, pp. 155 - 159. Available at: https://phaidra.univie.ac.at/detail/o:293834 (Accessed: 15 September 2023).

Jensen, R. C. and Hansen, H. (2020) 'Selecting Appropriate Words for Naming the Rows and Columns of Risk Assessment Matrices', *International Journal of Environmental Research and Public Health,* 17(15). Available at: https://doi.org/10.3390/ijerph17155521 (Accessed: 15 September 2023).

Johnston, L. (2018) 'Creating a holdings format profile and format matrix for risk-based digital preservation planning at the National Archives and Records Administration', *Proceedings of the fifteenth international conference on digital preservation: iPres 2018*, Boston and Cambridge (Massachusetts), 24 - 27 September, session 208.1. Available at: https://doi.org/10.17605/OSF.IO/CTW3G (Accessed: 15 September 2023).

Johnston, L. (2020) 'Challenges in preservation and archiving digital materials', *Information Services and Use,* 40(3), pp. 193 - 199. Available at: https://doi.org/10.3233/ISU-200090 (Accessed: 15 September 2023).

Kasanen, E., Lukka, K. and Siitonen, A. (1993) 'The Constructive Approach in Management Accounting Research', *Journal of Management Accounting Research,* Volume 5, pp. 243 - 264.

Kasperson, R. E. *et al*. (1988) 'The Social Amplification of Risk: A Conceptual Framework', *Risk Analysis,* 8(2), pp. 177 - 187. Available at: https://doi.org/10.1111/j.1539-6924.1988.tb01168.x (Accessed: 15 September 2023).

Kenney, A. R. and Personius, L. K. (1992) 'The Future of Digital Preservation', in B. Buckner Higginbotham and M. E. Jackson, eds. *Advances in Preservation and Access.* Westport: Meckler, pp. 195 - 212.

Kenney, A. R. *et al*. (2002) 'Preservation Risk Management for Web Resources: Virtual Remote Control in Cornell's Project Prism', *D-Lib Magazine,* 8(1). Available at: https://doi.org/10.1045/january2002-kenney (Accessed: 15 September 2023).

Kenney, A. R. and McGovern, N. Y. (2003) 'The Five Organisational Stages of Digital Preservation', in *Digital Libraries: A Vision for the 21st Century: A Festschrift in Honor of Wendy Lougee on the Occasion of her Departure from the University of Michigan.* Ann Arbor, MI: Michigan Publishing, University of Michigan Library. Available at: http://dx.doi.org/10.3998/spobooks.bbv9812.0001.001 (Accessed: 15 September 2023).

Klinke, A. and Renn, O. (2002) 'A New Approach to Risk Evaluation and Management: Risk-Based, Precaution-Based, and Discourse-Based Strategies', *Risk Analysis,* 22(6), pp. 1071 - 1094. Available at: https://doi.org/10.1111/1539-6924.00274 (Accessed: 15 September 2023).

Knight, G. and Pennock, M. (2009) 'Data Without Meaning: Establishing the Significant Properties of Digital Research', *International Journal of Digital Curation (IJDC),* 4(1), pp.159 - 174. Available at: https://doi.org/10.2218/ijdc.v4i1.86 (Accessed: 15 September 2023).

Knutas, A., Pourzolfaghar, Z. and Helfert, M. (2019) 'The Role and Impact of Descriptive Theories in Creating Knowledge in Design Science', in Holzinger, A., Silva, H., Helfert, M. (eds) *Computer-Human Interaction Research and Applications.*

CHIRA 2017. Communications in Computer and Information Science, vol 654, pp. 90 - 108. Springer, Cham. https://doi.org/10.1007/978-3-030-32965-5_5 (Accessed: 15 September 2023).

Kramer-Smyth, J., Gkremo, T. and Thompson, S. (2023) *Creating Digital Preservation Plans: Leveraging Expertise Across Your Organisation*. Submission to the 19th international conference on digital preservation: iPres 2023. Available at: https://ipres2023.us/timetable/event/tp-2/ (Accessed: 15 September 2023).

Krewski, D. *et al*. (2014) A framework for the next generation of risk science. *Environmental Health Perspectives,* 122(8), pp. 796-805. Available at: https://doi.org/10.1289/ehp.1307260 (Accessed: 15 September 2023).

Krisper, M. (2021) *Problems with Risk Matrices Using Ordinal Scales.* Graz University of Technology research paper. Available at: https://arxiv.org/abs/2103.05440 (Accessed: 15 September 2023).

Kuechler, W. and Vaishnavi, V. (2012) 'A Framework for Theory Development in Design Science Research: Multiple Perspectives', *Journal of the Association for Information Systems JAIS,* 13(6), pp. 395-423. Available at: https://doi.org/10.17705/1jais.00300 (Accessed: 15 September 2023).

Kuhn, T. (1962) *The Structure of Scientific Revolutions.* Chicago: The University of Chicago Press.

L'Hours, H. *et al*. (2014) *Cost Concept Model and Gateway Specification.* 4C project report. Available at: https://www.4cproject.eu/documents/D3.2%20Cost%20Concept%20Model%20and%20Gateway%20Specification.pdf (Accessed: 1 September 2023).

Langley, S. (2019) 'Digital Preservation Should Be More Holistic: A Digital Stewardship Approach', in *Digital Preservation in Libraries: Preparing for a Sustainable Future.* Chicago: ALA Editions, pp. 93 - 128.

Lavoie, B. F. (2008) 'The Fifth Blackbird: Some Thoughts on Economically Sustainable Digital Preservation', *D-Lib Magazine*, 14(3/4). Available at: https://doi.org/10.1045/march2008-lavoie (Accessed: 15 September 2023).

Lavoie, B. (2014) *The Open Archival Information System (OAIS) Reference Model: introductory guide.* Digital Preservation Coalition Technology Watch report. Available at: https://doi.org/10.7207/twr14-02 (Accessed: 15 September 2023).

Lavoie, B. and Dempsey, L. (2004) 'Thirteen Ways of Looking at...Digital Preservation', *D-Lib Magazine*, 10 (7/8). Available at: https://doi.org/10.1045/july2004-lavoie (Accessed: 15 September 2023).

Lawrence, G. W. *et al*. (2000) *Risk Management of Digital Information: A File Format Investigation,* Washington D.C.: Council on Library and Information

Resources. Available at: https://www.clir.org/wp-content/uploads/sites/6/pub93.pdf (Accessed: 15 September 2023).

Lazorchak, B. (2011) 'Digital Preservation, Digital Curation, Digital Stewardship: What's in (Some) Names?', *Library of Congress blog: The Signal,* 23 August Available at: https://blogs.loc.gov/thesignal/2011/08/digital-preservation-digital-curation-digital-stewardship-what%E2%80%99s-in-some-names/ (Accessed: 15 September 2023).

Lee, C. (2005) *Defining Digital Preservation Work: A Case Study of the Development of the Reference Model for an Open Archival Information System.* PhD thesis. University of Michigan.
Available at: https://deepblue.lib.umich.edu/handle/2027.42/39372
(Accessed: 1 September 2023).

Lee, J. S., Pries-Heje, J. and Baskerville, R. (2011) 'Theorizing in Design Science Research', *Proceedings of the 6th International Conference, DESRIST 2011,* Milwaukee, USA, May 5-6, pp. 1 - 16. Available at: https://doi.org/10.1007/978-3-642-20633-7 (Accessed 15 September 2023).

Leggett, E. (2021) *Digitization and Digital Archiving: A Practical Guide for Librarians.* 2nd ed. London: The Rowman and Littlefield Publishing Group, Inc.

LIFE Project (2010) *The LIFE3 project Final report.* LIFE3 project report. Available at: http://www.life.ac.uk/3/docs/life3_report.pdf
(Accessed: 1 September 2023).

Louis, A. (2008) 'What's Wrong with Risk Matrices?', *Risk Analysis,* 28(2), pp. 497 - 512. Available at: https://doi.org/10.1111/j.1539-6924.2008.01030.x (Accessed: 15 September 2023).

MacKrell, D., McDonald, C. and Gammack, J. (2017) 'An Information Systems PhD by Artefact and Exegesis?', *ACIS 2017 Proceedings,* Hobart, 5 - 6 December. Available at: https://aisel.aisnet.org/acis2017/74 (Accessed: 15 September 2023).

Maemura, E., Moles, N. and Becker, C. (2017) 'Organisational Assessment Frameworks for Digital Preservation: A Literature Review and Mapping', *Journal of the Association for Information Science and Technology,* 68(7), pp. 1619 - 1637. Available at: https://doi.org/10.1002/asi.23807 (Accessed: 15 September 2023).

March, S. T. and Smith, G. F. (1995) 'Design and natural science research on information technology', *Decision Support Systems,* Volume 15, pp. 251-266. Available at: https://doi.org/10.1016/0167-9236(94)00041-2 (Accessed: 15 September 2023).

Martens, F. and Rittenberg, L. (2020) *Risk Appetite: Critical to Success.* Committee of Sponsoring Organizations of the Treadway Commission *(*COSO) commissioned report. Available at:

https://www.worldcomplianceassociation.com/noticias/noticia_doc_wca__seg_220520_243166.pdf (Accessed: 15 September 2023).

Matthews, B. *et al*. (2009) 'Towards a Methodology for Software Preservation', *Proceedings of the Sixth International Conference on Preservation of Digital Objects: iPRES 2009*, San Francisco, 5 - 6 October. pp. 132 - 140. Available at: https://escholarship.org/uc/item/8089m1v1 (Accessed 15 September 2023).

May, P., Pennock, M. and Russo, D. (2019) 'The Integrated Preservation Suite: Scaled and automated preservation planning for highly diverse digital collections', *Proceedings of the 16th international conference on digital preservation: iPRES 2019*, Amsterdam, 16 - 20 September, pp. 142 - 154. Available at: https://ipres2019.org/static/proceedings/iPRES2019.pdf (Accessed 15 September 2023).

McGath, G. (2013) 'The Format Registry Problem', *Code4Lib,* Issue 19. Available at: https://journal.code4lib.org/articles/8029 (Accessed 15 September 2023).

McGovern, N. (2007) 'A Digital Decade: Where have we Been, and Where are we Going in Digital Preservation?', *RLG Diginews*, 11(1). Available at: http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070519/viewer/file2488.html#v10_n1_art2_fn1 (Accessed 15 September 2023).

McGovern, N. (2009) *Technology Responsiveness for Digital Preservation: A Model.* PhD thesis. University College London. Available at: https://discovery.ucl.ac.uk/id/eprint/18017/ (Accessed 1 September 2023).

McGovern, N. Y. *et al*. (2004) 'Virtual Remote Control: Building a Preservation Risk Management Toolbox for Web Resources', *D-Lib Magazine,* 10(4). Available at: https://doi.org/10.1045/april2004-mcgovern (Accessed 1 September 2023).

McGovern, N. Y. (2022), 'Digital Archive and Preservation Framework', P. C. Franks (ed), *The Handbook of Archival Practice*. Lanham, Maryland: Rowman & Littlefield, pp 220 - 225.

McHugh, A. (2016) *An ontology for risk management of digital collections.* PhD thesis. University of Glasgow. Available at: http://theses.gla.ac.uk/7757/ (Accessed 1 September 2023).

McLeod, R. (2008) 'Risk Assessment; using a risk based approach to prioritise handheld digital information', *Proceedings of the Fifth International Conference on Preservation of Digital Objects: iPRES 2008*, London, 29 - 30 September, pp. 127 - 133. Available at: https://www.bl.uk/ipres2008/ipres2008-proceedings.pdf (Accessed: 1 September 2023).

Merwood, H. (2020) 'Risk Alert: Insufficient Technical Metadata', *DPC blog,* 20 May. Available at: https://www.dpconline.org/blog/risk-alert-insufficient-technical-metadata (Accessed: 1 September 2023).

Mikelakis, M. and Papatheodorou, C. (2012) 'An ontology-based model for preservation workflows', *Proceedings of the Ninth International Conference on Preservation of Digital Objects: iPRES 2012*, Toronto, 1 - 5 October, pp. 21 - 28. Available at: https://phaidra.univie.ac.at/o:293677 (Accessed: 15 September 2023).

Milech, B. and Schilo, A. (2004) ''Exit Jesus': Relating the Exegesis and Creative/Production Components of a Research Thesis', *TEXT Journal of Writing and Writing Courses,* April, Issue 3. Available at: https://www.textjournal.com.au/speciss/issue3/milechschilo.htm (Accessed: 15 September 2023).

Mitcham, J., Currie, A. and Kilbride, W. (2021) *Safeguarding the Nation's Digital Memory Project: Evaluation Report for the National Archives.* Glasgow: DPC. Available at: http://doi.org/10.7207/op21-01 (Accessed: 1 September 2023).

Moeller, R. R. (2011) *COSO Enterprise Risk Management: establishing effective governance, risk, and compliance processes.* Chichester: Wiley. Available at: https://doi.org/10.1002/9781118269145 (Accessed: 1 September 2023).

Moulaison Sandy, H. and Corrado, E. M. (2018) 'Bringing Content into the Picture: Proposing a Tri-Partite Model for Digital Preservation', *Journal of Library Administration,* 58(1), pp. 1 - 17. Available at: https://doi.org/10.1080/01930826.2017.1385988 (Accessed: 1 September 2023).

NDSA (2022) *NDSA Levels of Digital Preservation.* Available at: https://ndsa.org/publications/levels-of-digital-preservation/ (Accessed: 1 September 2023).

nestor (2013) *Explanatory notes on the nestor seal for Trustworthy Digital Archives.* nestor working group materials 17. Available at: https://d-nb.info/1047613859/34 (Accessed: 1 September 2023).

Task Force on Digital Repository Certification (2007) *Trustworthy Repositories Audit and Certification: Criteria and Checklist.* OCLC/RLG report. Available at: https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf (Accessed: 1 September 2023).

Olivé, A. (2007) *Conceptual Modeling of Information Systems.* Berlin: Springer. Available at: https://doi.org/10.1007/978-3-540-39390-0 (Accessed: 1 September 2023).

OASIS Open (2020) *OASIS SOA Reference Model (SOA-RM) TC FAQ.* Guidance from the OASIS Service Oriented Architecture Reference Model Techical Committee. Available at: https://www.oasis-open.org/committees/soa-rm/faq.php (Accessed 1 September 2023).

Ostrom, L. T. and Wilhelmsen, C. A. (2019) *Risk Assessment: Tools, Techniques, and Their Applications.* 2 ed. Hoboken: Wiley.

Owens, T. (2018) *The Theory and Craft of Digital Preservation.* Baltimore: John Hopkins University Press.

Paté-Cornell, M. (1996) 'Uncertainties in risk analysis: Six levels of treatment', *Reliability Engineering and System Safety,* 54(2/3), pp. 95 - 111. Available at: https://doi.org/10.1016/S0951-8320(96)00067-1 (Accessed 1 September 2023).

Pearson, D. and Walker, M. (2007) *Report of the Format Notification and Obsolescence Service (AONS II).* Australian Partnership for Sustainable Repositories: AONS II project report. Available at: https://openresearch-repository.anu.edu.au/bitstream/1885/46645/4/report_aonsII.pdf (Accessed: 1 September 2023).

Pearson, D. and Webb, C. (2008) 'Defining File Format Obsolescence: A Risky Journey', *International Journal of Digital Curation (IJDC),* 3(1), pp.89 - 106. Available at: DOI: https://doi.org/10.2218/ijdc.v3i1.44 (Accessed: 1 September 2023).

Peffers, K. *et al*. (2006) 'The Design Science Research Process: A Model for Producing and Presenting Information Systems Research*', Proceedings of the First Design Research Information Systems and Technology Conference: DESRIST 2006,* Claremont, CA, 24 - 25 February, pp. 83-106. Available at: https://www.researchgate.net/publication/228650671 (Accessed: 1 September 2023).

Pennock, M. (2006a) 'Collaboration as the keystone for successful management of digital records', *ICA-SUV Seminar: Shared Concerns and Responsibility for University Archives*. Reykjavik, Iceland, September. Available at https://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/ICA-SUV-2006_mpennock_paper.pdf (Accessed: 1 September 2023).

Pennock, M. (2006b) 'Curating Emails: A life-cycle approach to the management and preservation of e-mail messages', in S.Ross, M.Day (eds), *DCC Digital Curation Manual*. Glasgow: Digital Curation Centre. Available at: https://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/curating-e-mails/curating-e-mails.pdf (Accessed: 1 September 2023).

Pennock, M. (2006c) *Digital Preservation - Continued access to authentic digital assets.* Jisc briefing paper. Available at: https://www.webarchive.org.uk/wayback/archive/20140614202005/http://www.jisc.ac.uk/publications/briefingpapers/2006/pub_digipreservationbp.aspx (Accessed: 15 September 2023).

Pennock, M. (2007) *Digital Curation: A Life-Cycle Approach to Managing and Preserving Usable Digital Information.* Prepublication article for *Library and Archives Journal*, Issue 1. Available at: https://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf (Accessed: 15 September 2023).

Pennock, M. (2008) *JISC Programme Synthesis Study: Supporting Digital Preservation and Asset Management in Institutions - Review part II, programme Synthesis.* Jisc programme evaluation report. Available at: https://www.webarchive.org.uk/wayback/archive/20140613220103/http://www.jisc.ac.uk/media/documents/programmes/preservation/404publicreport_2008.pdf (Accessed: 1 September 2023).

Pennock, M. (2013) *Web Archiving: A DPC Technology Watch Report.* Digital Preservation Coalition Technology Watch report. Available at: http://dx.doi.org/10.7207/twr13-01 (Accessed: 1 September 2023).

Pennock, M. (2018) *Digital Preservation Staffing and Skills,* Presentation to the DPC Memory Makers training event, Amsterdam, 29 - 30 October. Available online at: https://doi.org/10.5281/zenodo.8351313 (Accessed: 16 September 2023).

Pennock, M. (2020) 'So, have you got a Preservation Plan for that?', *OPF blog,* 27 July. Available at: https://openpreservation.org/blogs/so-have-you-got-a-preservation-plan-for-that/ (Accessed: 15 September 2023).

Pennock, M. and Kelly, B. (2006) 'Archiving web site resources: a records management view', *Proceedings of the 15th international conference on the World Wide Web: WWW06*, Edinburgh, 23 - 26 May, pp. 987-988. Available at: https://doi.org/10.1145/1135777.1135978 (Accessed: 15 September 2023).

Pennock, M. and Davis, R. (2009) 'ArchivePress: A Really Simple Solution to Archiving Blog Content', *Proceedings of the Sixth International Conference on Preservation of Digital Objects: iPRES 2009*, San Francisco, 5 - 6 October, pp. 148 - 154. Available at: https://escholarship.org/uc/item/7zs156mb (Accessed 15 September 2023).

Pennock, M., Jackson, A. N. and Wheatley, P. (2012) 'CRISP: Crowdsourcing Representation Information to Support Preservation', *Proceedings of the Ninth International Conference on Preservation of Digital Objects: iPRES 2012*, Toronto, 1 - 5 October, pp. 17 - 20. Available at: https://phaidra.univie.ac.at/o:293676 (Accessed: 15 September 2023).

Pennock, M., Wheatley, P. and May, P. (2014) 'Sustainability Assessments at the British Library: Formats, Frameworks and Findings', *Proceedings of the 11th international conference on digital preservation: iPRES 2014*, Melbourne, 6 - 10 October, pp. 141 - 148. Available at: https://ipres-conference.org/ipres14/sites/default/files/upload/iPres-Proceedings-final.pdf (Accessed: 15 September 2023).

Pennock, M. *et al*. (2016) 'The Flashback Project: rescuing disk-based content from the 1980's to the current day', *International Digital Curation Conference (IDCC) 2016*, Amsterdam, 22 - 25 February. Available at: https://zenodo.org/record/1321630#.Ym-WjdrMJhE (Accessed: 16 September 2023).

Pennock, M. and Smith, C. (2016) 'Managing an ISO 16363 Self-Assessment: A How To Guide', *International Digital Curation Conference (IDCC) 2016*, Amsterdam, 22 - 25 February. Available at: https://www.dcc.ac.uk/sites/default/files/documents/IDCC16/18_Managing_ISO16363.pdf (Accessed: 1 September 2023).

Pennock, M. and Coufal, L. eds. (2017) 'Special issue on Digital Preservation: A Global Concern', *Alexandria,* August, 27(2), pp. 63 - 65. Available at: https://doi.org/10.1177/0955749017725939 (Accessed: 15 September 2023).

Pennock, M. and Day, M. (2018) 'Adventures with ePub3: When Rendering Goes Wrong', *Proceedings of the fifteenth international conference on digital preservation: iPres 2018*, Boston and Cambridge (Massachusetts), 24 - 27 September, session 402.1. Available at: https://doi.org/10.17605/OSF.IO/94TEB (Accessed: 15 September 2023).

Pennock, M. and May, P. (2018) 'Scaled and automated preservation planning for highly diverse digital collections: the Integrated Preservation Suite', *Proceedings of the fifteenth international conference on digital preservation: iPres 2018*, Boston and Cambridge (Massachusetts), 24 - 27 September, session 402.1. Available at: https://osf.io/qxrhs (Accessed: 15 September 2023).

Pennock, M., Day, M. and Saramas, E. (2019) 'Malware Threats in Digital Preservation: Extending the Evidence Base', *Proceedings of the 16th international conference on digital preservation: iPRES 2019*, Amsterdam, 16 - 20 September, pp. 429 - 431. Available at: https://ipres2019.org/static/proceedings/iPRES2019.pdf (Accessed 15 September 2023).

Pennock, M. and May, P. (2019) 'Considerations on the Acquisition and Preservation of Mobile eBook Apps', *Proceedings of the 16th international conference on digital preservation: iPRES 2019*, Amsterdam, 16 - 20 September, pp. 212 - 220. Available at: https://ipres2019.org/static/proceedings/iPRES2019.pdf (Accessed 15 September 2023).

Pennock, M. *et al*. (2021) 'Back to Basics: The Minimum Preservation Tool', *Proceedings of the 17th International Conference on Digital Preservation: iPRES 2021,* Online, 19 - 22 October. Available at: https://phaidra.univie.ac.at/o:1424906 (Accessed: 15 September 2023).

PERICLES project (2016) *Deliverable 3.5 - Full Report on Digital Ecosystem Management.* PERICLES project report. Available at: https://web.archive.org/web/20170427214453/http://pericles-project.eu/uploads/files/PERICLES_WP3_D3_5-Full_Report_on_Digital_Ecosystem_Management_V1-0.pdf (Accessed: 1 September 2023).

Phillips, M. *et al*. (2013) 'The NDSA Levels of Digital Preservation: An Explanation and Uses', *Proceedings of the Archiving 2013 Conference*, Washington D.C., 2 - 5

April. Available at: https://doi.org/10.2352/issn.2168-3204.2013.10.1.art00047 (Accessed: 16 September 2023).

PLANETS project (2008) *Planets Planning Tool.* PLANETS project report. Available at: https://www.planets-project.eu/docs/reports/Planets_PP4-D4_PlanetsPlanningTool.pdf (Accessed: 1 September 2023).

Poole, A. H. (2016) 'The conceptual landscape of digital curation', *Journal of Documentation,* 72(5), pp. 961 - 986. Available at: https://doi.org/10.1108/JD-10-2015-0123 (Accessed: 16 September 2023).

Post, C. and Chassanoff, A. (2021) 'Beyond the workflow: archivists' aspirations for digital curation practices', *Archival Science,* Volume 21, p. 413–432. Available at: https://doi.org/10.1007/s10502-021-09365-0 (Accessed: 16 September 2023).

Potter, M. (2002) 'Researching Long Term Digital Preservation Approaches in the Dutch Digital Preservation Testbed (Testbed Digitale Bewaring)', *RLG Diginews,* 6(3). Available at: https://web.archive.org/web/20021213191632/http://www.rlg.org/preserv/diginews/diginews6-3.html#feature2 (Accessed: 1 September 2023).

Prat, N. *et al*. (2022) 'A Granular View of Knowledge Development in Design Science Research', *The Transdisciplinary Reach of Design Science Research: 17th International Conference on Design Science Research in Information Systems and Technology, DESRIST 2022*. St Petersburg, FL, USA, June 1 – 3. Lecture Notes in Computer Science, vol 13229. Springer, Cham. pp. 363-375. https://doi.org/10.1007/978-3-031-06516-3 (Accessed: 15 September 2023).

PREMIS Editorial Committee (2015) *PREMIS Data Dictionary for Preservation Metadata v3.0.* Available at: https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf (Accessed: 1 September 2023).

Purdy, G. (2011) 'Risk appetite - is using this concept worth the risk?' *NZ Society for Risk Management newsletter, RiskPost*. Available at: https://broadleaf.com.au/wp-content/uploads/2011/09/NZSRM_Risk-Appetite_Aug11_ver5.pdf (Accessed: 1 September 2023).

Ranger, J. (2014) 'Digital Preservation is People', *AVP blog*, 14 November. Available at: https://blog.weareavp.com/digital-preservation-is-people (Accessed: 1 September 2023).

Rechert, K., von Suchodoletz, D. and Welte, R. (2010) 'Emulation based services in digital preservation', *Proceedings of the 10th annual joint conference on Digital Libraries: JCDL '10*, Brisbane, 21 - 25 June, pp. 365 - 368. Available at: https://dl.acm.org/doi/10.1145/1816123.1816182 (Accessed: 16 September 2023).

Rechert, K., Falcoa, P. and Ensom, T. (2016) 'Towards a Risk Model for Emulation-based Preservation Strategies: A Case Study from the Software-based Art Domain', *Proceedings of the 13th international conference on digital preservation: iPRES*

2016, Bern, 3 - 6 October, pp. 139 - 148. Available at: https://phaidra.univie.ac.at/o:503169 (Accessed: 16 September 2023).

Rieger, O. Y. (2018) *The State of Digital Preservation in 2018: A Snapshot of Challenges and Gaps.* Ithaka S+R issue brief. Available at: https://doi.org/10.18665/sr.310626 (Accessed: 16 September 2023).

Rinehart, A., Prud'homme, P.A. and Huot, A. (2014) 'Overwhelmed to Action: Digital Preservation Challenges at the Under-resourced Institution', *OCLC Systems and Services,* 30(1), pp. 28 - 42. Available at: https://doi.org/10.1108/OCLC-06-2013-0019 (Accessed: 16 September 2023).

Risk Management Institute (2002) *A Risk Management Standard.* London: Institute of Risk Management. Available at: https://www.theirm.org/media/4709/arms_2002_irm.pdf (Accessed: 16 September 2023).

Rittel, H. and Webber, M. (1973) ' Dilemmas in a General Theory of Planning', *Policy Sciences,* 4(2), pp 155 - 169. Available at: https://www.jstor.org/stable/4531523 (Accessed: 18 September 2023).

RLG-OCLC (2002) *Trusted Digital Repositories: Attributes and Responsibilities.* Available at: https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf (Accessed: 1 September 2023).

Rog, J. and van Wijk, C. (2008) *Evaluating File Formats for Long-term Preservation.* Koninklijke Bibliotheek van Nederlands research report. Available at: https://web.archive.org/web/20200310053030/https://www.kb.nl/sites/default/files/docs/KB_file_format_evaluation_method_27022008.pdf (Accessed: 16 September 2023).

Romme, A. G. L. (2003) 'Making a Difference: Organization as Design', *Organization Science,* 14(5), pp. 558-573. Available at: https://www.jstor.org/stable/4135149 (Accessed: 16 September 2023).

Rosenblueth, A. and Wiener, N. (1945) 'The Role of Models in Science', *Philosophy of Science,* 12(4), pp. 316 - 321. Available at: https://www.jstor.org/stable/184253 (Accessed: 16 September 2023).

Rosenthal, D. S. H. *et al*. (2005) 'Requirements for Digital Preservation Systems: A Bottom-Up Approach', *D-Lib Magazine,* 11(11). Available at: https://doi.org/10.1045/november2005-rosenthal (Accessed: 16 September 2023).

Rosenthal, D. S. H. (2010) 'Format Obsolescence: Assessing the Threat and the Defences', *Library High Tech,* 28(2), pp. 195 - 210. Available at: https://doi.org/10.1108/07378831011047613 (Accessed: 1 September 2023).

Rosenthal, D. (2014) 'TRAC Audit: Lessons', *DSHR's blog*, 12 August. Available at: https://blog.dshr.org/2014/08/trac-audit-lessons.html (Accessed: 1 September 2023).

Ross, S. (2000) *Changing Trains at Wigan and the Future of Scholarship,* London: British Library National Preservation Office. Available at: https://www.researchgate.net/publication/31869566_Changing_Trains_at_Wigan_Digital_Preservation_and_the_Future_of_Scholarship (Accessed: 16 September 2023).

Ross, S. (2006) 'Approaching Digital Preservation Holistically', in M. Moss and A. Tough (eds), *Recordkeeping in a Hybrid Environment.* Oxford: Chandos Publishing, pp. 115 - 153.

Ross, S. (2012) 'Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries', *New Review of Information Networking,* 17(1), pp. 43 - 68. Available at: https://doi.org/10.1080/13614576.2012.679446 (Accessed: 16 September 2023).

Ross, S. and McHugh, A. (2006) 'The Role of Evidence in Establishing Trust in Repositories', *D-Lib Magazine,* 12 (7/8). Available at: https://doi.org/10.1045/july2006-ross (Accessed: 16 September 2023).

Rothenberg, J. (1990) 'Prototyping as Modeling: What is Being Modelled?', in H. G. Sol and K. M. Van Hee (eds), *Dynamic Modelling of Information Systems.* Amsterdam: Elsevier, pp. 335 - 359. Available at: https://doi.org/10.1016/C2009-0-12781-2 (Accessed: 16 September 2023).

Rothenberg, J. (1995) 'Ensuring the Longevity of Digital Information', *Scientific American,* 272/1, pp. 42-47. Available at: https://www.jstor.org/stable/24980135 (Accessed: 16 September 2023).

Rothenberg, J. and Bikson, T. K. (1999) *Carrying Authentic, Understandable and Usable Digital Records Through Time,* Santa Monica, CA: RAND Corporation. Available at: https://www.rand.org/pubs/rand_europe/RE99-016.html (Accessed: 16 September 2023).

Rusbridge, C. (2006) 'Excuse me... Some Digital Preservation Fallacies?', *Ariadne,* February, Issue 46. Available at: https://www.ariadne.ac.uk/issue/46/rusbridge/ (Accessed: 16 September 2023).

Ryan, H. (2014) *Who's afraid of file format obsolescence? Evaluating file format endangerment levels and factors for the creation of a file format endangerment index.* PhD dissertation. University of North Carolina. Available at: https://doi.org/10.17615/fhkk-dv90 (Accessed: 1 September 2023).

Saachi, S. (2015) *What Do We Mean by 'Preserving Digital Information'? Towards Sound Conceptual Foundations for Digital Stewardship.* PhD dissertation. Available at: https://doi.org/10.7916/D8WW7GMK (Accessed: 1 September 2023).

Sarker, S. (2007) 'Qualitative Research Genres in the IS Literature', *40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, Waikoloa, HI, USA, 3 - 6 January, pp. 244-244. Available at: https://doi.org/10.1109/HICSS.2007.456 (Accessed: 1 September 2023).

Saunders, M. N. K., Lewis, P. and Thornhill, A. (2019) *Research Methods for Business Students.* 8 ed. Harlow: Pearson Education Ltd.

Schwandt, T. A. (2001) *Dictionary of Qualitative Enquiry.* 2 ed. Thousand Oaks, California: Sage Publications.

Sierman, B. (2009) *Report on the Planets Functional Model.* PLANETS project report. Available at: https://planets-project.eu/docs/reports/Planets_PP7-D3-4_ReportOnThePlanetsFunctionalModel.pdf (Accessed: 1 September 2023).

Sierman, B (2019) 'Do FAIR data ever become heritage?', *Digital Preservation.nl blog,* 23 January. Available at: https://digitalpreservation.nl/seeds/do-fair-data-ever-become-heritage/ (Accessed: 18 September 2023).

Simon, H. A. (1996) *The Sciences of the Artificial.* 3 ed. Cambridge, MA: MIT Press.

Skinner, K. and Schultz, M. (2010) *A Guide to Distributed Digital Preservation.* Atlanta: Educopia Institute. Available at: https://metaarchive.org/wp-content/uploads/2017/03/A_Guide_to_Distributed_Digital_Preservation_0.pdf (Accessed: 1 September 2023).

Skødt, A. (2019) *Format Assessment Matrix v2.1.* The Danish National Archives assessment spreadsheet. Available at: https://github.com/the-danish-national-archives/concept-model/blob/main/P2%20Format%20Assessment/Matrix_Format%20Assessment%20v2.1.xlsx (Accessed: 16 September 2023).

Skødt, A. (2020) *Guide to Format Assessment.* The Danish National Archives memorandum, reference 19/06296. Available at: https://github.com/the-danish-national-archives/concept-model/blob/main/P2%20Format%20Assessment/Guide_Format%20Assessment%20v2.1.pdf (Accessed: 16 September 2023).

Skødt, A. (2022) 'Concept Model for Development of Preservation Plans', *Proceedings of the 18th International Conference on Digital Preservation: iPres 2022*, Glasgow, 12 - 16 September, pp. 477 - 478. Available at: http://doi.org/10.7207/ipres2022-proceedings (Accessed: 14 September 2023)

Slats, J. and Verdegem, R. (2005) 'Cost Model for Digital Preservation', *Proceedings of the 2005 DLM Forum Conference,* Budapest, 5 - 7 October. Available at: https://dlmforum.typepad.com/paper_remcoverdegem_and_js_costmodelfordigitalpreservation.pdf (Accessed: 16 September 2023).

Snodgras, A. and Coyne, R. (1997) 'Is Designing Hermeneutical?', *Architectural Theory Review,* 1(2), pp. 65 - 97. Available at: https://doi.org/10.1080/13264829609478304 (Accessed: 16 September 2023)

Society for Risk Analysis (2018) *Glossary.* Available at: https://www.sra.org/wp-content/uploads/2020/04/SRA-Glossary-FINAL.pdf (Accessed: 16 September 2023).

Spencer, R. (2022) 'Fractal in detail: What information is in a file format identification report?', *Code4Lib,* Issue 53. Available at: https://journal.code4lib.org/articles/16351 (Accessed: 16 September 2023).

Stanescu, A. (2004) 'Assessing the Durability of Formats in a Digital Preservation Environment: The INFORM Methodology', *D-Lib Magazine,* 10(11). Available at: https://doi.org/10.1045/november2004-stanescu (Accessed: 16 September 2023).

Starr, C. (2003) 'The precautionary principle versus risk analysis', *Risk Analysis,* 23(1). Available at: https://doi.org/10.1111/1539-6924.00285 (Accessed: 16 September 2023).

State Library of New South Wales (2022) *Digital Preservation Framework 2022 - 2024.* SL-NSW working paper. Available at: https://www.sl.nsw.gov.au/sites/default/files/digitalpreservationframework-2022-24.pdf (Accessed: 16 September 2023).

Steen, M. (2019) 'Preservation Planning, Beacons for a TDR: Three Cases on Archiving Emerging Media', *Proceedings of the 16th International Conference on Digital Preservation: iPRES 2019*, Amsterdam, 16 - 20 September, pp. 134 - 141. Available at: https://ipres2019.org/static/proceedings/iPRES2019.pdf (Accessed: 14 September 2023).

Steenbakkers, J. F. (2005) 'Digital Archiving in the Twenty-First Century', *Library Trends*, Summer, 54(1). Available at: https://doi.org/10.1353/lib.2006.0010 (Accessed: 16 September 2023).

Strodl, S., Petrov, P. and Rauber, A., (2011) *Research on Digital Preservation within projects co-funded by the European Union in the ICT Programme.* European Commission programme report. Available at: https://web.archive.org/web/20130319122613/http://cordis.europa.eu/fp7/ict/telearn-digicult/report-research-digital-preservation_en.pdf (Accessed: 16 September 2023).

Strodl, S. *et al*. (2013) *D4.6: Use Case Specific DP and Holistic Escrow.* TIMBUS project report. Available at: https://timbusproject.net/dmdocuments/D4.6_M24_Use%20Case%20Specific%20DP%20and%20Holistic%20Escrow.pdf
(Accessed: 1 September 2023).

Taherdoost, H. (2022) 'What are Different Research Approaches? Comprehensive Review of Qualitative, Quantitative, and Mixed Method Research, Their

Applications, Types, and Limitations', *Journal of Management Science and Engineering Research,* 5(1), pp. 53 - 63. Available at: https://doi.org/10.30564/jmser.v5i1.4538 (Accessed: 16 September 2023).

The National Archives (2023a) *DiAGRAM Glossary.* Available at: https://diagram.nationalarchives.gov.uk/glossary.html (Accessed: 1 September 2023).

The National Archives (2023b) *DiAGRAM website.* Available at: https://diagram.nationalarchives.gov.uk/ (Accessed: 1 September 2023).

Thirifays, A. *et al*. (2014) *D3.3—Curation Costs Exchange Framework.* 4C project report. Available at: https://www.4cproject.eu/d3-3-curation-costs-exchange-framework/ (Accessed: 16 September 2023).

Thomas, O. (2007) 'Version Management for Reference Models: Design and Implementation', in J. Becker and P. Delfmann (eds), *Reference Modeling: Efficient Information Systems Design Through Reuse of Information Models.* Münster: Physica-Verlag, pp. 1 - 26. Available at: https://doi.org/10.1007/978-3-7908-1966-3 (Accessed: 16 September 2023).

Thompson, K. M., Deisler, P. F. J. and Schwing, R. C. (2005) 'Interdisciplinary Vision: The First 25 Years of the Society for Risk Analysis (SRA), 1980–2005', *Risk Analysis,* 25(6), pp. 1333 - 1386. Available at: https://doi.org/10.1111/j.1539-6924.2005.00702.x (Accessed: 16 September 2023).

Thuan, N. H., Dreschler, A. and Antunes, P. (2019) 'Construction of Design Science Research Questions', *Communications of the Association for Information Systems,* 44(3), pp. 332 - 363. Available at: https://doi.org/10.17705/1CAIS.04420 (Accessed: 16 September 2023).

TIMBUS project (2014) *A Risk Management Approach to Preservation of Business Processes.* TIMBUS project report. Available at: https://www.timbusproject.net/dmdocuments/Vieira_WhitePaper_Risk%20Management_JB.pdf (Accessed: 30 August 2023).

Todd, M. (2009) *DPC Technology Watch report: File formats for Preservation.* Digital Preservation Coalition Technology Watch report. Available at: https://www.dpconline.org/docs/technology-watch-reports/375-file-formats-for-preservation/file (Accessed: 16 September 2023).

Tomiyama, T. *et al*. (2003) 'Abduction for Creative Design', *Papers from the 2003 AAAI Spring Symposium*, Palo Alto. AAAI. Available at: https://aaai.org/papers/0033-abduction-for-creative-design/ (Accessed: 16 September 2023).

Tylers, W. (1995) 'Thinking about Archival Preservation in the 90's and Beyond', *American Archivist,* 58(4), pp. 476 - 492. Available at: https://www.jstor.org/stable/40293944 (Accessed: 16 September 2023).

U.S National Archives and Records Administration (2023) *NARA File Format Risk Matrix.* US-NARA assessment spreadsheet. Available at: https://github.com/usnationalarchives/digital-preservation/blob/master/Digital_Preservation_Risk_Matrix/NARA_File_Format_Risk_Matrix_20230627.xlsx (Accessed: 16 September 2023).

U.S. National Archives and Records Service (1984) *White paper: Strategic Technology Considerations Relative to the Preservation and Storage of Human and Machine Readable Records,* Washington DC: NARS.

UK Government (2023) *The Orange Book - Management of Risk - Principles and Concepts.* Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1154709/HMT_Orange_Book_May_2023.pdf (Accessed: 16 September 2023).

ULCC (2009) *The AIDA self assessment toolkit.* AIDA project report. Available from Legal Deposit Library reading rooms at: https://www.webarchive.org.uk/wayback/en/archive/20141204025238/http://aida.jiscinvolve.org/wp/toolkit/ (Accessed: 16 September 2023).

Underdown, D. (2019) 'Digital archiving is a risky business', The National Archives' blog, 3 June. Available at: https://blog.nationalarchives.gov.uk/digital-archiving-is-a-risky-business/ (Accessed: 16 September 2023).

UNESCO (2001) *Final Report: International Round Table on 'Intangible Cultural Heritage – Working Definitions'.* Report from the International Round Table on 'Intangible Cultural Heritage – Working Definitions' Turin, Italy, 14 – 17 March. Available at: https://ich.unesco.org/doc/src/00077-EN.pdf (Accessed: 16 September 2023).

UNESCO (2023) *The Concept of Digital Preservation.* Available at: https://en.unesco.org/themes/information-preservation/digital-heritage/concept-digital-preservation (Accessed: 16 September 2023).

University of Manchester Library (2020) *Digital Preservation Statement March 2020.* UoML working paper. Available at: https://documents.manchester.ac.uk/display.aspx?DocID=58357 (Accessed: 16 September 2023).

van Aken, J. E. (2004) 'Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules', *Journal of Management Studies,* 41(2), pp. 219-246. Available at: https://doi.org/10.1111/j.1467-6486.2004.00430.x (Accessed: 16 September 2023).

van der Hoeven, J., Lohman, B. and Verdegem, J. (2007) 'Emulation for Digital Preservation in Practice: The Results', *International Journal of Digital Curation*

(IJDC), 2(2) pp 123 - 132. Available at: https://doi.org/10.2218/ijdc.v2i2.35 (Accessed: 16 September 2023).

van der Knijft, J. (2013) 'Assessing file format risks: searching for Bigfoot?', *OPF blog*, 30 September. Available at: https://openpreservation.org/blogs/assessing-file-format-risks-searching-bigfoot/ (Accessed: 16 September 2023).

van der Knijft, J. (2021) 'On the Significant properties of Spreadsheets', *Bitsgalore blog,* 24 September. Available at: https://www.bitsgalore.org/2021/09/24/on-the-significant-properties-of-spreadsheets (Accessed: 16 September 2023).

van der Werf-Davelaar, T. (1999) Long-Term Preservation of Electronic Publications: The NEDLIB project. *D-Lib Magazine*, 5(9). Available at: https://doi.org/ 10.1045/september99-vanderwerf (Accessed: 16 September 2023).

Venable, J., Pries-Heje, J. and Baskerville, R. (2016) 'FEDS: a Framework for Evaluation in Design Science Research', *European Journal of Information Systems,* Volume 25, pp. 77 - 89. Available at: https://doi.org/10.1057/ejis.2014.36 (Accessed: 16 September 2023).

Venable, J. R., Pries-Heje, J. and Baskerville, R. L. (2017) 'Choosing a Design Science Methodology', *ACIS 2017 Proceedings,* Hobart, 5 - 6 December. Available at: https://aisel.aisnet.org/acis2017/112 (Accessed: 16 September 2023).

Vermaaten, S., Lavoie, B. and Caplan, P. (2012) 'Identifying Threats to Successful Digital Preservation: the SPOT Model for Risk Assessment', *D-Lib Magazine,* 18 (9/10). Available at: https://doi.org/doi:10.1045/september2012-vermaaten (Accessed: 16 September 2023).

vom Brocke, J. and Buddendick, C. (2006) 'Reusable conceptual models – requirements based on the design science research paradigm', *Proceedings of the First Design Research Information Systems and Technology Conference: DESRIST 2006,* Claremont, CA, 24 - 25 February, pp. 576–604. Available at: https://www.researchgate.net/publication/228770897 (Accessed: 16 September 2023).

vom Brocke, J., Hevner, A. and Maedche, A. (2020a) *Design Science Research Cases.* Cham: Springer Nature. Available at: https://doi.org/10.1007/978-3-030-46781-4_1 (Accessed: 16 September 2023).

vom Brocke, J., Hevner, A. and Maedche, A. (2020b) 'Introduction to Design Science Research', in J. vom Brocke, A. Hevner and A. Maedche (eds) *Design Science Research. Cases.* Cham: Springer, pp. 1 - 12. Available at: https://doi.org/10.1007/978-3-030-46781-4_1 (Accessed: 16 September 2023).

vom Brocke, J. *et al*. (2020) 'Special issue Editorial: Accumulation and Evolution of Design Knowledge in Design Science Research: A Journey Through Time and Space', *Journal of the Association for Information Systems (JAIS),* 21(3), pp. 520 - 544. Available at: https://doi.org/10.17705/1jais.00611 (Accessed: 16 September 2023).

vom Brocke, J., Weber, M. and Grisold, T. (2021) 'Design Science Research of High Practical Relevance: Dancing through Space and Time', in S. Aier, P. Rohner, J. Schelp (eds), *Engineering the Transformation of the Enterprise.* Cham: Springer, pp. 115 - 135. Available at: https://doi.org/10.1007/978-3-030-84655-8_8 (Accessed: 16 September 2023).

Wacker, J. G. (2008) 'A Conceptual Understanding of requirements for Theory-Building Research: Guidelines for Scientific theory Building', *Journal of Supply Chain Management,* 44(3), pp. 5 - 15. Available at: https://doi.org/10.1111/j.1745-493X.2008.00062.x (Accessed: 16 September 2023).

Waddington, S. *et al*. (2016) 'PERICLES - Digital Preservation through Management of Change in Evolving Ecosystems', in S. Hamrioui (ed), *The Success of European projects using New Information and Communication Technologies.* Portugal: SCITEPRESS, pp. 51 - 74. Available at: https://www.researchgate.net/publication/317177717 (Accessed: 16 September 2023).

Waller, R. (2019) 'Collection Risk Assessment', in L. Elkin and C. A. Norris (eds), *Preventative Conservation Collection Storage.* New York: Society for the Preservation of Natural History Collections, pp. 59-90. Author content available at: https://www.researchgate.net/publication/335682123 (Accessed: 16 September 2023).

Webb, C., Pearson, D. and Koerbin, P. (2013) 'Oh, you wanted us to preserve that?!' Statements of Preservation Intent for the National Library of Australia's Digital Collections. *D-Lib Magazine*, 19 (1/2). Available at: https://doi.org/10.1045/january2013-webb (Accessed: 16 September 2023).

Weber, S. (2010) 'Design Science Research: Paradigm or Approach?', *AMCIS 2010 Proceedings,* Lima, Peru, 12 - 15 August. Available at: http://aisel.aisnet.org/amcis2010/214 (Accessed: 16 September 2023).

Westphal, M. *et al*. (2017) 'Future directions in risk science', *International Journal of Risk Assessment and Management,* 20(1 - 3), pp. 240 - 260. Available at: https://doi.org/10.1504/IJRAM.2017.082567 (Accessed: 16 September 2023).

Wheatley, P. (2018) 'A valediction for validation?', *DPC blog*, 11 October. Available at: https://www.dpconline.org/blog/a-valediction-for-validation (Accessed: 16 September 2023).

Wheatley, P. (2022) *DPC Technology Watch: A risk driven approach to Bitstream preservation.* Digital Preservation Coalition Technology Watch report. Available at: http://doi.org/10.7207/twgn22-02 (Accessed: 16 September 2023).

Wheatley, P. and Hole, B. (2009) LIFE3: Predicting Long Term Digital Preservation Costs, *Proceedings of the Sixth International Conference on Preservation of Digital*

Objects: iPRES 2009, San Francisco, 5 - 6 October, pp. 206 - 210. Available at: https://escholarship.org/uc/item/23b3225n (Accessed: 16 September 2023).

Wheatley, P. and Pennock, M. (2013) 'Supporting Practical Preservation Work and Making it Sustainable with SPRUCE', *Proceedings of the 10th International Conference on Preservation of Digital Objects: iPRES 2013*, Lisbon, 2 - 6 September, pp. 73 - 77. Available at: https://purl.pt/24107 (Accessed: 16 September 2023).

Wilkinson, M., *et al*. (2016) ' The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data,* Vol 3, Article number: 160018. Available at: https://doi.org/10.1038/sdata.2016.18 (Accessed: 30 September 2023).

Wilson, A. (2007) *Significant Properties Report.* INSPECT project report. Available at: https://web.archive.org/web/20090609141501/http://www.significantproperties.org.uk/documents/wp22_significant_properties.pdf (Accessed: 16 September 2023).

Winter, R. (2008) 'Design science research in Europe', *European Journal of Information Systems,* Volume 17, pp. 470-475. Available at: https://doi.org/10.1057/ejis.2008.44 (Accessed: 16 September 2023).

Winter, R., Gericke, A. and Bucher, T. (2009), 'Method versus Model - Two Sides of the Same Coin?' in Albani, A., Barjis, J., Dietz, J.L.G. (eds) *Advances in Enterprise Engineering III*. CIAO! EOMAS 2009 Proceedings. Lecture Notes in Business Information Processing, vol 34. Springer, Berlin. pp. 1 - 15. Available at: https://doi.org/10.1007/978-3-642-01915-9_1 (Accessed: 29 September 2023).

Yeo, C. (2019) *Principles of Risk Analysis: Decision Making under Uncertainty.* 2 ed. Boca Raton: Taylor and Francis.

Ylönen, M. and Aven, T. (2023) 'A framework for understanding risk based on the concepts of ontology and epistemology', *Journal of Risk Research,* Issue 6, pp. 581 - 593. Available at: https://doi.org/10.1080/13669877.2023.2194892 (Accessed: 16 September 2023).

Appendix A: Glossary of Key Terms

Term	Definition
Accessibility	The ability to access the digital content that is the meaningful focus of a preservation effort; a capability provided by the technological infrastructure; a target value.
Authenticity	The reliable and accurate rendering of a digital object; a property of the digital object; a target value.
Budget class	A class of risk source representing the sum of the financial envelope that supports an organisational digital preservation endeavour; associated with the organisational infrastructure risk originating entity.
Consequence	A negative outcome.
Content File(s) class	A class of risk source representing the digital file(s) in which an object and its intellectual content is encoded, typically in the structure of a given file format; associated with the digital content risk originating entity.
Contextualised Risk Source	A combination of a specific risk source instance with a specific factor, as relevant to a specific institutional setting.
Digital Content	The focus of a preservation endeavour; a risk originating entity.
Digital File(s)	The bitstream(s) that contain all of the encoded information for generating (and optionally also describing) a digital object.
Digital Object	The artefact that is the focus of the preservation effort.
Digital Preservation	The series of coordinated organisational and technological activities undertaken in an organisation throughout the lifecycle to ensure its digital content is retrievable, authentic, has integrity, and is accessible over time for current and future users.
Digital Preservation Risk	The potential for complete or partial loss of digital collection content in terms of its target values of retrievability, authenticity, integrity, accessibility, and longevity, arising from sub-optimised risk sources within the managed organisational and technological environment in which the content should otherwise be preserved.
Integrity	The binary sequence of a digital file remaining whole and unchanged since it was last confirmed; a property of a digital file; a target value.
Intellectual content	The meaningful knowledge or information consumed by humans or machines.

Legal class	A class of risk source representing the legislative and contractual framework that underpins digital preservation in an organisational context; associated with the organisational infrastructure risk originating entity.
Longevity	The endurance of content over time and throughout the lifecycle so it remains available for current and future users; a property and a capability; a target value.
Metadata class	A class of risk source representing information about digital content that supports its preservation and access; associated with the digital content risk originating entity.
Network class	A class of risk source representing the communication protocols in place to enable the exchange and transfer of data and resources across all the physical hardware utilised in a technological infrastructure; associated with the technological infrastructure risk originating entity.
Organisational Infrastructure	The organisational environment in which a managed digital preservation service occurs; a risk originating entity.
People class	A class of risk source representing the staffing or personnel resource associated with digital preservation; associated with the organisational infrastructure risk originating entity.
Physical Hardware class	A class of risk source representing all of the tangible machines, wiring, and other physical components needed to support a technological infrastructure; associated with the technological infrastructure risk originating entity.
Policy class	A class of risk source representing the policy ecosystem required to underpin digital preservation activities and draw on a series of appropriate and consistent principles; associated with the organisational infrastructure risk originating entity.
Preservation Objective	The goal of a digital preservation endeavour.
Preservation Plan	A risk treatment plan for a given scenario, developed in response to a risk assessment.
Processes and Workflows class	A class of risk source representing the processes and workflows that support digital preservation activities, both manual and automated; associated with both the organisational infrastructure and technological infrastructure risk originating entities.

Rendering Software class	A class of risk source representing software used to process digital object files in order to access and render a digital object and its intellectual content; associated with the technological infrastructure risk originating entity.
Retrievability	The ability to recover specific digital content files from their storage location with relative ease; a capability provided by the technological infrastructure; a target value.
Risk	A generic concept caused by uncertainty and associated with a negative outcome in relation to planned or specific objectives.
Risk Appetite	The threshold for an acceptable level of risk in relation to planned or specific objectives.
Risk Assessment	The process of identifying, analysing, and evaluating individuated risks.
Risk Description	A description of risk that associates a contextualised risk source with an uncertainty.
Risk Factor	A variable property of a risk source that can be optimised to reduce uncertainty and the likelihood or impact of a negative outcome.
Risk Matrix	A likelihood/impact framework against which to measure individuated risks.
Risk Originating Entity	An aspect of the operational digital preservation environment that contains risk sources.
Risk Source	A changeable element in the digital preservation environment that alone or in combination with others has the intrinsic potential to give rise to a negative outcome.
Risk Source Class	A conceptual grouping of similar risk sources.
Risk Source Instance	An individual instance of a risk source.
Risk Source Instance Type	Different types of risk sources within a given class.
Risk Statement	A full expression of risk that associates a risk description with a specific consequence.
Storage Media class	A class of risk source representing the physical devices upon which files are stored; associated with the digital content risk originating entity.
Strategy class	A class of risk source representing the framework of institutional strategies relevant to digital preservation; associated with the organisational infrastructure risk originating entity.

System Software class	A class of risk source representing software that facilitates management of a technological infrastructure and the digital objects held within; associated with the technological infrastructure risk originating entity.
Target Values	The goals of a preservation endeavour expressed as properties or capabilities of the environment and its content.
Technological Infrastructure	The technological environment in which digital content is acquired, managed, preserved, and made accessible from; a risk originating entity.
Uncertainty	A lack of knowledge in relation to a risk factor.

Appendix B: CHARM Risk Identification Framework Questions

This Risk Identification Framework represents Method Two, as described in this thesis. The questions from the framework are reproduced here for information purposes. The full template is available as a supplementary file to the CHARM 'How-To' Guide.

Digital Content

The Digital Content Entity contains three classes of risk source: Content File(s), Metadata, and Storage Media.

Content File(s)

Factor	**Question & Explanation**
Completeness	Do the files contain all of the intellectual content to which you expect to provide access? *Rendered objects can sometimes display or use information held externally, linked from the content files. If this additional information is important but not available then the authenticity of the rendered objects can be affected.*
Dependencies	Can you support any technical dependencies that affect how users access and interact with the rendered files? *Objects may have specific technical dependencies other than those relating to their format, for example relating to certain hardware or peripheral devices. If these are not supported then it can be difficult to provide an access environment that supports authentic rendering of objects.*
DRM	Are the files free from embedded DRM that might prevent you from opening them? *Files can have restrictions or controls on them to prevent unauthorised use, which can inhibit accessibility.*
Encoding	Do you know how the files in the collection are encoded, e.g. what file format (and version if relevant) they are in? *File formats need to be identified so that you can determine what software (and sometimes also hardware) is required to interpret them. Without this information, it can be more difficult to identify and provision access requirements that support authentic rendering of objects.*
Fixity	Has the integrity of the bit streams been confirmed?

	Changes to the integrity of a bitstream may adversely affect how a file is interpreted by software. If the integrity of the bitstream is damaged, the file may not render in an authentic manner even if its technical dependencies are all satisfied.
Malware	Are the content files free from malware? *Different types of malware exist, with many different types of undesirable consequences. If malware is present and is not appropriately managed then it can adversely impact on target values and cause other, reputational and organisational issues.*

Storage Media

Factor	**Question & Explanation**
Compatibility	Do you have compatible hardware and software to interact with this storage media? *Storage media needs to be read by appropriate physical drives and software. If compatible drives are not available when required then this can adversely affect retrievability and authenticity, and eventually also file integrity.*
Condition	Is the media free of obvious damage or degradation? *Storage media has a physical form that is susceptible to damage or decay. If it is not in good condition then this can prevent access to the stored files and damage their integrity.*
Location	Is the media held in a secure and safe location with appropriate environmental and access controls? *Storage media needs to be held in a secure location, accessible to authorised personnel and safe from traditional agents of deterioration. If this is not done properly then these can affect file integrity, longevity, retrievability and accessibility.*
Quality	Is the media of archival quality and durable for your expected storage timeframe? *Good quality storage media often has an expected 'shelf life' that indicates for how long it should remain reliable. If the shelf life is exceeded then its condition is likely to degrade,*

	which can inhibit access to the stored files and damage their integrity.

Metadata

Factor	Question & Explanation
Accuracy	Are you confident that metadata associated with the object is correct? *Accurate metadata informs the selection of management and preservation actions. If metadata is incorrect then it can lead to incorrect management and preservation actions, which can affect target values.*
Comprehensive-ness	Is sufficient metadata available for the object that meets your expectations and requirements? *Comprehensive metadata helps ensure that institutions have all the information they need to manage and preserve content. If this is not available then it can lead to uncertainty about the suitability of management and preservation actions, which can affect target values.*

Organisational Infrastructure

The Organisational Infrastructure Entity contains six classes of risk source: Strategy; Legal; Policy; People; Budget, and; Processes/Workflows.

Strategy

Factor	Question & Explanation
Prioritisation	Does your strategy sufficiently prioritise preservation? *Strategies define organisational priorities and identify the most important activities that an organisation will support. If a strategy does not clearly prioritise preservation then it can be difficult to ensure appropriate organisational support, which can manifest in different ways and adversely impact on target values.*

Policy

Factor	Question & Explanation
Clarity	Does your policy clearly support preservation? *Policies define operational principles, standards and goals. If a policy does not clearly support preservation then it can be difficult to ensure operational practices do the same, adversely affecting target values as relevant to the policy in question.*
Compliance	Do organisational practices comply with the policy? *Policies are a means to an end, not an end in and of themselves: they must be implemented. If organisational practices do not comply with policy then this can adversely affect target values as relevant to the policy or practice in question.*
Suitability	Are the preservation principles outlined in your policy suitable to achieve your goals? *Policies must outline good preservation principles that can help achieve target values across the lifecycle. If the preservation principles are unsuitable then this can adversely affect different target values depending on the specific principles in question.*

People

Factor	Question & Explanation
Capabilities	Do available personnel have the capabilities needed to implement the expected preservation practices? *People working on digital preservation activities need certain skills and knowledge to deliver their roles. A skills or knowledge gap can lead to errors or delays, affecting different target values depending on where the skills or knowledge shortage lies.*
Capacity	Do you have sufficient capacity within your personnel complement to support preservation? *Preservation requires people, either in dedicated roles or with preservation responsibilities as part of a wider role. If there is insufficient personnel capacity within the*

	organisation to support preservation then this can affect different target values depending on where the capacity shortage lies.

Budget

Factor	Question & Explanation
Quantity	Do you have sufficient budget available to implement your preservation policy over the next 3 - 5 years? *Preservation activities need funding. If there is insufficient budget available then this limits the activities that can be supported effectively, affecting different target values depending on where budget shortage lies.*

Legal

Factor	Question & Explanation
Permissions	Do you have the legal rights needed to support and enable your preservation activities? *Legislative and contractual frameworks can limit options for preservation by, for example, restricting re-use or retention of content and software. If preservation options are constrained in this way, it can impact on different target values depending on where the constraint lies.*

Processes & Workflows

Factor	Question & Explanation
Consistency	Are your manual processes and workflows consistent with each other and your policy? *Inconsistent processes and workflows can lead to inefficiencies and uncertainty about how they function. This can affect different target values, depending on the specific processes or workflows being assessed.*
Documentation	Are your manual processes and workflows clearly documented and is that documentation up to date? *Up to date documentation ensures third parties can understand how processes are intended to function, which is*

	important over time. If documentation is unavailable or out of date, then it can negatively affect target values depending on the processes concerned.
Effectiveness	Are your manual processes and workflows effective in delivering the expected outputs or outcomes? *Do your processes and workflows function as expected? If not then they can affect different target values, depending on the nature of the processes concerned.*

Technological Infrastructure

The Technological Infrastructure Entity contains five classes of risk source: Rendering Software, System Software, Physical Hardware, Network, and Processes/Workflows

System Software

Factor	**Question & Explanation**
Availability	Is the necessary system software available? *There are many different types of system software. If the software that you need isn't available then this can affect your target values, depending on the nature of the software or the function it should support. If you don't yet know what you need, make a note for further investigation.*
Compatibility	Is the system software that you need/have compatible with other components of your technological infrastructure? *Individual system software generally operates as part of a wider technological infrastructure, including hardware and other software. If it is incompatible with other parts of the infrastructure, then this can adversely impact on different target values depending on the software concerned.*
Configuration	Is your software configuration consistent with your preservation policy and security expectations? *Software configuration is the process of specifying changeable settings so that software performs as needed. If software is not configured properly then this can adversely affect its expected function and target values.*
Documentation	Is adequate and relevant documentation available so that you know how to use your software?

	Software documentation explains how to set up and use software properly. If this is missing or incorrect then it can be difficult to ensure the software can be used as expected, which can adversely impact different target values depending on the software concerned.
Quality	Is the software reliable and free from bugs or defects? *Good quality software works as expected, is well-tested, and is free from significant issues that otherwise affect its performance. If software is not of good quality then it is likely to demonstrate faults or glitches, which can adversely impact different target values depending on the function of the software concerned*.
Support	Is the software still appropriately supported by its originator or a third party*?* *Software support services can take different forms depending on the product and vendor or community, ranging from bug fixes, patches and updates, to troubleshooting and migration upgrades. If the support life of software is exceeded then this can result in problems affecting different target values, depending on the services and software concerned.*

Rendering Software

Factor	**Question & Explanation**
Availability	Is suitable rendering software available? *Having access to suitable rendering software is an important first step in enabling access to authentic representations of content. If appropriate rendering software is not available then it may be difficult, costly, or not possible to provide appropriate access. If you don't yet know what you need, make a note for further investigation.*
Compatibility	Is the rendering software that you need compatible with other components in your technological infrastructure? *Rendering software requires deployment on an appropriate technological platform, including both hardware and other software. If they are not compatible then this can adversely affect its functionality and inhibit access or affect other target values, depending on the software concerned.*

Documentation	Is adequate and relevant documentation available so that you - and your users - know how to use the software? *Software documentation explains how to set up and use software properly. If this is missing or incorrect then it can be difficult to ensure the software works as expected, which can adversely affect its expected function, impacting on accessibility and authenticity of rendered content.*
Quality	Is your rendering software reliable and free from bugs or defects? *Good quality software works as expected, is well-tested, and is free from significant issues that otherwise affect its performance. If software is not of good quality then it is likely to demonstrate faults or glitches, which can adversely affect how content is provided to users, its authenticity and its accessibility.*
Support	Is your rendering software still appropriately supported by its originator or a third party? *The availability of software support services reflects the currency and usability of software. They are typically associated with software that is still in (relatively) widespread use. A lack of support for software indicates that it may be obsolete, which can make it difficult to deploy and affect target values relating to authenticity and accessibility.*
Updates/ upgrades	Are updates/upgrades backwards compatible? *Backwards compatibility generally means that new versions of software will support files created or used with previous versions of the same software. If backwards compatibility is limited, then upgrades or updates to rendering software may change how they interact with older files and affect the authenticity and/or accessibility of rendered objects.*

Physical Hardware

Factor	**Question & Explanation**
Availability	Is suitable hardware available? *Different types of hardware support different functions. If the hardware that you need isn't available then this can affect your ability to achieve your target preservation values. If you*

	don't yet know what you need, make a note for further investigation.
Compatibility	Is the hardware that you need compatible with other components of your technological infrastructure? *Hardware is usually used in combination with other hardware and software. If compatible items are not available when required then this can adversely impact on your ability to achieve your target preservation values.*
Condition	Is the hardware in good condition and free from defects? *Hardware has a physical form that is susceptible to damage or decay. If it is not in good condition then this can prevent access to systems and content, affecting target values.*
Location	Is the hardware in a secure and safe location with appropriate environmental and access controls? *Hardware needs to be held in a secure location, accessible to authorised personnel and safe from traditional agents of deterioration. If this is not done properly then these can affect file integrity, longevity, retrievability and accessibility.*
Quality	Is the hardware of good quality and durable for your expected timeframe? *Good quality hardware is usually durable and reliable for a foreseeable period of time. Lesser quality hardware may not last as long. Hardware failures can negatively affect different target values, depending on the function of the hardware concerned.*
Quantity	Do you have enough of the hardware to meet your needs? *The amount of different hardware you need will vary depending on what you need to do. If you do not have enough to meet your demands, then this can result in problems affecting different target values, depending on the hardware concerned.*
Support	Is the hardware you have or need still appropriately supported by the vendor or a third party? *Hardware support can range from troubleshooting to warrantees, availability of replacement parts, and compatibility with other products. If the support life is*

	exceeded then this can result in problems affecting different target values, depending on the hardware concerned.

Network

Factor	Question & Explanation
Capacity	Does your network have sufficient capacity to complete your processes in the time required? *Network capacity enables transfer of content or information across different aspects of the infrastructure. Insufficient capacity can lead to bottlenecks or failures, which can affect different target values, depending on the process affected.*
Security	Is your network securely configured and maintained? *Good network security configuration limits opportunities for internal or external attack, as well as meaningful errors. Attacks can exploit weaknesses and affect content longevity and other target values, depending on the nature of the attack.*

Processes & Workflows

Factor	Question & Explanation
Consistency	Are your automated processes and workflows consistent with each other and your policy? *Inconsistent processes and workflows can lead to inefficiencies and uncertainty about how they function. This can affect different target values, depending on the specific processes or workflows being assessed.*
Documentation	Are your automated processes and workflows clearly documented and is that documentation up to date? *Up to date documentation ensures third parties can understand how automated processes are intended to function, which is important over time. If documentation is unavailable or out of date, then it can negatively impact on target values depending on the nature of the processes concerned.*

Effectiveness	Are your automated processes and workflows effective in delivering the expected outputs? *Do your processes and workflows function as expected? If not then they can affect different target values, depending on the nature of the processes concerned.*

Appendix C: CHARM Risk Assessment Spreadsheet (RAS) Tables

The Risk Assessment Spreadsheet is used in Method Three, as described in the thesis. The tables from the spreadsheet are reproduced here for illustrative purposes. The full template is available as a supplementary file to the CHARM 'How-To' Guide.

Scenario table

Assessment Name:	

Assessment Type:	
Assessment Scope:	
Assessment Trigger:	
Date of Assessment:	
Assessor(s):	

Scope description:	

Risk Appetite for Scope:	
Risk Appetite Description:	
Justification for Appetite:	

Assessment table

			Contextualised Risk Source		Risk Statement				Risk Score			
#	Class	Instance Type	Instance	Factor	Description	Consequence	Affected values	Status	Impact	Likelihood	Overall	Mitigation Constraints
1												
2												
3												
4												
5												
6												
7												
8												
9												
10												
11												