Article

JIS

Journal of Information Science
39(4) 442–458

DOI: 10.1177/0165551513480107
jis.sagepub.com
SAGE

Sustaining accessibility of information through digital preservation: A literature review

Daniel Burda
SAP Research, Switzerland

Frank Teuteberg
University of Osnabrück, Germany

Abstract
Today valuable business, scientific and cultural information is created and stored digitally. However, the sustained availability and accessibility of digital information cannot be taken for granted as they are threatened by rapidly changing technologies and the associated risk of obsolete software and hardware. Digital preservation (DP) comprises methods and techniques geared towards securing long-term access to digital information. In this study we undertake a vital step to shed light on the area of DP by conducting a systematic literature review. We analysed 122 publications with respect to various aspects such as drivers, stakeholders or applied research methods. The findings point towards some unsolved organizational issues such as lacking methods to support cost–benefit analysis or DP decision-making. We thus propose a set of research questions based on our findings and suggest a research agenda.

Keywords
digital preservation; long-term archiving; research agenda; systematic literature review

1. Introduction

In this information age data[1] is perceived as both an end-product itself and a factor of production within the production process of other goods or decisions in an organization [1]. Valuable business, scientific and cultural information assets are created and accessed digitally within various types of organizations and across the public and private sectors [2]. However, the diffusion of information technology and our dependence on the exponentially growing volume of data also pose challenges for organizations and the information society [3]. While digital information has gained the status of an intangible asset, its sustained accessibility cannot be taken for granted [4, 5]. Rather, it has to be actively secured against the threats imposed by rapidly changing technologies and organizational change. Digital preservation (DP) may provide appropriate means of securing long-term access to digital information. The rising awareness of the challenges of preserving digital information has led to a growing interest in the research community and various publicly funded projects in DP in recent years [6].

Digital preservation can be understood as 'the ability to sustain the accessibility, understandability and usability of digital objects in the distant future regardless of changes in technologies and in the ''designated communities'' (data consumers) that use these digital objects' [7]. DP comprises techniques like migration or emulation to ensure long-term access to digital information [8]. The meaning of long-term has been defined in the Open Archival Information System standard as 'long enough to be concerned with the impacts of changing technologies, including support for new media and data formats or with a changing user community. Long-Term may extend indefinitely' [9]. That is, the meaning of long-term is rather concerned with the threatening effects of rapidly changing technologies and the lifetime of storage

Corresponding author:
Daniel Burda, SAP Research, Althardstrasse 80, CH-8105 Regensdorf, Switzerland.
Email: daniel.burda@sap.com

media on the accessibility of digital information than the notion of time in the sense of several years, decades or even centuries. In this vein, DP can be thought of as a set of methods and techniques geared towards the mitigation of the 'deleterious effects of technology obsolescence' on the accessibility of digital information [10].

We may conceive DP as an organizational issue of decision-making concerning what (and what not) to preserve while considering both organizational constraints such as costs or compliance objectives and technological aspects. However, we assert that DP has as yet been addressed in rather a domain-centric way that focuses on the development and evaluation of preservation techniques or strategies. Despite the vast number of studies, mainly driven by the digital library domain, we see a lack of research examining DP from an organizational point of view.

It is thus the intention of this study to investigate how and to what extent DP is currently addressed within research. Towards this end, we conduct a systematic review of literature from the area of computer science and management information systems (MIS) while focusing on organizational aspects of DP. This paper is structured as follows: In Section 2 our underlying research methodology and literature search strategy are described. Section 3 describes the findings of our review. Based on those findings we derive a set of research questions and contribute an agenda for future research in Section 4. Section 5 concludes the main contributions and further points to the limitations of this study.

2. Research methodology

2.1. Overall approach: systematic literature review

A review of relevant literature is crucial for any academic research project and demonstrates that the proposed research extends the existing body of knowledge. It should systematically analyse and synthesize quality literature, provide a solid foundation to a particular research question, present the concluding implications for researchers and practitioners and motivate future research [11, 12]. In this line of thinking we follow Bandara et al. [13], who propose a systematic and tool-support approach to review literature. The authors present a framework consisting of four phases: (1) identification and extraction of articles; (2) preparation for analysis; (3) actual coding; and (4) analysis and write-up.

However, in comparison to other well-established approaches (see, e.g. [12, 14]), the definition of the review scope and an explication of a research agenda are not clearly delineated. We thus incorporate those aspects into our literature review by adding phase (0) in which the scope of the review is described and phase (5) that outlines a research agenda. This synthesized approach is depicted in Figure 1. Considering that our review scope has already been described in the first section, we elaborate the next phases in the following sections.

2.2. Literature search and extraction of articles

To examine existing research, we queried the following databases, providing us with access to a wide range of sources from several computer science disciplines and MIS comprising major journals and conference proceedings: (1) ACM Digital Library; (2) IEEE Xplore Digital Library; (3) EBSCOhost (Business Source Premier Database); and (4) Association for Information Systems Electronic Library (AISel). To focus our review on high-quality MIS literature, we

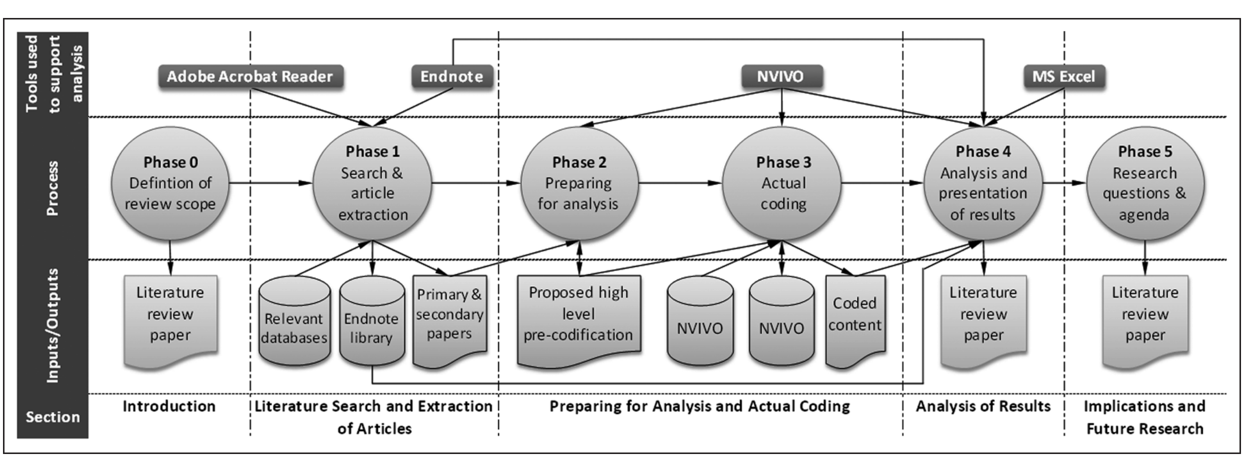

Figure 1. Research approach (cf. [12–14]).

Table 1. Summary of search results and number of analysed articles.

Database	ACM	IEEE	EBSCOhost	AISel	∑
Limiters	Limit to ACM publications	Limit to IEEE and IET publications	Limit to publications of the top 20 MIS journals	Limit to publications of the CAIS, ICIS, ECIS, AMCIS	
Search date	22 August 2011	22 August 2011	22 August 2011	22 August 2011	
Results	45	68	18	0	131
Available for download	44 (−1)	68	18	0	130 (−1)
After analysing abstract	27 (−17)	46 (−22)	13 (−5)	0	86 (−44)
After backward/forward search	40 (+13)	61 (+15)	21 (+8)	0	122 (+36)

restricted the queries to the top 20 MIS journals according to the AIS journal ranking [15] and the proceedings of A-ranked conferences such as ICIS or ECIS according to the ERA conference ranking [16].

Previously performed test queries to the databases have shown that DP specific search terms led to very few or no results. Thus, we queried for the general terms 'digital preservation', 'digital curation' or 'long-term archiving' included in the abstract of the paper and did not restrict the timeframe of the search in order to gain exhaustive results. The search resulted in 131 sources of which we were able to download 130 owing to access restrictions. In a second step we evaluated these 130 papers regarding their relevance to the topic of digital preservation by reading the paper's abstract. This led to a reduction of relevant sources from 130 to 86. In a third step we conducted both backward and forward reference searches [11]. That is, we reviewed the references of the analysed papers (backward) and explored further sources that cite the papers yielded in our search (forward). This step led to an increase of an additional 36 to a total of 122 papers ranging from 1996 to 2011.[2] Table 1 summarizes our search approach and lists the number of results retrieved.

2.3. Preparing for analysis and actual coding

Content analysis can be understood as a scientific research technique to gain 'replicable and valid inferences from text' [17], which enables us to extract trends, characteristics, patterns or concepts out of the extracted sources [18]. One crucial task of the content analysis is to develop categories that capture and structure the essential content in a systematic and pursuable way [19]. In order to analyse the selected publications with respect to our particular review scope, while trying to be as unbiased and close as possible to the contents [20], we used a combined approach by applying deductive high-level categories first. Second, we either inductively developed corresponding subcategories from the analysed content or utilized existing frameworks to structure our findings in a systematic way.

The definition of our high-level categories is derived from an analytical framework based on the work of Dibbern et al. [21], which serves as a pre-codification scheme for content analysis and thus can be conceived as our theoretical 'sensitizing device' guiding the investigation [22]. While conceiving DP as an important and multifaceted decision that an organization deliberately has to make [23], we adapted the framework to the particular context of DP as shown in Table 2. Based on this pre-codification scheme, we analysed all 122 publications by reading their introduction, discussion and conclusion sections [24]. To ease the coding and to increase the ability to systematically 'extract adequate meaning from the underlying data' [25], we used the qualitative data analysis tool 'NVIVO'[3] for content analysis. NVIVO is a commercial software package that allows one to systematically capture, code and analyse literature within one single repository. Various types of data, such as PDF, Microsoft Word or HTML documents, can be imported. This textual data can then be analysed, coded, retrieved, reviewed, recoded and eventually exported. All data within NVIVO is organized

Table 2. Framework applied for analysis of extracted publications.

Perspective	Stages	Research questions
Motivation and drivers	Why?	Why DP?
Stakeholders and scope	Who?	Who are the stakeholders of DP?
	What?	What is preserved?
Requirements	Which?	Which requirements are relevant for DP?
Research methods and artefacts	How?	How does the research community address DP?

Table 3. Drivers for digital preservation.

Dimension	Common positions and various expressions of drivers		#S
Technological	Common position	'[. . .] the rapid changes in technologies, file formats and information systems make the longevity of digital information a challenging problem' [28]	46
	Various expressions	Digital data is more fragile than paper [29] Storage media, hardware and software become obsolete [30]	
Social/cultural	Common position	'We live in a digital world. With an increasing amount of information being created, stored and distributed in digital formats, preservation of digital information is a central concern' [31]	44
	Various expressions	Prevent collective body of knowledge and cultural heritage from vanishing [32] Data must be preserved for the benefit of humankind [33]	
Legal/compliance	Common position	'Regulatory compliance and legal issues require preservation of email archives, medical records, financial accounts, aircraft designs, oil-field data and more' [7]	10
	Various expressions	Sarbanes Oxley Act, HIPPA requires data to be kept over decades [34] Compliance legislations are emerging worldwide [33]	
Learning/re-use	Common position	'The long term value of data rests in their potential as evidence, their reuse possibilities [. . .]' [35]	10
	Various expressions	Support understanding a product, its design and manufacture [36] Preserved data represents a crucial baseline for future science [56]	
Economic/financial	Common position	'Private companies facing discovery orders for their digital information in connection with lawsuits often find that recovering this information costs more than the computer system itself' [37]	8
	Various expressions	Data is increasingly important to economy [38] Data can be unique, irreplaceable and a valuable asset [39]	

in 'Documents' and 'Nodes'. Documents simply represent the textual data that is examined by the researcher. On the other hand, nodes are virtual containers that can be created and organized in branches to store ideas or categories. Nodes capture certain aspects of the data and contain all the information that has been assigned to a specific category during the process of coding [13].

In NVIVO, we coded each finding by separating the underlying data into topics referring to any of the stages of our framework into the corresponding nodes in NVIVO [26]. Having assigned the findings into the respective high-level categories based on the pre-codification scheme, we further conceptually summarized and labelled all items with equivalent or similar meanings to form subcategories [20]. In an iterative approach of reduction and generalization of the findings, we grouped and regrouped items, renamed or removed categories aiming for a manageable number of exhaustive and mutually exclusive categories [17] while linking them to each specific research question stated in Table 2.

3. Findings

3.1. Why digital preservation? (drivers)

Beginning with the question 'why' one would employ DP enables us to explore the various drivers expressing the need for DP and, at the same time, to reason about related threats. To systematically summarize the key drivers of DP, we developed a set of five dimensions drawing from the PEST analysis [27].

Table 3 shows our results along the five dimensions, including a definition of the common position based on a representative citation and various expressions of that driver. In order to assess the dimensions' respective importance, we also indicated the number of sources (#S) in which we identified at least one driver, subsumable under one of the dimensions.

3.2. Who preserves what? (stakeholders and scope)

In this subsection we present the findings related to the stakeholders and scope of DP that are addressed in the analysed articles. We decided to combine these two dimensions in our analysis since we believe that we will thereby acquire more fruitful insights into the area of DP rather than by simply itemizing the relevant stakeholders and types of information that are preserved. In the context of DP we define a stakeholder as any group or individual ('who') that can affect, is affected by or has an interest [40] in preservation of a particular type of digital information ('what'). In a few cases the article did not clearly mention the stakeholder for the preservation of a particular artefact. In these cases we assigned the stakeholders according to the context of the article or the affiliation of the article's authors. The results along the two dimension of stakeholders (who) and type of preserved information (what) are presented in Table 4.

3.3. Which requirements are relevant for digital preservation?

We define a non-functional requirement following Glinz [63, p. 25] as 'an attribute of or a constraint on a system'. Following this definition, the analysis of the non-functional requirements enables us to understand which aspects are crucial for successful digital preservation and hence have to be taken into consideration when dealing with DP. To gather the non-functional requirements, we investigated the yielded papers with respect to words ending with the suffix '-ility' or '-ity' [41] or statements that explicitly express non-functional requirements, for example, 'Engineering Archives require formats that can be accessible for 30-to-50 year lifespans' [36].

We captured 152 items and coded them following our coding approach into a specific node in NVIVO. Figure 2 depicts our findings. With a total of 83 mentions, equalling more than 53% of all captured items, 'authenticity', 'accessibility' and 'integrity' can be considered the most important non-functional requirements for digital preservation. 'Authenticity' means that a preserved object is 'what it purports to be' [42]. 'Accessibility' refers to the need for a continuous and timely access to the digital object [43], and 'integrity' describes the condition that 'data is unchanged from its source and has not been accidently or maliciously modified' [4].

The ability of a digital object to survive a system's lifetime can be understood as the 'reliability' requirement [44], which has been mentioned 18 times. 'Data privacy' shall protect preserved objects' confidentiality from unauthorized access through the application of access controls or encryption mechanisms [43]. 'Trustworthiness' refers to the provider of preservation services who has to demonstrate a proper and auditable preservation process and accountability in order to gain the trust of his consumer [23]. Further, DP has to ensure that the preserved objects are intelligible and usable in their original and intended context by future consumers [7]. Other non-functional requirements that we have captured point to the performance of DP solutions, their scalability [44] and cost-effectiveness [8].

3.4. How does the research community address digital preservation?

Figure 3 provides an overview of the research methods applied in the yielded papers. As 23 of the 122 analysed papers apply more than one research method, and hence are assigned to more than one category, the total number of assignments (*n*) depicted in Figure 3 is 151.

The spectrum of research methods used for categorization was adapted by the authors from the compilation presented by Palvia et al. [45]. The analysis of the applied research methods shows that commentary and argumentative research followed by conceptual modelling/analysis and prototyping clearly dominate the area of DP. Empirical approaches such as field and laboratory experiments, case and field studies, surveys and interviews or action research are less represented. Reference modelling and mathematical models are utilized just twice, respectively once.

The first three categories are considered to emphasize ideas rather than on observations and are intended to be practical advice and recommendations for action. Further, they are recognized to design and describe models or prototype solutions by synthesizing existing knowledge to primarily provide ground for future research. On the other hand the use of empirical approaches suggests that efforts on testing developed conceptual work in a more rigorous setting have been undertaken and hence may be regarded as an indicator of maturity [21, 46]. Based on the assumption that the selection of a research method depends on the current knowledge of a topic [47], our findings indicate that the progress in the area of DP is still at an early stage.

Another perspective, which supports to assess how a particular field approaches inherent issues, lies in the exploration of developed IT artefacts that are regarded as an important element of a specific knowledge base [48]. We present major artefacts that have been created in the area of DP so far and structure our findings based on a classification scheme of artefacts presented by March and Smith [49]. In case of the artefact type 'constructs/concepts', we focus on those that can be regarded as the major antecedents for models, methods and instantiation, although they are mainly expressed in an informal way. The findings, includng a definition of the artefact type (cf. [49], pp. 256–258), DP-specific examples and a reference to the underlying paper, are presented in Table 5.

Table 4. Stakeholders of digital preservation and artefacts in scope of digital preservation.

Who?										
What?	Research, science community		Public interest organizations		Enterprises		Governments		Individuals	
	Example	Source	Example	Source	Example	Source	Example	Source	Example	Source
Scientific and research artefacts	Mathematical simulations Longitudinal studies Earth observation data	S5, S9, S17, S18, S34, S35, S38, S39, S47, S48, S56, S70, S86, S99, S104	Critical scientific data collections Genealogical data Engineering data	S18, S19, S23, S32, S37, S56, S72, S92, S93, S101, S109	Research data	S19, S38 Pharmaceutical research data			Personal medical records	S42
Documents Cultural heritage data	Publications Bibliographies Culturally significant sites Unique historical artefacts	S9, S15, S18, S76 S1, S26, S41, S50, S52, S55, S62, S70, S83, S95, S110	Federal records Census data Historical writings Digitized paintings, sculptures	S18, S45, S72, S92, S93 S17, S18, S19, S22, S27, S38, S39, S74, S122	Business contracts Service agreement	S5, S18, S42, S78, S97, S108	Deed/will records Legislative records	S64, S72, S108	Insurance contracts Personal documents	S68, S77, S99, S115
Multimedia data	Multi-media objects Photographs, movies	S9, S33, S48, S63, S85, S91, S100	Digital photographs Congressional hearings video	S15, S18, S69, S92	Digital photography Motion picture	S15, S18, S19	Legislative proceedings, public meetings recording	S64		
Web content Product related data	Email	S76	Web sites	S19, S71, S75, S79, S80, S81	Web content CAD, design data Manufacturing plans Assembly plans	S105 S3, S19, S38, S71, S72, S98, S99, S108	Web sites, email	S64	Web sites, email blogs	S75, S79, S80, S115
Financial data					Financial accounts Financial records	S18, S19, S99	Fiscal data	S64	Financial contracts Financial/tax data	S18, S46, S77, S115
Software	Programming code Video games Databases	S15, S48, S76	Video games Databases	S14, S15			Transactional databases	S64		

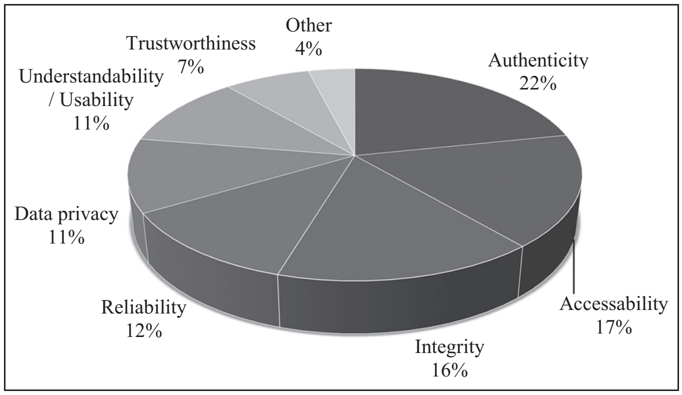

Figure 2. Proportions of non-functional requirements relevant for digital preservation.

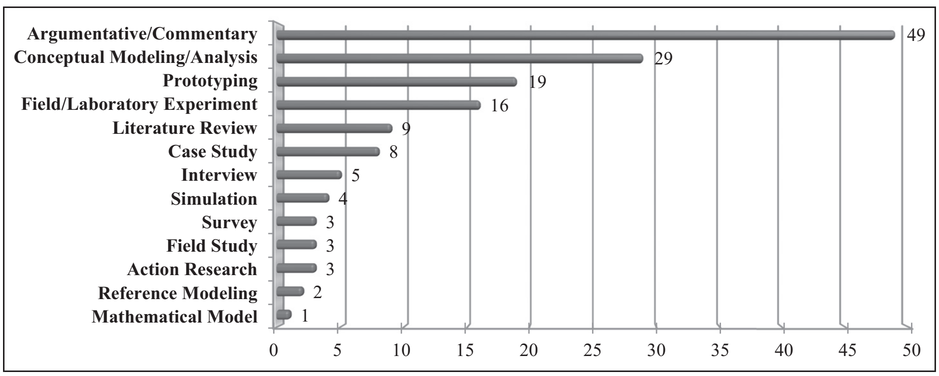

Figure 3. Overview of the applied research methods in the underlying 122 articles.

4. Research agenda outline

4.1. Research questions

In a final step, we analysed the yielded papers with respect to discussed key topics and open issues to identify areas that are still under-researched, and hence, form a good starting point for future research. Towards this end, we focused on the examination of the discussion, conclusion and outlook sections of the underlying papers. Despite the numerous works that have been published to date, our review points towards some salient issues and unsolved problems. While acknowledging those issues, we present some of the many arising research questions that scholars might analyse in Table 6.

To structure our research questions, we assigned them to both topic areas consisting of organizations, people and technology and to common organizational research themes that are often used in MIS research [50, 51]. As can be seen from Table 6, we find various questions that can be assigned to the organizational domain. In line with our findings (see, e.g., Section 3.4), this focus indicates that prior research focused rather on technological aspects and lacks examination of DP in the context of an organization by the means of empirical studies.

Moreover, it has to be noted that questions with regards to the justification of DP from an economical or compliance viewpoint have not yet been investigated. However, in practice, economically legitimated decisions based on clear business cases and documented decision processes towards or against an IT investment play a decisive role [52]. Organizations have to decide what information is worth being sustained and is essential to ensure economic success and compliance or to support innovation. For example, firms are held accountable for the way they preserve, manage and retrieve information which, in turn, has an impact on their risk exposure and legal costs [53]. To be able to make reasonable decisions towards the question ‘what information is worth preserving?’, several requirements and constraints should be considered to prevent the ultimate risk of information loss and its particular consequences for the organization.

Table 5. Artefacts developed in the area of digital preservation.

Artefact	Definition	Examples	References
Construct/concept	Conceptualization used to describe problems within the domain. They may be expressed highly formalized or informal	Risks, threats	S7, S8, S9, S22, S64, S88, S105, S106, S114
		Trust	S18, S43, S46, S60
		Data deluge	S18, S61, S64, S89, S117, S120
		Longevity	S89, S107, S108
Model	Expresses relationships among constructs and represents how things are	Conceptual models	S10, S11, S21, S25, S29, S49, S85, S87, S96, S111, S116, S118
		Analytic models	S5
		Architecture models	S2, S3, S12, S39, S40, S43, S45, S51, S69, S91, S92, S93, S103
Method	A set of steps used to perform a task	Frameworks	S34, S63, S104
		Guidelines, principles	S6, S8, S12, S14, S15, S16, S19, S28, S42, S60, S64, S66, S90, S94, S97, S101, S122
		Algorithms	S23, S50, S62, S118
		Preservation methods	S8, S67, S68, S84, S88, S120, S117
		Standards	S2, S24, S28, S38, S72, S73, S93, S94, S98, S99, S116
Instantiation	Realization of an artefact in its environment. Used to demonstrate feasibility of underlying models and methods	Multimedia illustrations	S1, S26, S41, S52, S53, S55, S119
		File formats	S72, S99, S120
		Prototypes	S4, S9, S13, S24, S30, S31, S40, S47, S70, S71, S73, S100, S102, S103, S115
		Preservation software	S15, S32, S39, S44, S58, S79, S80, S82, S99, S110, S112, S118

4.2. Elements of a research agenda

When endeavouring to respond to the research questions described above, we suggest the development of a reference model following a design science approach [50] as one of the many potential next steps. Reference models are considered beneficial for the design of organizations and application systems and are thought to have both a recommendatory and explanatory character for the selected application area [54]. They are appropriate for serving as universally valid models for the purpose of comparison or the deduction of recommendations for action [55].

While two of the analysed articles apply the method of reference modelling, they focus on architecture models [56, 57]. To the authors' knowledge no DP reference model has been published to date that addresses organizational concerns considering aspects such as costs, risks, decision criteria, roles or compliance issues associated with DP.

Therefore, the proposed research agenda aims to develop a reference model for DP focusing on organizational aspects to explain relevant constructs and their dependencies. Figure 4 illustrates one way of developing a reference model for DP by applying a multi-method approach structured into four major levels: (1) identification/structuring level; (2) construction level; (3) validation/evaluation level; and (4) continuous improvement level.

The first level represents the outset of any academic project in which extant research is reviewed to uncover the 'state of the field' as well as areas and gaps where further research is needed [12]. This step aims to constitute a firm foundation for advancing knowledge and can be achieved by the means of a literature review in support of methods like content analysis 'in which text (notes) are systematically examined by identifying and grouping themes and coding, classifying and developing categories' [45].

The second level comprises the activities in course of the construction of the reference model. This may be based on a combination of deductive and inductive elements, drawing from preliminary considerations, the results of the prior literature review as well as additional library research of both academic and practitioner sources in the pursuit of a high degree of universality as well as linkage to theory [58].

The third level aims to validate and evaluate the constructed reference model with regards to the application domain, focusing on criteria such as relevance, universal applicability and robustness [59]. Validation can be achieved by means of analytical and theoretical research methods, which are primarily based on deductive conclusions drawing from specific theories that are applicable to the specific area. On the other hand empirical examination is suggested, aiming to validate the model based on experiences [60]. Towards this end, we believe case studies and action research to be adequate means

Table 6. Research questions derived from findings.

Area/topic		Research questions	Reference	Potential outcomes
People	Roles	Who is responsible for digital preservation and who is supposed to do it?	S15, S18, S19, S94	Role model
Organizations	Value of IT investment	How can the long-term value of digital information be quantified? How can costs for DP be quantified? What are major cost drivers? How can DP economically be rationalized? Which decision criteria are relevant for a decision towards/against DP of information?	S18, S19, S25, S37, S44, S67, S72, S77, S109	Information value assessment model Cost–benefit assessment DP decision criteria/ model
	Risk/compliance management	How can the risk of information loss owing to technological uncertainty and obsolescence be managed? How can DP be linked into enterprise/ IT risk management and governance processes? How can DP support compliance?	S8, S12, S15, S19, S44, S72, S77, S104, S109	Risk model Alignment of business objectives and DP Compliance model
Technology	Architectures, applications	How can a DP solution be integrated in an existing IT landscape? How can long-term information security be realized? How can a whole business process including its origin context be preserved? Which functional/end-user requirements are relevant for a DP software application? What is the impact of cloud computing on DP?	S18, S43, S72, S99, S101, S104, S114	Architecture model Information security model Software prototype

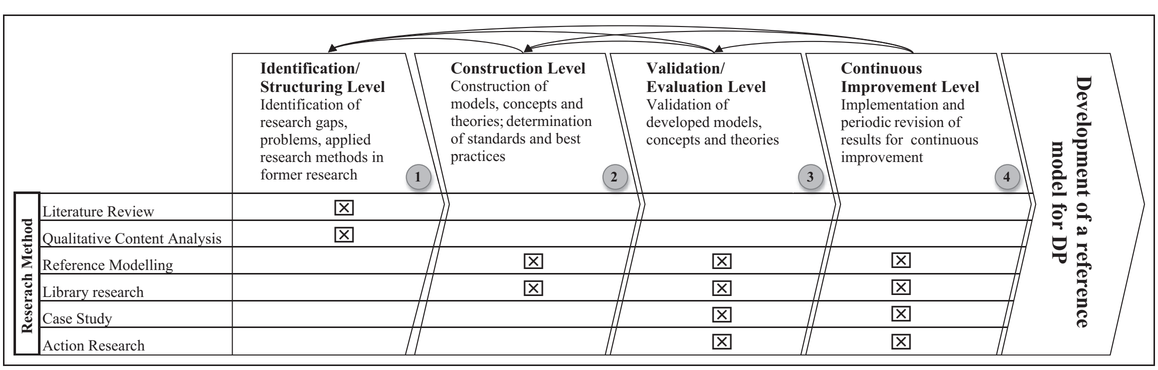

Figure 4. Elements of a research agenda for digital preservation (cf. [58]).

since they are considered particularly appropriate to develop a rich understanding of information systems in the context of complex social organizations and their processes [61, 62].

Based on the evaluation results, the developed reference model can be adjusted and continuously improved by means of additional iterations and, for example, in cooperation with business partners and practitioners. This continuous improvement is represented by the fourth level of our proposed approach and is characterized by periodic evaluation and revision, leading to interrelated processes between the various proposed levels.

5. Conclusion and limitations

The present study was designed to determine to what extent and how DP is currently addressed within the computer science and MIS field by means of a systematic literature review. Our findings show that DP has gained little attention in MIS research compared with the computer science discipline so far. The scientific contribution of this study lies in the elaboration of a set of research questions that have not yet been addressed and the proposal of a research agenda.

We conducted a content analysis on 122 articles from 1996 to 2011 and utilized an analytic framework as well as inductive item grouping to structure our findings. We thoroughly documented our review approach and identified the major drivers, stakeholders, digitally preserved artefacts and requirements for DP. Moreover, we analysed the applied research methods and IT artefacts produced to date. Based on our findings, we contribute a set of research questions that point towards some salient organizational issues in the context of DP and thus provide a set of interesting research opportunities. Finally, we propose an agenda for future research that outlines a multi-method approach for the objective of constructing a reference model for DP. We describe steps for the construction and validation of the reference model and point out adequate research methods. Owing to its explanatory character, a reference model can contribute to improving the understanding of DP in a holistic way, including organizational aspects and, as such might be beneficial for both the research and practitioner communities.

In addition to the theoretical and practical contributions, this review has some general limitations associated with the selected research method. These should be considered when interpreting the results. One limitation of a literature review lies in the paper selection process, which requires judgment on the sources to be included and excluded. However, we tried to minimize this limitation by both following a proven course of action and focusing on the extraction of high-quality literature [11, 12], while providing a detailed description of our search strategy and yielded results.

Another limitation originates from the content analysis and coding of the underlying sources. This process involves cognitive tasks like assigning, summarizing and synthesizing content and as such cannot be completely free of any bias. Furthermore, 16 articles did not clearly mention their applied research method. In these cases, we assigned the article to a particular category, which might be subjective to our interpretation. The same limitation applies to the assignment of stakeholders, which were not always clearly delineated and thus were assigned based on our understanding. However, we tried to mitigate these risks by using NVIVO for coding, which allowed us to analyse the underlying paper in a repeatable way. We cross- and double-checked our coding with built-in functions of NVIVO and by means of several manual reviews. Finally, we had to abstain from referencing every single underlying source in the illustration of our findings to preserve their legibility. Despite these limitations, we believe that this study offers an important step in the exploration of DP and may serve as a starting point for future research.

Acknowledgements

This work has partly been funded by the TIMBUS project, co-funded by the European Union under the 7th Framework Programme for research and technological development and demonstration activities (FP7/2007-2013) under grant agreement no. 269940. The authors are solely responsible for the content of this paper.

Notes

1. We use the terms ‘data’ and ‘information’ interchangeably, understood as any type of ‘information’ that can be digitally stored.
2. The Appendix contains a full listing of all 122 analysed articles. All references to a source in the Appendix are prefixed with ‘S’ (e.g. S23) in this paper to avoid confusion with the citations.
3. A free trial version of NVIVO can be downloaded at: http://www.qsrinternational.com/products_nvivo.aspx.

References

[1] Kooper M, Maes R and Lindgreen E. On the governance of information: Introducing a new concept of governance to support the management of information. *International Journal of Information Management* 2010; 31: 195–200.
[2] Dappert A and Enders M. Digital preservation metadata standards. *Information Standards Quarterly (ISQ)* 2010; 22: 4–13.
[3] Hilbert M and López P. The world’s technological capacity to store, communicate, and compute information. *Science* 2011; 332: 60–65.
[4] Berman F. Got data?: A guide to data preservation in the information age. *Communications of the ACM* 2008; 51: 50–56.
[5] Khatri V and Brown CV. Designing data governance. *Communications of the ACM* 2010; 53: 148–152.
[6] Strodl S, Petrov P, Rauber A et al. *Research on digital preservation within projects co-funded by the European Union in the ICT Programme*, http://www.ifs.tuwien.ac.at/~strodl/paper/Report%20-%20Research%20on%20Digital%20Preservation.pdf (2011, accessed November 27, 2012).

[7] Rabinovici-Cohen S, Baker MG, Cummings R, Fineberg S and Marberg J. Towards SIRF: Self-contained information retention format. In: *Proceedings of the 4th annual international conference on systems and storage (SYSTOR '11)*. Haifa: ACM, 2011.
[8] Lee KH, Slattery O, Lu R, Tang X and McCrary V. The state of the art and practice in digital preservation. *Journal of Research of the National Institute of Standards and Technology* 2002; 107: 93–106.
[9] CCSDS. *Reference model for an open archival information system (OAIS)*, Blue Book. Consultative Committee for Space Data Systems, 2002.
[10] Gladney HM. *Preserving digital information*. New York: Springer, 2007.
[11] Levy Y and Ellis TJ. A systems approach to conduct an effective literature review in support of information systems research. *Informing Science: International Journal of an Emerging Transdiscipline* 2006; 9: 181–212.
[12] Webster J and Watson RT. Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly* 2002; 26: xiii–xxiii.
[13] Bandara W, Miskon S and Fielt E. A systematic, tool-supported method for conducting literature reviews in information systems. In: *Proceedings of the European conference on information systems*, Helsinki, 2011.
[14] Vom Brocke J, Simons A, Niehaves B, Reimer K, Plattfaut R and Cleven A. Reconstructing the giant: On the importance of rigour in documenting the literature search process. In: *Proceedings of the European conference on information systems*, Verona, 2009.
[15] Association for Information Systems, *MIS journal rankings*, http://ais.affiniscape.com/displaycommon.cfm?an=1&subarticlenbr=432 (2012, accessed December 2, 2012).
[16] Excellence in Research for Australia (ERA), *ERA – ranked conference list*, http://www.arc.gov.au/xls/ERA2010_conference_list.xls (2010, accessed November 30, 2012).
[17] Krippendorff K. *Content analysis: An introduction to its methodology*, 2nd edn. Thousand Oaks, CA: Sage, 2004.
[18] Steininger D, Huntgeburth J and Veit D. A systemizing research framework for Web 2.0. In: *Proceedings of the European conference on information systems*, Helsinki, 2011.
[19] Mayring P. Qualitative content analysis. *Forum: Qualitative Social Research* 2000; 1: 1–10.
[20] Corbin JM and Strauss A. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology* 1990; 13: 3–21.
[21] Dibbern J, Goles T, Hirschheim R and Jayatilaka B. Information systems outsourcing: A survey and analysis of the literature. *ACM SIGMIS Database* 2004; 35: 6–102.
[22] Klein HK and Myers MD. A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly* 1999; 23: 67–93.
[23] Becker C and Rauber A. Preservation decisions: Terms and conditions apply. In: *Proceedings of the 11th annual international ACM/IEEE joint conference on digital libraries*. Ottawa: ACM, 2011, pp. 67–76.
[24] Swanson EB and Ramiller NC. Information systems research thematics: Submissions to a new journal, 1987–1992. *Information Systems Research* 1993; 4: 299–330.
[25] Leech NL and Onwuegbuzie AJ. An array of qualitative data analysis tools: A call for data analysis triangulation. *School Psychology Quarterly* 2007; 22: 557.
[26] Urquhart C. An encounter with grounded theory: Tackling the practical and philosophical issues. In: Trauth EM (ed.) *Qualitative research in IS: Issues and trends*. Hershey, PA: IGI Global, 2001, pp. 104–140.
[27] Fahey L and Narayanan V. *Macroenvironmental analysis for strategic management*. St Paul, MN: West, 1986.
[28] Becker C, Rauber A, Heydegger V, Schnasse J and Thaller M. A Generic XML language for characterising objects to support digital preservation. In: *Proceedings of the ACM symposium on applied computing (SAC '08)*. Fortaleza: ACM, 2008, pp. 402–406.
[29] Eastwood T. Appraising digital records for long-term preservation. *Data Science Journal* 2004; 3: 202–208.
[30] Gladney HM and Lorie RA. Trustworthy 100-year digital objects: Durable encoding for when it's too late to ask. *ACM Transactions on Information Systems* 2005; 23: 299–324.
[31] Hart PE and Liu Z. Trust in the preservation of digital information. *Communications of the ACM* 2003; 46: 93–97.
[32] Mikeal A, Green C, Maslov A, Phillips S and Leggett J. Preserving the scholarly side of the web. In: *Proceedings of the 4th Latin American Web congress (LA-Web '06)*. New York: IEEE, 2006, pp. 162–171.
[33] Factor M, Naor D, Rabinovici-Cohen S et al. Preservation datastores: Architecture for preservation aware storage. In: *Proceedings of the 24th IEEE conference on mass storage systems and technologies (MSST 2007)*. New York: IEEE, 2007, pp. 3–15.
[34] Baker M, Shah M, Rosenthal DSH, et al. A fresh look at the reliability of long-term digital storage. In: *Proceedings of the 1st ACM SIGOPS/EuroSys European conference on computer systems (EuroSys '06)*. New York: ACM, 2006, pp. 221–34.
[35] Rusbridge C, Burnhill P, Ross S, et al. The digital curation centre: A vision for digital curation. In: *Proceedings of the IEEE international symposium on mass storage systems and technology 2005*. New York: IEEE, 2005, pp. 31–41.
[36] Regli WC, Grauer M and Kopena JB. A framework for preservable geometry-centric artifacts. In: *Proceedings of the SIAM/ACM joint conference on geometric and physical modeling (SPM '09)*. San Francisco, CA: ACM, 2009.
[37] Chen S-S. The paradox of digital preservation. *Computer* 2001; 34: 24–28.

[38] Jianhai R and McDonough JP. Preserving born-digital cultural heritage in virtual world. In: *Proceedings of the IEEE international symposium on IT in medicine education (ITIME '09)*. New York: IEEE, 2009, pp. 745–748.
[39] Hedges M, Hasan A and Blanke T. Curation and preservation of research data in an iRODS data grid. In: *Proceedings of the IEEE international conference on e-science and grid computing*. New York: IEEE, 2007, pp. 457–464.
[40] Freeman RE. *Strategic management: A stakeholder approach*. Boston, MA: Pitman, 1984.
[41] Chung L and do Prado Leite J. On non-functional requirements in software engineering conceptual modeling: Foundations and applications. In: Borgida A, Chaudhri V, Giorgini P and Yu E (eds) *Conceptual modeling: Foundations and applications*. Berlin: Springer, 2009, pp. 363–379.
[42] Muir A. Legal deposit and preservation of digital publications: A review of research and development activity. *Journal of Documentation* 2001; 57: 652–682.
[43] Storer MW, Greenan K and Miller EL. Long-term threats to secure archives. In: *Proceedings of the 2nd ACM workshop on storage security and survivability (StorageSS '06)*. Alexandria, VA: ACM, 2006.
[44] Barateiro J, Antunes G, Freitas F and Borbinha J. Designing digital preservation solutions: A risk management-based approach. *International Journal of Digital Curation* 2010; 5: 4–17.
[45] Palvia P, Pinjani P and Sibley EH. A profile of information systems research published in information & management. *Information & Management* 2007; 44: 1–11.
[46] Alavi M and Carlson P. A review of MIS research and disciplinary development. *Journal of Management Information Systems* 1992; 8: 45–62.
[47] Benbasat I, Goldstein DK and Mead M. The case research strategy in studies of information systems. *MIS Quarterly* 1987; 11: 369–86.
[48] Hevner A and Chatterjee S. Design science research in information systems. in: hevner a and chatterjee s (eds) *design Research in Information Systems*. Boston, MA: Springer, 2010, pp. 9–22.
[49] March ST and Smith GF. Design and natural science research on information technology. *Decision Support Systems* 1995; 15: 251–266.
[50] Hevner AR, March ST, Park J and Ram S. Design science in information systems research. *MIS Quarterly* 2004; 28: 75–105.
[51] Sidorova A, Evangelopoulos N, Valacich JS and Ramakrishnan T. Uncovering the intellectual core of the information systems discipline. *MIS Quarterly* 2008; 32: 467–482.
[52] Baskerville R. Information systems security design methods: Implications for information systems development. *ACM Computing Surveys (CSUR)* 1993; 25: 375–414.
[53] Volonino DL, Sipior JC and Ward BT. Managing the lifecycle of electronically stored information. *Information Systems Management* 2007; 24: 231–238.
[54] Fettke P and Loos P. Classification of reference models: A methodology and its application. *Information Systems and E-Business Management* 2003; 1: 35–53.
[55] Thomas O. Understanding the term reference model in information systems research: History, literature analysis and explanation business process management workshops. In: Bussler C and Haller A (eds) Berlin: Springer, 2006, pp. 484–496.
[56] Antunes G, Barateiro J and Borbinha J. A reference architecture for digital preservation. In: *Proceedings of the 7th international conference on preservation of digital objects (iPRES2010)*, Vienna, 2010.
[57] Huhnlein D, Korte U, Langer L and Wiesmaier A. A comprehensive reference architecture for trustworthy long-term archiving of sensitive data. In: *Proceedings of the 3rd international conference on new technologies, mobility and security (NTMS)*. Cairo: IEEE, 2009, pp. 1–5.
[58] Martens B and Teuteberg F. Why risk management matters in IT outsourcing – A systematic literature review and elements of a research agenda. In: *Proceedings of the European conference on information systems*, Verona, 2009.
[59] Mišic VB and Zhao JL. Evaluating the quality of reference models. In: *Proceedings of the 19th international conference on conceptual modeling ER'00*. Salt Lake City, UT: Springer, 2000, pp. 484–98.
[60] Fettke P and Loos P. Ontological evaluation of reference models using the Bunge–Wand–Weber Model. In: *Proceedings of the Americas conference on information systems (AMCIS)*, Tampa, FL, 2003.
[61] Baskerville R and Myers MD. Special issue on action research in information systems: Making IS research relevant to practice. *MIS Quarterly* 2004; 28: 329–335.
[62] Yin RK. *Case study research: Design and methods*. Los Angeles, CA: Sage, 2009.
[63] Glinz M. On non-functional requirements. In: *Proceedings of the 15th IEEE international requirements engineering conference (RE '07)*. New Delhi: IEEE, 2007, pp. 21–26.

Appendix

Table A1. List of all 122 analysed articles.

No.	Reference
S1	Andrade BT, Bellon ORP, Silva L and Vrubel A. Enhancing color texture quality of 3D models for digital preservation of indigenous ceramic artworks. In: *Proceedings of the IEEE 12th international conference on computer vision workshops*, 2009, pp. 980–987.
S2	Antunes G, Barateiro J, Becker C, Borbinha J and Vieira R. Modeling contextual concerns in enterprise architecture. In: *Proceedings of the 15th IEEE international enterprise distributed object computing conference workshops (EDOCW 2011)*, Helsinki, 2011, pp. 3–10.
S3	Antunes G, Barateiro J and Borbinha J. A reference architecture for digital preservation. In: *Proceedings of the 7th international conference on preservation of digital objects (iPRES2010)*, Vienna, 2010.
S4	Antunes G, Barateiro J, Cabral M, Borbinha J and Rodrigues R. Preserving digital data in heterogeneous environments. In: *Proceedings of the 9th ACM/IEEE-CS joint conference on digital libraries (JCDL '09)*. New York: ACM, 2009, pp. 345–348.
S5	Baker M, Shah M, Rosenthal DSH, et al. A fresh look at the reliability of long-term digital storage. In: *Proceedings of the 1st ACM SIGOPS/EuroSys European conference on computer systems (EuroSys '06)*. New York: ACM, 2006, pp. 221–234.
S6	Barateiro J, Antunes G and Borbinha J, *Addressing digital preservation: Proposals for new perspectives*, http://cs.harding.edu/indp/papers/barateiro7.pdf (2009, accessed September 25, 2011).
S7	Barateiro J, Antunes G and Borbinha J. Long-term security of digital information: Assessment through risk management and enterprise architecture. In: *Proceedings of the IEEE international conference on computer as a tool (EUROCON 2011)*, 2011, pp. 1–4.
S8	Barateiro J, Antunes G, Freitas F and Borbinha J. Designing digital preservation solutions: A risk management-based approach. *International Journal of Digital Curation* 2010; 5: 4–17.
S9	Barateiro J, Borbinha J, Antunes G and Freitas F. Challenges on preserving scientific data with data grids. In: *Proceedings of the 1st ACM workshop on data grids for eScience (DaGreS '09)*. New York: ACM, 2009, pp. 17–22.
S10	Beagrie N and Carpenter L. Development of digital preservation environments by the UK Joint Information Systems Committee (JISC). In: *Proceedings of the IEEE international symposium on mass storage systems and technology*, 2005, pp. 74–77.
S11	Becker C, Antunes G, Barateiro J, Vieira R and Borbinha J. Modeling digital preservation capabilities in enterprise architecture. In: *Proceedings of the 12th annual international conference on digital government research (dgo 2011)*, Maryland, 2011, pp. 84–93.
S12	Becker C, Barateiro J, Antunes G, Borbinha J and Vieira R. On the relevance of enterprise architecture and IT governance for digital preservation. In: Janssen M, Scholl H, Wimmer M and Tan Y-h (eds) *Electronic government*. Berlin: Springer, 2011, pp. 332–344.
S13	Becker C, Kulovits H, Rauber A and Hofman H. Plato: A service oriented decision support system for preservation planning. In: *Proceedings of the 8th ACM/IEEE-CS joint conference on digital libraries (JCDL '08)*. New York: ACM, 2008, pp. 367–370.
S14	Becker C and Rauber A. Requirements modelling and evaluation for digital preservation: A COTS selection method based on controlled experimentation. In: *Proceedings of the ACM Symposium on Applied Computing (SAC '09)*. New York: ACM, 2009, pp. 401–402.
S15	Becker C and Rauber A. Preservation decisions: Terms and conditions apply. In: *Proceedings of the 11th annual international ACM/IEEE joint conference on digital libraries (JCDL '11)*. New York: ACM, 2011, pp. 67–76.
S16	Becker C, Rauber A, Heydegger V, Schnasse J and Thaller M. A Generic XML language for characterising objects to support digital preservation. In: *Proceedings of the ACM symposium on applied computing (SAC '08)*. Fortaleza: ACM, 2008, pp. 402–406.
S17	Berman F. One hundred years of data. In: *Proceedings of the international conference on digital government research (dgo '06)*. New York: ACM, 2006, pp. 3–4.
S18	Berman F. Got data?: A guide to data preservation in the information age. *Communications of the ACM* 2008; 51: 50–56.
S19	Blue Ribbon Task Force, *Sustainable economics for a digital planet: Ensuring long-term access to digital information – final report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access*, http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf (2010, accessed September 21, 2011).
S20	Borbinha J. It is the time for the digital library to meet the enterprise architecture. In: Goh D, Cao T, Soelvberg I and Rasmussen E (eds) *Asian digital libraries: Looking back 10 years and forging new frontiers*. Berlin: Springer, 2007, pp. 176–85.
S21	Brocks H, Kranstedt A, Jäschke G and Hemmje M. Modeling context for digital preservation. In: Szczerbicki E and Nguyen N (eds) *Smart information and knowledge management – Advances, challenges, and critical issues*. Berlin: Springer, 2010, pp. 197–226.
S22	Campisi P, Maiorana E, Teri E and Neri A. Challenges to long term digital preservation: A glimpse of Italian experience. In: *Proceedings of the 16th international conference on digital signal processing*, 2009, pp. 1–8.
S23	Cartledge CL and Nelson ML. Unsupervised creation of small world networks for the preservation of digital objects. In: *Proceedings of the 9th ACM/IEEE-CS joint conference on digital libraries (JCDL '09)*. New York: ACM, 2009, pp. 349–352.

(continued)

Table A1. (continued)

No.	Reference
S24	Chang F-C, Chang C-Y and Hang H-M. A study on the meta-data design for long-term digital multimedia preservation. In: *Proceedings of the international conference on intelligent information hiding and multimedia signal processing (IIHMSP '08)*, 2008, pp. 95–8.
S25	Chen S-S. The paradox of digital preservation. *Computer* 2001; 34: 24–28.
S26	Cheng G, Jia Z, Lu X and Li A. Key technology in three-dimensional modeling for cultural heritages. In: *Proceedings of the international conference on information engineering and computer science (ICIECS 2009)*, 2009, pp. 1–4.
S27	Da N, Han R and Tang J. Study on system and methods of digital preservation and renovation of shadow play in Liaoning Province. In: *Proceedings of the international conference on system science, engineering design and manufacturing informatization (ICSEM)*, 2010, p 200–3.
S28	Dappert A and Enders M. Digital preservation metadata standards. *Information Standards Quarterly (ISQ)* 2010; 22: 4–13.
S29	Dappert A and Farquhar A. Modelling organizational preservation goals to guide digital preservation. *International Journal of Digital Curation* 2009; 4: 119–134.
S30	de la Rosa JL, Trias A, Ruusalepp R et al. Agents that supply knowledge exchange in long-term digital preservation. *eChallenges* 2010: 1–9.
S31	de la Rosa JL, Trias A, Ruusalepp R, et al. Agents for social search in long-term digital preservation. In: *Proceedings of the 6th international conference on semantics knowledge and grid (SKG)*, 2010, pp. 363–366.
S32	Dong H, Yu S and Jiang Y. Knowledge representation of Chinese genealogical record of VIPs in KMT and CPC. In: *Proceedings of the 9th international conference on hybrid intelligent systems (HIS '09)*, 2009, pp. 116–120.
S33	Dougherty MT, Folk MJ, Zadok E, et al. Unifying biological image formats with HDF5. *Communications of the ACM* 2009; 52: 42–47.
S34	Doyle J, Viktor HL and Paquet E. Long term digital preservation – An end user's perspective. In: *Proceedings of the 2nd international conference on digital information management (ICDIM '07)* 2007, pp. 146–151.
S35	Duerr RE, Cao P, Crider J, Folk M, Lynnes C and Yang MQ. Ensuring long-term access to remotely sensed data with layout maps. *IEEE Transactions on Geoscience and Remote Sensing* 2009; 47: 123–129.
S36	Duranti L. The long-term preservation of accurate and authentic digital data: The INTERPARES Project. *Data Science Journal* 2005; 4: 106–118.
S37	Eastwood T. Appraising digital records for long-term preservation. *Data Science Journal* 2004; 3: 202–208.
S38	Factor M, Naor D, Rabinovici-Cohen S, Ramati L, Reshef P and Satran J. The need for preservation aware storage: A position paper. *ACM SIGOPS Operating Systems Review* 2007; 41: 19–23.
S39	Factor M, Naor D, Rabinovici-Cohen S, et al. Preservation datastores: Architecture for preservation aware storage. In: *Proceedings of the 24th IEEE conference on mass storage systems and technologies (MSST 2007)* 2007, pp. 3–15.
S40	Ferreira M, Baptista AA and Ramalho JC. An intelligent decision support system for digital preservation. *International Journal on Digital Libraries* 2007; 6: 295–304.
S41	Geller T. Rebuilding for eternity. *Communications of the ACM* 2010; 53: 19–21.
S42	Gladney H. Principles for digital preservation. *Communications of the ACM* 2006; 49: 111–116.
S43	Gladney HM. Trustworthy 100-year digital objects: Evidence after every witness is dead. *ACM Transactions on Information Systems* 2004; 22: 406–436.
S44	Gladney HM and Lorie RA. Trustworthy 100-year digital objects: Durable encoding for when it's too late to ask. *ACM Transactions on Information Systems* 2005; 23: 299–324.
S45	Goth G. Long-term data preservation efforts ramping up. *IEEE Software* 2002; 19: 98–102.
S46	Hart PE and Liu Z. Trust in the preservation of digital information. *Communications of the ACM* 2003; 46: 93–97.
S47	Hedges M, Hasan A and Blanke T. Management and preservation of research data with iRODS. In: *Proceedings of the first ACM workshop on cyberinfrastructure: Information management in eScience (CIMS '07)*. New York: ACM, 2007, pp. 17–22.
S48	Hedges M, Hasan A and Blanke T. Curation and preservation of research data in an iRODS data grid. In: *Proceedings of the IEEE international conference on e-Science and grid computing* 2007, pp. 457–464.
S49	Higgins S. The DCC curation lifecycle model. *International Journal of Digital Curation* 2008; 3: 134–140.
S50	Hu S-x, Zha H-b and Zhang A-w. Real 3D digital method for large-scale cultural heritage sites. In: *Proceedings of the 9th international conference on information visualisation*, 2005, pp. 503–508.
S51	Huhnlein D, Korte U, Langer L and Wiesmaier A. A comprehensive reference architecture for trustworthy long-term archiving of sensitive data. In: *Proceedings of the 3rd international conference on new technologies, mobility and security (NTMS)* 2009, pp. 1–5.
S52	Ikeuchi K, Nakazawa A, Hasegawa K and Ohishi T. Representing cultural heritage in digital forms for VR systems through computer vision techniques. In: *Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004)*, 2004, pp. 1–6.
S53	Ikeuchi K, Nakazawa A, Nishino K and Oishi T. Creating virtual Buddha statues through observation. In: *17th international conference on computer vision and pattern recognition workshop (CVPRW '03)*, 2003.
S54	Inouye A, Hedstrom M, Flecker D and Levy D. A digital strategy for the Library of Congress. *Communications of the ACM* 2001; 44: 43.

(continued)

Table A1. (continued)

No.	Reference
S55	Izani M, Bridges A, Razak A, Calder B and Grant M. Digital preservation of a Famosa Fortress in Malaysia. In: *Proceedings of the 2nd workshop on eheritage and digital art preservation (eHeritage '10)*. New York: ACM, 2010, pp. 7–12.
S56	Jacobs JA and Humphrey C. Preserving research data. *Communications of the ACM* 2004; 47: 27–29.
S57	Jianhai R and McDonough JP. Preserving born-digital cultural heritage in virtual world. In: *Proceedings of the IEEE international symposium on IT in medicine education (ITIME '09)*. New York: IEEE, 2009, pp. 745–8.
S58	Jin X, Jiang J and Min G. Automatic digital preservation solutions enabled by Web services amp and intelligent agents. *eChallenges* 2010: 1–8.
S59	Jordan C, McDonald RH, Minor D and Kozbial A. Cyberinfrastructure collaboration for distributed digital preservation. In: *Proceedings of the 4th IEEE international conference on eScience (eScience '08)*, 2008, pp. 408–409.
S60	Kaczmarek JS, Habing TG and Eke J. Repository software evaluation using the audit checklist for certification of trusted digital repositories. In: *Proceedings of the 6th ACM/IEEE-CS joint conference on digital libraries (JCDL '06)*, 2006, pp. 107–108.
S61	Kanipe J. Modeling the astronomical. *Communications of the ACM* 2010; 53: 13–15.
S62	Kumar S, Snyder D, Duncan D, Cohen J and Cooper J. Digital preservation of ancient cuneiform tablets using 3D-scanning. In: *Proceedings of the 4th International Conference on 3-D Digital Imaging and Modeling (3DIM)*, 2003, pp. 326–333.
S63	Kuo T-Y and Hsiao W-H. A novel preservation framework for digital image archiving system. In: *Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP '08)*, 2008, pp. 760–763.
S64	Kwon H, Pardo TA and Burke GB. Building a state government digital preservation community: Lessons on interorganizational collaboration. In: *Proceedings of the international conference on digital government research (dgo '06)*. New York: ACM, 2006, pp. 277–284.
S65	Lawrence GW, Kehoe WR, Rieger OY and Walters WH. *Risk management of digital information: A file format investigation*. Washington, DC: Council on Library and Information Resources, 2000.
S66	Lee CA, Tibbo HR and Schaefer JC. Defining what digital curators do and what they need to know: The DigCCurr Project. In: *Proceedings of the 7th ACM/IEEE-CS joint conference on digital libraries (JCDL '07)*. New York: ACM, 2007, pp. 49–50.
S67	Lee KH, Slattery O, Lu R, Tang X and McCrary V. The state of the art and practice in digital preservation. *Journal of Research of the National Institute of Standards and Technology* 2002; 107: 93–106.
S68	Levy DM. Heroic Measures: Reflections on the possibility and purpose of digital preservation. In: *Proceedings of the 3rd ACM conference on digital libraries (DL '98)*. New York: ACM, 1998, pp. 152–161.
S69	Li C, Xing C, Huang MB, Dong L and Zhang M. A practice of Web-based sustainable preservation risk management systems for digital libraries in China. In: *Proceedings of the 7th Web information systems and applications conference (WISA)*, 2010, pp. 69–72.
S70	Lindley A, Jackson AN and Aitken B. A collaborative research environment for digital preservation – The planet's testbed. In: *Proceedings of the 19th IEEE international workshop on enabling technologies 2010: Infrastructures for collaborative enterprises (WETICE)*, 2010, pp. 197–202.
S71	Lorie RA. A methodology and system for preserving digital data. In: *Proceedings of the 2nd ACM/IEEE-CS joint conference on digital libraries (JCDL '02)*. New York: ACM, 2002, pp. 312–319.
S72	Lubell J, Subrahmanian E, Regli W and Sudarsan R, *Long term knowledge retention workshop summary – report 7386*, http://www.ukoln.ac.uk/events/ltkr-2007/ltkr-workshop-summary%202006.pdf (2007, accessed September 23, 2011).
S73	Ma N, Li C, Jiang A and Xing C. Design and implementation of open source based digital preservation experimental platform (THDP). In: *Proceedings of the 9th international conference for young computer scientists (ICYCS 2008)*, 2008, pp. 959–964.
S74	Mannoni B. Bringing museums online. *Communications of the ACM* 1996; 39: 100–105.
S75	Marshall CC. How people manage personal information over a lifetime. In: Jones WP and Teevan J (eds) *Personal information management*. Seattle, WA: University of Washington Press, 2007, pp. 57–75.
S76	Marshall CC. From writing and analysis to the repository: Taking the scholars' perspective on scholarly archiving. In: *Proceedings of the 8th ACM/IEEE-CS joint conference on digital libraries (JCDL '08)*. New York: ACM, 2008, pp. 251–260.
S77	Marshall CC, Bly S and Brun-Cottan F. The long term fate of our digital belongings: Toward a service model for personal archives. In: *Proceedings of the IS&T archiving 2006*. Ottawa: Society for Imaging Science and Technology, 2007.
S78	Marshall CC and Golovchinsky G. Saving private hypertext: Requirements and pragmatic dimensions for preservation. In: *Proceedings of the 15th ACM conference on hypertext and hypermedia (HYPERTEXT '04)*. New York: ACM, 2004, pp. 130–138.
S79	McCown F, Marshall CC and Nelson ML. Why web sites are lost (and how they're sometimes found). *Communications of the ACM* 2009; 52: 141–145.
S80	McCown F and Nelson ML. Lazy preservation and Warrick. *Communications of the ACM* 2009; 52: 142.
S81	McCown F, Smith JA and Nelson ML. Lazy preservation: Reconstructing websites by crawling the crawlers. In: *Proceedings of the 8th annual ACM international workshop on Web information and data management (WIDM '06)*. New York: ACM, 2006, pp. 67–74.

(continued)

Table A1. (continued)

No.	Reference
S82	McHugh A, Ruusalepp R, Seamus R and Hofman H *DCC and DPE digital repository audit method based on risk assessment (DRAMBORA)*, http://www.repositoryaudit.eu/download/ (2007, accessed September 25, 2011).
S83	Mendes CM, Drees DR, Silva L and Bellon OR. Interactive 3D visualization of natural and cultural assets. In: *Proceedings of the 2nd workshop on eHeritage and digital art preservation (eHeritage '10)*. New York: ACM, 2010, pp. 49–54.
S84	Mikeal A, Green C, Maslov A, Phillips S and Leggett J. Preserving the scholarly side of the Web. In: *Proceedings of the 4th Latin American Web congress (LA-Web '06)*, 2006, pp. 162–171.
S85	Mikroyannidis A, Ong B, Ng K and Giaretta D. Ontology-based temporal modelling of provenance information. In: *Proceedings of the 14th IEEE Mediterranean electrotechnical conference (MELECON 2008)*, 2008, pp. 176–181.
S86	Miller SP, Detrick RS and Helly J. DIGARCH project highlights multi-institutional testbed for scalable digital archiving. In: *Proceedings of the international conference on digital government research (dgo '06)*. New York: ACM, 2006, pp. 289–290.
S87	Mois M, Klas C-P and Hemmje ML. Digital preservation as communication with the future. In: *Proceedings of the 16th international conference on digital signal processing*, 2009, pp. 1–8.
S88	Muir A. Legal deposit and preservation of digital publications: A review of research and development activity. *Journal of Documentation* 2001; 57: 652–682.
S89	National Science Board. *Long-lived digital data collections: Enabling research and education in the 21st century*, www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf (2005, accessed September 28, 2011).
S90	nestor Arbeitsgruppe Vertrauenswürdige Archive. *nestor-Kriterien – Kriterienkatalog vertrauenswürdige digitale Langzeitarchive – Version 2*, http://edoc.hu-berlin.de/docviews/abstract.php?id=29332 (2008, accessed September 28, 2011).
S91	Ng K, Mikroyannidis A, Ong B and Giaretta D. Practicing ontology modelling for preservation of interactive multimedia performances. In: *Proceedings of the international conference on automated solutions for cross media content and multi-channel distribution (AXMEDIS '08)*, 2008, pp. 276–281.
S92	Nguyen QL, Lake A and Huber M. Content server architecture pattern for evolvability and scalability. In: *Proceedings of the 5th international conference on software engineering advances (ICSEA)* 2010, pp. 149–154.
S93	Nguyen QL, Lake A and Huber M. Evolvable and scalable system of content servers for a large digital preservation archive. In: *Proceedings of the 4th annual IEEE systems conference*, 2010, pp. 306–310.
S94	Nicholson D and Dobreva M. Beyond OAIS: Towards a reliable and consistent digital preservation implementation framework. In: *Proceedings of the 16th international conference on digital signal processing*, 2009, pp. 1–8.
S95	Pasko G, Pasko A, Vilbrandt C and Ikedo T. Virtual Shikki and Sazaedo: Shape modeling in digital preservation of Japanese lacquer ware and temples. In: *Proceedings of the spring conference on computer graphics*, 2001, pp. 147–154.
S96	Perez RF, Perez O, Portela O, et al. Evaluation model for long term data archiving systems in the context of Earth observation. In: *Proceedings of the 27th IEEE symposium on mass storage systems and technologies (MSST)*, 2011, pp. 1–5.
S97	Porekar J, Saljic S, Klobucar T and Jerman-Blazic A. Technical patterns for long term trusted archiving. In: *Proceedings of the 3rd international conference on digital society (ICDS '09)*, 2009, pp. 241–246.
S98	Project LOTAR. *Long term archiving and retrieval of product data within the aerospace industry (LOTAR) – technical aspects of an approach for application – White Paper*, http://www.prostep.org/fileadmin/freie_downloads/WhitePaper-Studien/ProSTEP-iViP_White-Paper_LOTAR_1.0.pdf (2002, accessed September 28, 2011).
S99	Rabinovici-Cohen S, Baker MG, Cummings R, Fineberg S and Marberg J. Towards SIRF: Self-contained information retention format. In: *Proceedings of the 4th annual international conference on systems and storage (SYSTOR '11)*. New York: ACM, 2011.
S100	Rajasekar A, Marciano R, Moore R, et al. Building a demonstration prototype for the preservation of large-scale multimedia collections. In: *Proceedings of the international conference on digital government research (dgo '06)*. New York: ACM, 2006, pp. 287–288.
S101	Rank R. Enterprise IT Support for NOAA archives. In: *Proceedings of the IEEE international geoscience and remote sensing symposium (IGARSS 2007)*, 2007, pp. 4025–4028.
S102	Rawassizadeh R. A holistic multipurpose life-log framework. In: *Proceedings of the 12th ACM international conference adjunct papers on ubiquitous computing (Ubicomp '10)*. New York: ACM, 2010, pp. 481–484.
S103	Rechert K, von Suchodoletz D and Welte R. Emulation based services in digital preservation. In: *Proceedings of the 10th annual joint conference on digital libraries (JCDL '10)*. New York: ACM, 2010, pp. 365–368.
S104	Regli WC, Grauer M and Kopena JB. A framework for preservable geometry-centric artifacts. In: *Proceedings of the SIAM/ACM joint conference on geometric and physical modeling (SPM '09)*. New York: ACM, 2009.
S105	Rosenthal DSH. Keeping bits safe: How hard can it be? *Communications of the ACM* 2010; 53: 47–55.
S106	Rosenthal DSH. Author's response. *Communications of the ACM* 2011; 44: 7.
S107	Rothenberg J. Ensuring the longevity of digital information. *International Journal of Legal Information* 1998; 26: 1–22.
S108	Rothenberg J. *Avoiding technological quicksand: Finding a viable technical foundation for digital preservation – A report to the Council on Library and Information Resources*. Washington, DC: Council on Library and Information Resources, 1999.
S109	Rusbridge C, Burnhill P, Ross S, et al. The digital curation centre: A vision for digital curation. In: *Proceedings of the international symposium on mass storage systems and technology*, 2005, pp. 31–41.

(continued)

Table A1. (continued)

No.	Reference
S110	Schmidt R, Sadilek C and King R. A service for data-intensive computations on virtual clusters. In: *Proceedings of the first international conference on intensive applications and services (INTENSIVE '09)*, 2009, pp. 28–33.
S111	Schott M, Dittmann J, Vielhauer C, Krätzer C and Lang A. Integrity and authenticity for digital long-term preservation in iRODS grid infrastructure. In: *Proceedings of the 6th international workshop for technical, economic and legal aspects of business models for virtual goods incorporating the 4th international ODRL workshop*, 2008, pp. 90–104.
S112	Seegraef C, Lochter F, Lesch L and Wachter J. Clearinghouse for seismic data. In: *Proceedings of the IEEE international geoscience and remote sensing symposium 1999 (IGARSS '99)*, 1999, pp. 1164–1166.
S113	Spring AP, Peters C and Minns T. Using mid-range laser scanners to digitize cultural-heritage sites. *IEEE Computer Graphics and Applications* 2010; 30: 15–19.
S114	Storer MW, Greenan K and Miller EL. Long-term threats to secure archives. In: *Proceedings of the 2nd ACM workshop on storage security and survivability (StorageSS '06)*, 2006, pp. 9–16.
S115	Strodl S, Motlik F, Stadler K and Rauber A. Personal & SOHO archiving. In: *Proceedings of the 8th ACM/IEEE-CS joint conference on digital libraries (JCDL '08)*. New York: ACM, 2008, pp. 115–123.
S116	Strodl S and Rauber A. Preservation planning in the OAIS model. *New Technology of Library and Information Service* 2008; 1: 61–68.
S117	Thibodeau K. Overview of technological approaches to digital preservation and challenges in coming years. *The State of Digital Preservation: An International Perspective* 2002: 4–32.
S118	Vignatti T, Bona LCE, Sunye MS and Vignatti AL. Long-term digital archiving based on selection of repositories over P2P networks. In: *Proceedings of the 9th IEEE international conference on peer-to-peer computing (P2P '09)*, 2009, pp. 194–203.
S119	Vrubel A, Bellon ORP and Silva L. A 3D Reconstruction pipeline for digital preservation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR 2009)*, 2009, pp. 2687–2694.
S120	Waugh A, Wilkinson R, Hills B and Dell'Oro J. Preserving digital information forever. In: *Proceedings of the fifth ACM conference on digital libraries (DL '00)*. New York: ACM, 2000, pp. 175–184.
S121	Woods K, Lee CA and Garfinkel S. Extending digital repository architectures to support disk image preservation and access. In: *Proceedings of the 11th annual international ACM/IEEE joint conference on digital libraries (JCDL '11)*. New York: ACM, 2011, pp. 57–66.
S122	Yu F. Case study of digital preservation for e-heritage: Digital preservation project for the collection of Penhas family. In: *Proceedings of the 4th international conference on management of e-commerce and e-government (ICMeCG)*, 2010, pp. 71–75.