Preservation research and sustainable digital libraries

Seamus Ross[1], Margaret Hedstrom[2]

[1] HATII, University of Glasgow, UK
[2] University of Michigan, USA

Abstract. The National Science Foundation and DELOS, the European Commission sponsored Network for Digital Libraries, supported a working group to define a research agenda for digital archiving and preservation (DAP-WG) within the context of digital libraries. The report of this group, *Invest to Save*, has laid out a range of research challenges that need to be addressed if we are to make progress in the development of sustainable digital libraries. DAP-WG considered archiving and preservation needs and the research that had been conducted to address these. It concluded that research in this domain could benefit from being expanded and refocused—new research communities must be engaged, the approaches to conducting the research must be made more rigorous, and a significant shift in what was being researched needed to be taken. The Group identified twenty-two key research activities worthy of investigation.

Keywords: Preservation – Digital archiving – Digital library – Needs analysis

Introduction

Commercial enterprises, government agencies, and private citizens need digital archiving mechanisms in order to retain access to their records in the face of constantly changing information and communication technologies (ICTs). Digital content is being created at exponentially accelerating rates and is finding a role in more and more spheres of human activity. Indeed, the need for digital preservation touches all our lives, whether we work in business or public sector institutions, engage in e-commerce, participate in e-government, or use a digital camera. In all these instances we use, trust and create e-content, and expect that this content will remain accessible to allow us to validate claims, trace what we

have done, or pass a record to future generations [6]. Digital entities are fragile, prone to corruption, susceptible to misidentification, frequently poorly described (if at all), difficult to track, and hard to maintain because of media, hardware and software obsolescence. Yet, they are at the heart of the digital environment that is fundamentally reshaping how society produces, disseminates, uses, and repurposes information and knowledge. This transformation requires effective digital preservation and curation solutions as part of the infrastructure for a knowledge-driven economy. If cultural institutions are to extend their missions of preserving society's cultural heritage and intellectual capital into the digital age they need more reliable and more affordable digital archiving methods, systems, and technologies. The availability of these depends upon innovative research followed by imaginative development and deployment of products and tools. In an effort to address the challenges surrounding long-term digital preservation and curation the United States National Science Foundation's (NSF) Digital Library Initiative and the European Union under the Fifth Framework Programme Network of Excellence for Digital Libraries (DELOS) agreed to establish a joint working group [A].

The Digital Archiving and Preservation Working Group (DAP-WG) examined the available research, considered the issues that must be addressed if our society is to have the ability to preserve digital information resources for future reuse, and set out priorities for research to ensure that the challenges were addressed and opportunities opened up. The members of the DAP-WG concluded that investing in well-founded and focused research now will save money in the long term, for government, industry, and the citizen. This investment will create new commercial opportunities in the form of both services and products. Our report *Invest to Save* [6], noted that digital archiving challenges threatened the successful

expansion of the information society, its accountability, and its memory. Opportunities for reuse of existing data are often overlooked, while potential commercial developments it could generate are missed. For example, e-government transactions produce digital records that must be organised, managed, and retained for purposes of audit and public accountability. E-commerce depends upon access to records, sometimes, as in the case of pharmaceuticals and aerospace manufacturers, for decades, for purposes of auditing and compliance with laws and regulations. If they are to serve the needs of their current and future users digital libraries must develop digital repositories that will enable them to ingest, make accessible, conserve, and preserve digital entities.

Current approaches to digital preservation and curation are limited. They are labour intensive. They depend upon awareness of many types of risks, including technology obsolescence, deliberate alteration, interruptions in management and funding for curation, and incompetent handling. Their scalability has not been demonstrated. They have neither been tested against a diversity of types digital entities nor repeatedly tested against the same digital entities. The fundamental issues of digital entity interpretability and trustworthiness have not been given generic solutions. Automating digital preservation and curation processes has not been successfully achieved. So, as imaginative and groundbreaking as research in the area has been to date it has not laid the foundation to meet these challenges, although it has provided an indication of what shape these foundations might have. We took as a given that technology will continue to evolve, new kinds of digital entities will be developed, user expectations both of the kinds of material with which they can work and of the ways they can work with them will rise, and the quantities and diversity of materials that will need to be preserved will keep on expanding. New research projects in digital preservation and archiving need to: (a) concentrate on a specific aspect of digital preservation; (b) focus on tangible deliverables; (c) place more emphasis on engineering and computer science that is informed by archival issues and long-term digital preservation requirements, and (d) recognise that digital entities are the raw materials for the industries of the 21st century – intellectual capital. The problem is not just technological. If it were, the problems associated with long-term use would be – and would have been – more readily solvable. Organisational, legal, cultural, social, and financial obstacles also need to be addressed.

Method of working

Early in 2002 our working group was established by the NSF and DELOS. We designed the working group to include a balance of experts in preservation studies, library and archival studies, and information and computing science from Europe and North America. Members included Kevin Ashley (University of London Computing Centre), Birte Christensen-Dalsgaard (Statsbiblioteket, Denmark), Wendy Duff (University of Toronto), Henry Gladney (HMG Consulting), Claude Huc (French Space Agency), Anne R. Kenney (Cornell University), Reagan Moore (San Diego Supercomputer Center), and Erich Neuhold (Fraunhofer, Darmstadt Germany). The working group was co-chaired by Margaret Hedstrom (University of Michigan) and Seamus Ross (HATII, University of Glasgow). DAP-WG met three times between April and November 2002: twice in Washington DC and once in Paris. Members of the working group were tasked with specific research between meetings. At its first meeting the group developed a matrix of current digital preservation projects and research initiatives, discussed our initial impression of the gaps in current research, and developed a work schedule. Three month later, in Paris, the DAP-WG focused on developing a conceptual framework, reviewing the work done to date, drafting an outline for the final report and sketching its content, as well as finalising a work plan for writing the report. We further developed these at a final meeting in Washington in November 2002. This was followed by online exchange of material and discussion to construct the final version of the report. The final report was endorsed in full by nine Working Group members, one member dissented on several points.

Why is preservation necessary

Society benefits from the availability of preservation mechanisms when unique and vulnerable digital entities need to be secured, when long-term accessibility to and for researchers and the public needs to be guaranteed, when accountability of individuals, institutions, and governments can be fostered, when economic and societal advantages can be created through the reuse of information, and when legal constraints make preservation essential. We recognised six primary reasons why technologies, practices, and methods to enable the preservation of digital entities were essential. First we accepted the notion that the *protection and conservation of cultural memory is a societal good*. Archives, libraries, museums, and other memory institutions acknowledge preservation as one of their core roles. They foresee the need to prepare for the deluge of digital materials. They will only be able to respond if suitable preservation systems and technologies are available. Long term access to digital materials enables cross-disciplinary collaboration. International scientific collaborations have *benefited from the availability of data repositories* and more and more investigations depend on re-using data sets collected from earlier experiments or observations and analysing the data in new ways and in conjunction with other datasets [13, 14]. Third, *accountability* of regulated industries, financial institutions, hospitals and clinics, and public bodies depends upon the guaranteeing the authenticity and integrity of

digital materials over time. The main economic benefits of long-term preservation *depend on the re-use* of stored digital information that is impossible to reproduce or far too costly to regenerate. For example, research institutes such as space agencies, nuclear-physical or geo-physical institutions, or social field research institutes preserve scientific and sociological empirical data for secondary evaluation, re-purposing as research problems change, and enabling longitudinal studies of change over time. Fifth, *effective and affordable digital preservation strategies* and systems will underpin the move from an industrial to a knowledge economy. Finally, sustainable digital libraries depend upon the *availability of preservation tools and services*. Increasingly nations are recognising the need to establish legislation to enable their national institutions to ensure their digital documentary heritage is secured for posterity. The institutions charged with delivering on this legislative mandate are finding it difficult because the core technologies and skills are not available.

Constraints

At the outset, members of DAP-WG agreed that developing a research agenda, setting priorities, and indicating the most effective types of projects and research methods was best done if we were first to characterise the space in which any research the agenda proposed would be promoted, conducted, exploited, and eventually deployed. Digital preservation *depends upon a long-term perspective*, where long-term means a period of time long enough to be concerned with the impact of changing technology on the materials in a repository [3, 11]. While only some digital entities will have long-term value we do not have the luxury of time in appraisal of digital materials for preservation that we did with physical objects. One crucial way to address this is to stage preservation action at the point when the digital materials are first created, but intervention will be necessary at subsequent stages. Preservation of the digital entities has little value if they cannot be interpreted and trusted. This makes establishing mechanisms and approaches to the maintenance of the *authenticity and integrity of digital entities* essential. The available methods from audit trails to trusted repositories have not been sufficiently elaborated, developed, tested, or evaluated. Research in preservation needs to come to grips with the problem of *scalability*. It ranges from the variability in the complexity of digital resources to the difficulties associated with managing the rate at which material is ingested into repositories, to the fact that the scale of the activity increases as the numbers of digital objects accumulate. As digital entities reside in an environment susceptible to constant technological change, digital preservation is not a one off event, but must be *continuous and dynamic*. Preservation can not begin at the point at which a digital entity finds its way to a memory institution because all aspects of the life-cycle of a digital entity impacts on what should be preserved, how it should be preserved, and what information exists about its context of creation and use. So, research into preservation must be conducted within the *context of a lifecycle*. The requirement that preservation be addressed at many different stages in the lifecycle of a digital entity and the variety of skills required to make this happen also means that memory institutions, governments, and commercial organisations need to recognise their *shared roles and responsibilities*, including issues of costs and who will pay. The diversity in technologies, organisational contexts, types of digital entities, cultures, legislative frameworks, and purposes for which digital entities are created (from electronic records to digital art) means that a *variety of approaches to preservation* need to be established. As new technologies, methods, and even concepts for creating and exploiting digital information emerge, curatorial processes will require regular rethinking and revision. Similarly, this diversity of contexts, obstacles, and objectives will require that research teams *have a multi-disciplinary character* (e.g. archival science, computer science, information studies, organisational behaviour and human computer interaction). Theoretical and methodological research into digital preservation *can not be separated from practice*; the results of research have value in this context only when they can be deployed, whether this happens as new ways of working or commercial products (or both). The sheer quantities of material that need to be preserved and the pervasiveness of the need (whether for economic, audit, or compliance reasons) for methods and technologies to enable its preservation makes *preservation a high-priority area for research.*

Previous work

DAP-WG began by examining existing and ongoing research in the context of the rapid developments in technology and application environments. There are plethora a recent reviews of thinking in the area of digital preservation [16] and projects that promote research and awareness in the areas of digital preservation. These include CEDARS, CAMiLEON, ERPANET, InterPARES, and PADI [B]. This research has been funded by a range of programmes from the US National Science Foundation (NSF), the UK Joint Information Systems Committee (JISC), the Deutsche Forschungsgemeinschaft (DFG), the European Union (EU), the Social Science and Humanities Research Council (SSHRC) of Canada, the U.S. National Historical Publications and Records Commission, and the Library of Congress's National Digital Information Infrastructure and Preservation Program (NDI-IPP) [5, 9] [C]. Two decades of research has done much to raise awareness of the digital preservation challenge and encouraged some organisations to adopt policies and procedures to improve the longevity of their digital resources. It has contributed to improvements in practice including

enhancing metadata creation, increasing the use of standards, and helped to establish widespread recognition that intervention needs to come early in the life of digital entities (if not before they are even created). Another significant stream of research has been carried out within cultural institutions, sometimes in collaboration with industry partners, to develop solutions to operational problems in these institutions. National, regional, and University archives and libraries in Australia, Canada, Belgium, Denmark France, Germany, the Netherlands, New Zealand, Sweden, Switzerland, the U.K., the U.S., and elsewhere have investigated the development of institutional repositories, preservation, and strategies for web archiving [D].

Funded research programmes, within the academic environment and in cultural institutions have resulted in publications, software, demonstrator systems, archive and digital library services [1, 4, 10, 15], and guidelines, training materials and handbooks. These are designed to assist practioners address immediate preservation challenges [2]. Commercial (e.g. pharmaceuticals, petrochemical), scientific (e.g. genome sequencing, seismic surveys), and public sector (e.g. national archives and libraries) organisations have amassed experience in overcoming obstacles to preservation, but the results and accumulated knowledge have not always been disseminated as widely or in the same ways as formally funded research. These approaches have resulted in products that are either guidelines and high-level abstract models (e.g. OAIS), or very specific solutions tied to a particular format of material or institutional context. At institutional level research is usually only sufficient to meet current business needs and little consideration is given to its wider relevance or applicability. Moreover, little of this work enters the public domain. The lack of broad recognition by memory institutions, commercial enterprises, and funding agencies of the significance of the problem also has meant that research as been 'sporadic', fragmented, small scale, uncoordinated, and repetitious.

Preservation of the bitstream has been the focus of much research effort, and this has resulted in a substantial body of knowledge. From the vantage of digital libraries the real challenges lie in further development of work related to models for repositories or preservation processes, schema for preservation metadata, or general guidelines for content creation. Format-centered work has given us good ideas about how to handle the preservation of static documents or structured numeric data. But new formats emerge regularly and, more importantly, there had been little research on complex entities despite the fact that almost all digital entities are now complex amalgams of a number of information types. Consider a Web page that contains text, image, metadata and a small program, or a spreadsheet that interacts with a database to produce real-time reports and also contains embedded spoken and written annotations. The challenge here is not simply to preserve each component (speech, text, image, database, program), but to preserve the whole and the interrelationships among the parts in ways that ensure accessibility in the future when models and methods of human computer interaction will be very different from those we experience today.

The digital archiving and preservation research agenda

Building on the analysis of the work to date and an assessment of the rapidly evolving landscape of information technologies the DAP-WG concluded that the research needs to be classified into three areas. Preservation Strategies – Emerging Research Domains, identifies new problems that result from constantly evolving technology. Re-engineering Preservation Processes, sketches ways in which archival processes might be re-engineered to reduce significantly the costs of long-term data management and preservation. Finally, Preservation Systems and Technology, discusses tools and technologies that are needed to support these process transformations. In the second and third sections we have identified research that is urgent and will deliver benefits quickly. This is in contrast to the ideas set out in section one which, although promising, are likely to require several rounds of substantial funding before they deliver exploitable commercial products. Here we only highlight key issues. We have not attempted to provide a detailed report of the work of DAP-WG, but to give an overview.

Preservation strategies – Emerging research domains

Repositories such as data archives provide fundamental resources to future researchers. In the area of repositories we identified four areas for additional research and development: elaboration of existing repository models, software repositories, format repositories, and repositories of peripheral devices. For example, format repositories will provide crucial information about the structural and semantic properties of digital objects. File format registries enable users (whether human or machine), to identify the format of unknown files, to verify whether a file is the format that it purports to be, to assess the viability and implications of transforming from one file format to another, to provide an information resource to support the investigations of file format risk, and to store information about how to render an object from a particular format.

The preservation community has accumulated substantial knowledge about media, but we lack *archival media* which is sustainable without intervention for 100s of years. Engineering and materials science research could do much to deliver this. Bitstream preservation has been shown to be viable and very inexpensive (even if it could be enhanced by archival media), but little work has been done on developing *salvage and rescue* techniques to enable raw data streams to be analysed and the original meaningful (e.g. logical) units they represent re-

constructed. Storage abstractions provide the ability to characterise digital entities independently of underlying software and hardware infrastructure. This implies the ability to create infrastructure-independent representations of digital entities. No formal ways to express the *functionality and behaviour of digital entities* exist. These could contribute to benchmarks, the measurement of consistency of performance across migrations or emulations and enable us to reconstruct systems and automatically verify the validity of the reconstruction.

Context-aware digital entities that know what they are (e.g. their file type), can observe the state of other objects, know where they are and where their metadata are, and can communicate with their originator or manager could underpin automated archiving strategies. Research into *accelerated aging tests* is needed to develop formal methods for predicting the viable life of specific media, formats, software applications, and standards. As archives preserve *intellectual capital* more robust ways to describe digital entities and relate them to their larger organisational, technical, and domain context will be required to open new opportunities for content owners and users.

Re-engineering preservation processes

Most research on digital preservation has attempted to map traditional preservation practices for physical materials onto digital content. Digital preservation strategies necessitate a complete redesign and re-engineering of digital preservation processes. DAP-WG concluded that automating processes must be a key feature of this re-engineering if preservation activities are to be made efficient and cost-effective. Longevity of digital entities can be enhanced if preservation functionality is built into systems used to create and manage them. This can only happen if research delivers *preservation modeling methodologies* capable of formally representing preservation functions and processes. Moreover, most modeling work has attempted to replicate traditional processes. Formal modeling of preservation processes is the necessary first step in re-engineering them and developing tools for automatic management of digital entities. Current digital longevity processes require extensive human intervention, which does not scale to the size or complexity of the digital content that needs to be preserved. Research to *automate processes* from ingest, to validation, metadata creation, data management, and migration is needed to make preservation efficient and affordable.

Preservation processes often depend on transforming the original bitstream. Unless handled correctly, such transformations call into question the *authenticity, quality and trustworthiness* of the preserved entity. Most digital preservation research to date has examined either large sets of homogeneous data or small collections of heterogeneous material. This raises a series of issues concerning the *scalability* of current models and methods,

including the maximum archive size, the ingest rate, and the rate at which digital materials can be normalised or migrated.

A corollary of trustworthiness and scalability are mechanisms to determine *collection completeness, and to detect and correct anomalies*. Current processes are manual. Can automated processes be developed and integrated into the accessioning process that would provide better detection of problems with collections or when they are incomplete? New methods for distributed storage offer potential benefits to long-term preservation, such as increased redundancy and lower storage costs. Grid technologies, for example, provide support for methods that fragment and distribute digital entities. Numerous research questions arise as to how we can use these storage models in the preservation arena and reduce the risks that their use poses.

Preservation systems and technology

Systems, tools, and technologies to support digital preservation depend in part, on producing formal models of preservation processes. Their availability is an area with great commercial potential. Different formats require different kinds of strategic approaches to ensure that they can be accessed in the future. Some formats may have properties that enable or put preservation at risk. We need predictive measures to enable developers to assess the preservation impact of attributes of formats in advance of their completed development. *Managing Complex and Dynamic Digital Entities* has been little studied. For instance, how can complex and dynamic entities be authenticated and their integrity verified? How can dynamic entities be accessioned and managed in an archive? To what aspects of a dynamic document should metadata be attached and what metadata would be required? Creating metadata is labour intensive and expensive. Preservation of digital entities requires mechanisms for *automatic or semi-automated creation* and authoring of the metadata. Based on existing technologies for extraction of low-level features and structural and administrative information, the challenge is the development of methods and approaches for creation of metadata supporting the understandability of digital entities. The meaning of metadata itself changes over time, what we might describe as 'metadata drift'. Among the research issues in this area are metadata schema and ontology evolution mapping to ensure that, over time, metadata and underlying ontologies do not lose their meaning. Tools are needed to track the provenance of metadata schema.

It will not always be possible or economical to preserve all of the features and functionality of original digital entities, but we lack metrics for measuring what is *acceptable loss*. Even where we can retain digital entities new ways of presenting them must be investigated. *Repurposing* archives and using their contents in unforeseen scenarios requires more than the usual access func-

tions supporting search and presentation of found items. Parallel research on discovery, search, retrieval and presentation can be applied to this problem. Digital entities worthy of preservation come in all languages, but there has been little research on issues of *multilingualism*.

Towards a sustainable infrastructure for digital archiving

Nearly a decade ago, the Task Force on Archiving Digital Information identified the need for a *deep infrastructure* for digital archiving and long-term preservation [17]. Sufficiently robust, effective, and affordable digital archiving strategies ultimately depend on an infrastructure of common standards, methods, and tools. The recommendations for research outlined by our working group will facilitate discovery and development of key infrastructure components. For example, a relatively small number of repositories are needed to maintain software, formats, or peripheral devices for use by a much larger user community. Digital archiving standards could form the basis for commercial products and services for which the potential market is substantial. Products that would enable individual content creators to produce digital objects that are archive-ready with little or no additional effort could have an enormous impact on driving down preservation costs.

The DAP-WG recognised that an infrastructure for digital archiving requires much more than research and development of tools and technologies. Developing organisational capacity capable of sustaining digital entities depends upon further research into how these technologies can be effectively deployed, how content creators can be encouraged to use them, and on investigations into their scalability so that they are applicable for use by individuals and by organisations of all sizes. The processes by which many organisations archive material are heterogeneous and their structures are usually incompatible. Ways need to be found to overcome these inconsistencies if institutions are to be able to exchange and re-use their assets and if users are to be able to search across diverse and distributed repositories.

A number of legal and regulatory frameworks from data protection to privacy legislation to intellectual property rights (IPR) issues could have an impact on what digital entities can be preserved, how they can be preserved, and what kinds of access can be provided to them. Research at international, EU, and national levels is needed to clarify this and to track the implications of proposed legislation on long-term archiving. Changes may be needed in legislation to create a legal environment(s) conducive to preservation and reuse of digital information.

Expertise in the preservation of digital entities is limited and currently there are few educational programmes that provide opportunities for the next generation of information specialists and computing scientists to develop knowledge in this domain. Further development of digital archiving capabilities will require trained professionals with knowledge and skills in the technological, curatorial, and legal aspects of preservation, along with specific domain knowledge. The ongoing continuing professional development (CPD) activity, such as that run by ERPANET in Europe, needs to be extended, deepened, and made relevant to broader communities. Existing CPD programmes focus mainly on library, archives (including records managers), and museum professionals. This training needs to be supported by research that improves our understanding of workflow and how the special requirements for handling born digital (and digitised) materials can be meet within the organisational framework.

Conclusions

Research into digital preservation and curation is still in its early days. Whereas many other areas of research have the development of new knowledge as a primary objective, digital preservation studies must also deliver practical results. In designing research projects emphasis must be put on the production of exploitable outcomes through the development of methods and practical tools. While in 2004 revenue from digital preservation and curation products and services is small by comparison to data storage products and services, the potential of the market is substantial. The growth of data-driven research based on reuse of individual and dispersed data sets, of which bioinformatics is a proven example, and the need for sustainable digital libraries will be instrumental in producing a market for the results of digital archiving and preservation research over the coming five to ten years. The research proposed by the DAP-WG will produce prototypes and pilots. It will not generate commercial products. The transition to, what we termed, 'industrial strength' products will require substantial additional investment. Planning for commercial development of the outcomes of research undertaken in this area should begin at the very outset of every research project. Applicants for funding should indicate how their results would be represented in public domain solutions and commercial products. It would be helpful if the ways research were evaluated were to shift so that, in addition to measuring whether or not a project achieved its initial objectives, its impact over a longer period were to be assessed. Impact will be seen in new hardware and software products, and services ranging from repositories to value-added services. The markets range from public and private institutions to citizens.

As with any research programme, The NSF/DELOS DAP-WP research agenda is not risk free. Indeed some of the research areas it recommends may take several research project failures before they produce exploitable results. That said, in assessing the risks of the research

proposed here it is essential not to err on the side of caution; we must ask whether or not we are taking enough risk. In the case of digital preservation and curation the long term sustainability of digital entities poses sufficient risk to lead us to conclude that we must entertain high risk research alongside lower risk developments.

Adequate preservation technologies, practices, and methods are in the public interest. Data, information, and knowledge are vital resources. They fortify democracy. They enable corporate and public sector accountability. They promote economic growth and underpin the confidence of citizens in e-government. The scale of investment that is needed can only come from public sources, although industry should also be making substantial investments in archiving technologies and services. If even the existing level of research investment, as meagre as it is, is to be maximised, it is essential that preservation and curation research be informed and co-ordinated by a common research programme. The Digital Archiving and Preservation Working Group of the National Science Foundation and the European Union's Network of Excellence DELOS has developed this agenda. National funding agencies in North America, Europe, Asia, Australia and elsewhere supporting research that directly or indirectly addresses the digital preservation domain need to be encouraged to do so in the context of this agenda.

Notes

[A] National Science Foundation (NSF) Award #0207482 and the European Commission under IST-1999-12262 (DELOS). The opinions expressed in this report are those of the authors and do not necessarily reflect the views of the National Science Foundation or the European Commission.

[B] `http://www.leeds.ac.uk/cedars`;
`http://129.11.152.25/CAMiLEON//dh/ep5.html`;
`http://www.erpanet.org`;
`http://www.interpares.org/reports.htm`;
`http://www.nla.gov.au/padi`

[C] `http://www.nsf.gov/`;
`http://www.jisc.ac.uk`; `http://www.dfg.de/`;
`http://www.cordis.lu/ist/ka3/digicult/research.htm`;
`http://www.sshrc.ca`;
`http://www.archives.gov/grants/`;
`http://www.digitalpreservation.gov/`

[D] The literature is extensive and here we note activity in a dozen countries. In each case we could have cited numerous other projects or publications. Links and references to further examples can be found at the ERPANET (`http://www.erpanet.org`), and at the PADI – Preserving Access to Digital Information (`http://www.nla.gov.au/padi`) websites. *Australia*: Public Records Office Victoria. Victoria Electronic Records Strategy (VERS), `http://www.prov.vic.gov.au/vers/published/publcns.htm`, Margaret Phillips, PANDORA, Australia's Web Archive, and the Digital Archiving System that Supports It, DigiCULT.info (`http://www.digicult.info`), Issue 6, December 2003, Warwick Cathro, Colin Webb, and Julie Whiting, (2001), *Archiving the web – The PANDORA Archive at the National Library of Australia* presented at Preserving the present for the future – Strategies for the Internet, Copenhagen (18–19 June 2001), `http://www.nla.gov.au/nla/staffpaper/2001/cathro3.html`; *Belgium*: City Archives of Antwerp. DAVID: Digitale Archivering in Vlaamse Instellingen en Diensten (Digital Archiving in Flemish Institutions and Administrations), `http://www.antwerpen.be/david/`; *Canada*: `http://www.collectionscanada.ca/cidl/preserv-conserv/preserv-e.htm`; *Denmark*: Netarchive.dk, `http://www.netarchive.dk/index-en.htm`; *France*: Julien Masanès, (2002), 'Towards Continuous Web Archiving: First Results and an Agenda for the Future', *D-Lib Magazine*, 9.12 (December), `http://www.dlib.org/dlib/december02/masanes/12masanes.html` and S. Abiteboul, G. Cobena, J. Masanès, and G. Sedrati, (2002), 'A First Experience in Archiving the French Web', M. Agosti and C. Thanos (eds.), *ECDL 2002*, LNCS 2458, 1–15. `ftp://ftp.inria.fr/INRIA/Projects/verso/gemo/GemoReport-229.pdf`; *Germany*: DOMEA: Dokumentenmanagement und Elektronische Archivierung `http://www.uni-koblenz.de/fvi/projekte/domea.html`; *Netherlands*: Digitale Duurzaamheid (Testbed Digitale Bewaring – Digital Preservation Testbed) `http://www.digitaleduurzaamheid.nl`, The NEDLIB Guidelines – Setting up a Deposit System for Electronic Publications, (NEDLIB Report Series, 5, Koninklijke Bibliotheek, 2000), `http://www.kb.nl/coop/nedlib/`; *New Zealand*: S Ross 2003, *Digital Library Development Review*, National Library of New Zealand, (Wellington), ISBN 0-477-02797-0, `http://www.natlib.govt.nz/files/ross_report.pdf`; *Sweden*: DiVA – Digitala vetenskapliga arkivet (Digital Scientific Archive), `http://publications.uu.se/epcentre/` (There are a large number of publications from this project, see for example: Eva Müller, Uwe Klosa, Peter Hansson, Stefan Andersson, Erik Siira Archiving Workflow Between a Local Repository and the National Archive (2003-08-18: ECDL 2003), `http://publications.uu.se/epcentre/conferences/ecdl2003/archiving_ECDL_2003.pdf`, Kulturarw3 at the National Library of Sweden `http://kulturarw3.kb.se/html/kulturarw3.eng.html`; *Switzerland*: ARELDA Project of the Swiss Federal Archives (`http://www.bundesarchiv.ch/`); *United Kingdom*: SHERPA project, `http://www.sherpa.ac.uk`; The National Archives, `http://www.nationalarchives.gov.uk/preservation/` and its work on electronic records `http://www.nationalarchives.gov.uk/electronicrecords/advice/guidelines.htm`. See also The National Digital Archive of Datasets (NDAD), `http://ndad.ulcc.ac.uk/`; *United States*: The National

Archives (`http://www.archives.gov`) and the San Diego Super Computer Centre (e.g. `http://www.sdsc.edu/NHPRC`), W. Y. Arms, (2001), *Web Preservation Project: final report*. (Washington, D.C.: Library of Congress), `http://www.loc.gov/minerva/webpresf.pdf`

References

1. Abrams SL, Seaman D (2003) Towards a Global Format Registry. In: IFLA2003, Berlin. `http://www.ifla.org/IV/ifla69/papers/128e-Abrams_Seaman.pdf`
2. Beagrie N, Jones M (2001) Preservation Management of Digital Materials: A Handbook. London. `http://www.dpconline.org/graphics/handbook/index.html`
3. Consultative Committee for Space Data Systems [CCSDS, Panel 2] (January 2002) CCSDS 650.0-B-1: Reference Model for an Open Archival Information System (OAIS). Blue Book. Issue 1, 1-1. `http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf`
4. CEDARS Project (2000) Metadata For Digital Preservation The Cedars Project Outline Specification Draft For Public Consultation. `http://www.leeds.ac.uk/cedars/metadata.html`
5. Friedlander A (2002) The National Digital Information Infrastructure Preservation Program. D-Lib Magazine 8.4, April 2002. `http://www.dlib.org/dlib/april02/friedlander/04friedlander.html`
6. Hedstrom M, Ross S, Ashley K, Christensen-Dalsgaard B, Duff W, Gladney H, Huc C, Kenney AR, Moore R, Neuhold E (2003) Invest to Save: Report and Recommendation of the NSF-DELOS Working Group on Digital Archiving and Preservation. Pisa & Washington DC. `http://delos-noe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/digitalarchiving/Digitalarchiving.pdf`
7. Hedstrom M (2003) It's About Time: Research Challenges in Digital Archiving and Long-term Preservation. Final Report on a Workshop on Research Challenges in Digital Archiving, 12–13 April 2002. Washington, DC, Library of Congress.
8. InterPARES Preservation Task Force Final Report (2003) `http://www.interpares.org/book/interpares_book_f_part3.pdf`
9. Library of Congress (2002) Plan for the National Digital Information Infrastructure and Preservation Program, Washington D.C. `http://www.digitalpreservation.gov/index.php?nav=3&subnav=1`
10. Payette S, Staples T (2002) The Mellon Fedora Project: Digital Library Architecture Meets XML and Web Services. In: Agosti M, Thanos C (eds) ECDL 2002, LNCS, vol 2458, pp 406–421
11. Lavoie B, Dale R (2002) Preservation Metadata and the OAIS Information Model. A Metadata Framework to Support the Preservation of Digital Objects. A Report by the OCLC/RLG Working Group on Preservation Metadata, June. `http://www.oclc.org/research/pmwg/pm_framework.pdf`
12. Research Issues in Electronic Records (1991) St. Paul: Minnesota Historical Society for the National Historical Publications and Records Commission
13. Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure (2003) Arlington, Virginia, USA: National Science Foundation, January. `http://www.communitytechnology.org/nsf_ci_report/`
14. Schade D (2002) The Virtual Observatory: The Future of Data and Information Management in Astrophysics. In: 2002 CODATA Conference, Montreal
15. Steenbakkers J (2000) In: The NEDLIB Guidelines – Setting up a Deposit System for Electronic Publications. NEDLIB Report Series, 5, Koninklijke Bibliotheek
16. Tibbo HR (2003) On the Nature and Importance of Archiving in the Digital Age. Advances in Computers 17:1–67
17. Waters D, Garrett J (2003) Preserving Digital Information: Report of the Task Force on Archiving of Digital Information. The Commission on Preservation and Access and The Research Libraries Group. `http://www.rlg.org/ArchTF/`