Aligning National Approaches to Digital Preservation

Nancy Y. McGovern, Volume Editor
Katherine Skinner, Series Editor

EDUCOPIA INSTITUTE PUBLICATIONS
Atlanta, Georgia

Aligning National Approaches to Digital Preservation

Nancy Y. McGovern, Volume Editor
Katherine Skinner, Series Editor

EDUCOPIA INSTITUTE PUBLICATIONS
Atlanta, Georgia

EDUCOPIA INSTITUTE PUBLICATIONS
Atlanta, Georgia
www.educopia.org/publications

References to websites (URLs) were accurate at the time of writing. Neither the authors nor the editors are responsible for URLs that may have expired or changed since the manuscript was prepared.

Publication and Cataloging Information:

ISBN: 978-0-9826653-1-2

Volume Editor: Nancy Y. McGovern

Series Editor: Katherine Skinner

Publisher: Educopia Institute Publications
Atlanta, GA 30309
USA

STANDARDS ALIGNMENT

Raivo Ruusalepp (Tallinn University)
Christopher A. Lee (University of North Carolina at Chapel Hill)
Bram van der Werf (Open Planets Foundation)
Matthew Woollard (UK Data Archive, Economic and Social Data Service)

Abstract

A standard is a specification of precise criteria designed to be used consistently and appropriately. This essay discusses standards that are relevant to digital preservation. We start by presenting an overview of what has been achieved in terms of standards and standards alignment and follow this with a discussion of what we perceive to be the main challenges in aligning standards for preservation requirements. We then propose an agenda for action for the coming five years.

Introduction

Standards cover a variety of topics and issues; they may be *normative*—setting requirements for quality and actions, or *informative*—describing and guiding the use of methods. In all cases they represent agreements that are generally, but not always, considered to be best practice.

A standard of any form or type represents a statement by its authors, who believe that their work will be understood, accepted, and implemented by the market. This belief is tempered by the understanding that the market will work in its own best interests, even if those best interests do not coincide with the standard. A standard is also one of the agents used by the standardization process to bring about market change (Cargill, 2011).

Standards "embody the outcomes of negotiations that are simultaneously technical, social, and political in character" (Edwards, 2004, p.827). Standards usually are designed by institutions and individuals who have aligned their interests and have been able to reach a consensus on these negotiating factors. Standards are often categorized along one dimension as being either *de jure* (what the law states) or *de facto* (what actually happens in practice), and along another dimension as being either

open or closed (proprietary). The concept of openness itself addresses several different factors; for example, a proprietary standard may be public, but not open (Tiemann, 2005). Some technical texts that are widely used as standards are publically available specifications (PAS) or requests for comments (RFC), while other standards are established by institutions and intended only for their own internal use, but can have a wider applicability.

An important distinction is between anticipatory and *ex poste* standards (Byrne and Golderb, 2002; Schumny, 2002). The former are introduced before products are developed, while the latter are codifications of characteristics reflected in existing products. Anticipatory standards development is a "future oriented and self-creating process of defining standards: writing for the future now" (Bonino and Spring, 1999, p.101).

Other sources of guidance are not official standards, but since they are a form of agreement for controlling activities that may lead to repeatable ways of carrying out activities, they should be mentioned briefly here. Legislation is not usually seen as a standard, but it can impinge on some institutions' digital curation activities. National differences in copyright and data protection legislation can produce different processes being applied across boundaries.

A major trend since the mid-1980s, has been that parts of the information and communication technology (ICT) industry, sometimes combined with university researchers, have moved away from well-established formal standards development organizations (SDO) such as the International Organization for Standardization (ISO) and have instead formed more *ad hoc* consortia in order to establish specific standards or classes of standards (Weiss and Cargill, 1992; Updegrove, 1995; Cargill, 1999). While industry consortia can often act much more quickly than SDOs in the development of standards, consortia are likely to have much less incentive than publicly funded SDOs to develop standards that require significant time and energy, have little immediate financial payoff (Spring and Weiss, 1994), or are designed to have very wide applicability.

Carl Cargill (1997) describes a chain of standards at increasing levels of specificity. The highest level of standard is a *reference model*, which characterizes a problem space, providing fundamental concepts and terminology. At the next level down, an *industry consensus* standard describes a subset of the functions or

capabilities identified in the reference model. At the third level, a *functional profile* describes a set of functions from the industry standard for a specific, but large, class of users. Next down, the fourth category of standard is the *systems profile*, which describes the system requirements of a smaller group of users than that addressed by the functional profile. Finally, a specific organization that has its own needs and requirements can often be addressed by a *document* or set of documents that specifies the implementation in its particular organizational and technical context.

Standardization can help to set the direction of product development within an industry, but it can also help to shape and reinforce particular approaches to professional work. For example, the development of formal management hierarchies and the systematic management movement in the 19th century were based on an intersection between standardized metrics, tools and resources; and the differentiated professional status of managers and engineers who had the expertise to control and coordinate the use of such metrics, tools and resources (Cargill, 1989; Yates, 1989; Zuboff, 1988; Chandler, 1980).

Digital preservation relies on interoperability between computer systems and is thus dependent on standards. Standards are essential for the ability of software and hardware to exchange and use information. They can be seen as tools that can help to make digital collections accessible, sustainable and interoperable. The difficulty usually lies in the selection of the appropriate combination of standards and, if necessary, their customization to suit the specific needs of the organization using them. Navigating at least 200 standards[1] that are related to preservation and digital curation can be a daunting task. Attempts to maintain a community-based standards watch have faced issues of sustainability.[2]

The literature on the preservation and management of digital objects includes many references to the important role of standards (Walch, 1990). Several authors have identified standards and

[1] There is no register of standards relating to digital curation. This number is based on those (current) standards known to the authors.

[2] For example, updates to the DIFFUSE project ended in 2009, and updates to the Preserving Access to Digital Information (PADI) page on standards likewise ended when the National Library of Australia ceased to support it in 2010. See DCC DIFFUSE Standards Frameworks: http://www.dcc.ac.uk/resources/standards/diffuse/ and PADI Standards http://www.nla.gov.au/padi/topics/43.html (both last accessed 03-21-2012).

standardization as important components of professional education of those responsible for managing and preserving digital resources (Gilliland-Swetland, 1993; Hedstrom, 1993; Walch, 1993). In 1992, Charles Dollar recommended that archivists "identify archival functional requirements" and then participate in standards development organizations in order "to ensure that these functional requirements are incorporated into" relevant standards (Dollar, 1992, p.81). Soon after, David Bearman (1994) presented standards as one of the four "tactics" for achieving the functional requirements for evidence in record-keeping.

In addition to developing standards, many sources have also suggested the value of adopting standards in order to facilitate long-term access to digital objects. According to one early report on electronic records in the US federal government, "Machine incompatibility…will undoubtedly be solved both by standardization and by development of universal conversion machines" (Jacobs, 1961, p. 11). Although this prediction seems overly optimistic in retrospect, there is still considerable hope for the role of standards within the digital preservation literature. A federal report in the US entitled "Taking a Byte out of History" (1990), indicated, "Sometimes, files can be readily converted to a format that uses generic software and standard hardware. When this is possible, specific software and hardware are not needed to ensure long-term access" (p. 3). Dollar and Weir (1991) argued that open standards help to address problems of interoperability over time, much as they support interoperability across systems at a given point in time. Stielow (1992) argued, "Electronic preservation has a chance of success only at the place where standards exist and where we can reasonably project some constancy over time" (p. 334). In 1996, the Task Force on Archiving of Digital Information argued for the potential value of incorporating "data standards" into digital preservation strategies. Dollar (1999) presented standards and open systems as vital components of a digital preservation strategy, though he also raised warnings about the danger of adopting standards that do not ultimately win out in the market. The most outspoken critic of reliance on standards is Jeff Rothenberg, Senior Computer Scientist at the RAND Corporation. He has warned that standards, like proprietary formats, will become obsolete over time and has suggested that "standards may play a minor role in a long-term solution by providing a way to keep metadata and annotations readable" (1999, p. 12).

Digital Preservation Standards

The development of standards within a particular domain is often regarded as a sign of maturity of that domain. Preservation in general and digital preservation more specifically are quite established in this regard with families of standards that are interlinked and that stem from common antecedents. As an example, the Task Force on Archiving of Digital Information's report *Preserving Digital Information* (1996) is notable for its influence on standards currently in use.

Preservation Standards from the Analog Era

Description standards traditionally used in libraries, museums, archives, research data centers, or heritage institutions have often been extended to accommodate their use for accessing digital objects. Significance of metadata standards in ensuring interoperability in the digital environment has grown rapidly and a new brand of standards for metadata exchange has emerged (see below) that did not exist in the analog era.

Standards have also emerged to guide the conversion of analog materials into digital objects (e.g., ISO/TR 13028, 2010). A whole series of standards and technical reports on scanning of microfilms and paper documents has been published by the ISO technical committee on document management applications.[3]

Reference Models for Digital Repositories

The first international standard to describe a digital archive system was the Reference Model for an Open Archival Information System (OAIS) (ISO 14721, 2003) that has become the "ur-standard" for many other standards that have emerged to address digital repositories and digital preservation.[4] The OAIS Reference Model defines the processes required for effective long-term preservation and access to information objects and establishes a common language to describe them. It does not specify an

[3] These are currently shown at the links available at: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_tc_browse.htm?commid=53650 (last accessed 03-21-2012). Details of the committee are available at: http://www.iso.org/iso/iso_technical_committee.html?commid=53650 (last accessed 03-21-2012).

[4] For example, PAIMAS (Producer-Archive Interface Methodology Abstract Standard), the PREMIS (Preservation Metadata Implementation Strategies) metadata data dictionary and TRAC (Trustworthy Repositories Audit and Certification) are all influenced by the OAIS Reference Model.

implementation, but instead provides the framework to make a successful implementation possible, through describing the basic functionality and types of information required for a preservation environment. OAIS identifies mandatory responsibilities and interactions of producers, consumers, and managers of both paper and digital records. It provides a standardized method to describe repository functionality by providing detailed models of archival information and archival functions. Although the OAIS grew out of a standards body—the Consultative Committee for Space Data Systems (CCSDS)—that is focused on terrestrial and space data, its development took on a much wider scope, involving and gaining visibility among a much broader set of stakeholders than simply members of the CCSDS (Lee, 2009). The parties involved in the creation of the OAIS attempted to make it applicable to a wide variety of repository types. In this way, OAIS became a *lingua franca* for archival information systems that has since become widely adopted because it enables effective communication among projects on a national and international scale (Klump, 2011).

The OAIS reference model represents a rare case in the history of use of ICT methods—a model that found broad acceptance across a diversity of audiences and professional communities; from around 1999 to the present, it has appeared in numerous presentations on digital preservation. The exploratory, catalytic, and standard-setting model has contributed to the discussions and the exchange of conceptual and practically realizable ideas within the community of preservation specialists (Oßwald, 2010). The OAIS reference model has been revised by CCSDS and at the time of writing is available as a draft revised standard (CCSDS, 2009).

The OAIS reference model is supplemented by another standard—Producer-Archive Interface Methodology Abstract Standard (PAIMAS) (ISO 20652, 2006)—that describes the workflow of negotiating and coordinating the submission and transfer of objects to an archive. Standards for transfer and ingest of electronic records are also beginning to emerge at a national level (e.g. DIN 31645, 2011).

Other digital repository models have been developed and, although they have not been issued as formal standards, their *de facto* impact on the preservation community and other interest groups is significant. The InterPARES project Chain of Preservation Model (InterPARES, 2007) covers all stages in the

life of digital records, from their creation, through their maintenance by their creator, and during their appraisal, disposition, and long-term preservation as authentic memorials of the actions and matters of which they are a part. The model places the function of preserving records into the context of other business functions of an organization.

The Digital Library Reference Model originally created by the EU-funded DELOS project (DELOS, 2007) and developed further by the DL.org project (DL.org, 2011) provides a formal and conceptual framework describing the characteristics of a digital library management system. The model seeks to overcome the heterogeneity of existing digital library systems and provides a conformance checklist.

The UK-based Digital Curation Centre has published its Curation Lifecycle Model (DCC, 2009) that provides a graphical, high-level overview of the stages required for successful curation and preservation of data from initial conceptualization or receipt. The model can be used to plan activities within organizations to ensure that all of the necessary steps in the curation lifecycle are covered. It is important to note that the model is an ideal and focuses on interrelationships between stages of curation work.

Digital preservation standards have also influenced communication among those outside of the digital preservation domain. For example, a report of the National Academy of Sciences' on ensuring the integrity, accessibility, and stewardship of research data (Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age, 2009) makes use of terms including ingest, data producer, and other phrases which may once have been considered exclusive to those working in digital repositories. Standards can help to provide a common language to work across domains.

Digital Repository Audit Methods

One of the first uptakes of the OAIS reference model was for establishing conventions for determining the trustworthiness of repositories. To determine whether an archive or repository is following practices that will ensure long-term digital preservation required a community consensus. As stated some years earlier by the Task Force on Archiving of Digital Information:

> A critical component of the digital archiving infrastructure is the existence of a sufficient number of

> trusted organizations capable of storing, migrating, and providing access to digital collections. […] A process for certification of digital archives is needed to create an overall climate of trust about the prospects of preserving digital information (Task Force, 1996).

The models that emerged, especially OAIS (because it is a reference model and not a process model) lacked the granularity required for an auditable certification process. Individual, emerging standards lacked a framework for what constituted a trustworthy repository, and the community remained incapable of coming to a collective agreement on an exact definition of "trusted archives" as called for by the task force (Dale, Gore, 2010).

In 2003, the Research Libraries Group (RLG) and US National Archives and Records Administration (NARA) established a joint Digital Repository Certification Task Force with membership from the US, UK, France, and the Netherlands, representing multiple domains including archives, libraries, research laboratories, and data centers from government, academic, non-profit, e-science, and professional organizations (Ambacher, 2007). The task force developed an audit checklist for digital repositories that was published in 2007 as the Trustworthy Repositories Audit and Certification (TRAC) checklist (RLG/NARA, 2007). It presents almost 90 organizational, technological and digital object management criteria for digital repositories. Many are based heavily on the principles, terminology and functional characteristics outlined in the OAIS reference model. The Center for Research Libraries received a grant from the Andrew W. Mellon Foundation to investigate the means for audit and certification of digital archives and to complete a series of test audits to inform the investigation. As a result of these tests, two digital repositories were "certified" by CRL on behalf of its membership as trustworthy digital repositories in 2010 and 2011.[5]

[5] The CRL's advice on certification and assessment can be found at: http://crl.edu/archiving-preservation/digital-archives/certification-and-assessment-digital-repositories (last accessed 03-21-2012). This certification process predated the formal issuance of the Trustworthy Repository Audit and Certification (TRAC) ISO standard in 2011 (ISO 16363: *Audit and certification of trustworthy digital repositories*, described below), but was based upon the 2007 TRAC checklist.

In 2004 the German Network of Expertise in Long-term Storage of Digital Resources (nestor) established a working group[6] on the certification of trustworthy archives. Building on the RLG/NARA draft version of TRAC checklist, the nestor group focused on identifying features and values that are relevant for evaluating both existing and planned digital object repositories. The first version of the nestor criteria for auditing digital preservation repositories was released in 2006 (nestor, 2006) with an update in 2008 (nestor, 2008). This checklist covers the technical, organizational, and financial characteristics of a digital repository. It is structured similarly to the TRAC checklist, but additionally provides examples and perspectives that are of particular relevance to the legal and economic contexts and operational situation in Germany. On the conclusion of the nestor project, work on the trustworthiness criteria was transferred to the German national standards body[7] and a new version of the criteria was published as a national standard DIN 31644 (2010).

In February 2007 the DigitalPreservationEurope project (DPE) and the UK Digital Curation Centre (DCC) published their joint work on digital repository assessment methods as the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) (Hofman et al., 2007). This tool presents a methodology for repository self-assessment and characterizes digital curation as a risk management activity; the job of a digital curator is to rationalize the uncertainties and threats that inhibit efforts to maintain digital object authenticity and understandability, transforming them into manageable risks. An online assessment tool was released in 2008 to guide and document the repository assessment.[8]

The Data Archiving and Networked Services (DANS) in the Netherlands published 17 guidelines in 2008 to help data archiving institutions to establish trustworthy digital repositories for research data. An international editorial board modified these guidelines to deal more broadly with the different needs of a wider audience.

---

[6] See information regarding the nestor working group on the certification of trustworthy archives (an English-language version is at the bottom of the page): http://www.langzeitarchivierung.de/Subsites/nestor/DE/Arbeitsgruppen/AGZertifizierung.html (last accessed 05-21-2012).

[7] The nestor "Trusted digital long-term repositories" working group of the DIN standardization committee "Records management and long-term preservation of digital information objects" (NABD 15).

[8] DRAMBORA Interactive can be accessed at: http://www.repositoryaudit.eu/ (last accessed 03-21-2012).

The Data Seal of Approval (DSA) method (DANS, 2009) characterizes the repository audit as a three-stage process in which a repository carries out its own assessment and is then peer-reviewed by a member of the international DSA assessment group. The reviewer recommends to the board whether the guidelines have been complied with and whether the DSA logo can be awarded to the data repository (Harmsen, 2008, p. 1). A number of organizations have already been through this process.[9]

An international joint effort[10] undertaken to develop a set of criteria on which full audit and certification of digital repositories can be based resulted in a 2011 ISO standard in support of the OAIS reference mode. The OAIS reference model standard included a roadmap for follow-on standards which included "standard(s) for accreditation of archives." The ISO 16363 (2011) *Audit and certification of trustworthy digital repositories* is based on the previously mentioned 2007 TRAC checklist, but with more detailed specification of criteria by which digital repositories are to be audited. The scope of the checklist is explicitly the entire range of digital repositories; its criteria are empirically derived and consistent measures of effectiveness have been ascertained. TRAC's evaluative metrics should be used to judge the overall suitability of a repository as being trustworthy to provide a preservation environment that is consistent with the goals of the OAIS. Separately, individual metrics or measures from TRAC can be used to identify possible weaknesses or pending declines in repository functionality.

The same working group has also developed another standard, *Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories* (ISO 16919, 2011). This standard is meant primarily for those setting up and managing organizations that perform the auditing and certification of digital repositories. The standard provides normative rules against which an organization providing audit and certification of digital repositories may be judged, and it describes the auditing process. A team of experts conducted a series of pilot

[9] A list of repositories which have achieved the DSA is at: http://assessment.datasealofapproval.org/seals/ (last accessed 03-21-2012).
[10] The Digital Repository Audit and Certification Wiki is at: http://wiki.digitalrepositoryauditandcertification.org/bin/view (last accessed 03-21-2012).

audits in spring and summer of 2011, to test the methodology promoted by the ISO 16363 standard.[11]

In 2010 a memorandum of understanding (MoU) was signed between three groups working on standards for trusted digital repositories. The CCSDS, the Data Seal of Approval Board, and the DIN "Trustworthy Archives—Certification" Working Group together defined a framework consisting of a sequence of three levels. These levels, in increasing trustworthiness, are documented as follows:

- "Basic Certification" should be granted to repositories that obtain DSA certification through a process of self-audit and the public release of a peer-reviewed statement from another organization which has previously received the DSA;
- "Extended Certification" is granted to Basic Certification repositories that also perform a structured, externally reviewed and publicly available self-audit based on ISO 16363 or DIN 31644; and
- "Formal Certification" is granted to repositories that in addition to Basic Certification obtain full external audit and certification based on ISO 16363 or equivalent DIN 31644.

This MoU was witnessed by the European Commission, but not explicitly endorsed by it.

With the increasing maturity of these two draft international standards, the standardization process for trustworthy digital repositories will have completed its first cycle. Much like the DCC's Curation Lifecycle Model, this cycle of understanding and standardization will continue as an iterative process. With a stable base that includes a process model, relevant standards, and best practices for individual parts of the process, measures of "trustworthiness" will continue to develop as the community's experience with and expertise in digital preservation grows (Dale, Gore, 2010).

Preservation Metadata

Among the many classes of metadata and description standards, preservation metadata has emerged as a separate category. Digital curation requires a provenance mechanism to

[11] These audits were undertaken as part of the EU-funded APARSEN project. Full details of the outcomes will be published later in 2012.

record preservation actions that have been applied to digital objects over time. Early conceptualizations of preservation metadata saw it as "all of the various types of data that allows the re-creation and interpretation of the structure and content of digital data over time" (Ludäsher, Marciano, and Moore, 2001). Thus, preservation metadata spans the popular division of metadata into descriptive, structural, and administrative categories. Understood in this way, it is clear that such metadata must support an extremely wide range of functions, including: discovery, the technical rendering of objects, the recording of contexts and provenance, and documenting the relevant repository policies in place at any given time and the repository actions taken to ensure data integrity. The wide range of functions that preservation metadata is expected to support means that the definition (or recommendation) of standards is not a simple task. The situation is complicated further by the knowledge that different kinds of metadata will be required to support different digital preservation strategies and that metadata standards themselves need to evolve over time (Day, 2005). The OAIS reference model has also become an influential source for preservation metadata standards—Preservation Metadata Implementation Strategies (PREMIS) being the most widely adopted of them.

National and research libraries began to develop preservation metadata standards in the late 1990s with the publication of a number of draft element sets. The National Library of Australia produced the first of these (NLA, 1999), quickly followed by the Cedars and NEDLIB projects (Russell, et al., 2000; Lupovici, Masanès, 2000). An international working group sponsored by RLG and the Online Computer Library Center (OCLC) then built upon these (and other) proposals to produce a unified *Metadata Framework to Support the Preservation of Digital Objects* (OCLC/RLG Working Group on Preservation Metadata, 2002). The National Library of New Zealand, finding past work too theoretical, developed its own preservation metadata element set in 2003 (NLNZ, 2003). While the earlier initiatives had all been informed by the (then) evolving OAIS reference model, the OCLC/RLG Metadata Framework was explicitly structured around the OAIS information model (Day, 2005).

The OAIS reference model provides a functional and information model for a digital archive but it does not define what specific metadata should be collected or how it should be implemented in order to support preservation goals. The

OCLC/RLG Working Group on Preservation Metadata: Implementation Strategies (PREMIS) published its first proposal for core preservation metadata elements in 2005 as the *PREMIS Data Dictionary for Preservation Metadata* (PREMIS Working Group, 2005). The *Data Dictionary* defined preservation metadata as "the information a repository uses to support the digital preservation process," specifically that "metadata supporting the functions of maintaining viability, renderability, understandability, authenticity, and identity in a preservation context" (PREMIS Working Group, 2005). PREMIS defines a common data model to encourage a shared way of thinking about and organizing preservation metadata. The PREMIS data model contains five types of entities: Intellectual Entities, Objects, Rights, Agents and Events. The semantic units that describe the entities in this data model are rigorously defined. PREMIS supports specific implementations through guidelines for metadata management and use, and it puts an emphasis on enabling automated workflows. It makes, however, no assumptions about specific technology, architecture, content type, or preservation strategies. As a result, it is "technically neutral" and supports a wide range of implementation architectures (Dappert, Enders, 2010). PREMIS (2011) is currently maintained by the US Library of Congress and has been translated into multiple languages, including French, German, Italian, and Spanish.

While PREMIS is in use internationally, in Germany a national standard Long Term Preservation Metadata for Electronic Resources (LMER) is more commonly used for preservation metadata. LMER is a standard of the German National Library and is based on a data model originally developed by the National Library of New Zealand. As with PREMIS, each metadata element is associated with a particular type of entity, which in LMER are objects, processes, files, and metadata modification.

Metadata Encoding and Transmission Standard (METS) is a standard for encoding in Extensible Markup Language (XML) the metadata describing or characterizing digital objects. It provides a flexible means of associating all the metadata about a digital object with the object—it is a "container format" specifying how different kinds of metadata can be packaged together (Caplan, 2008). One extension of METS—METSRights provides for the documentation

of the intellectual rights associated with a digital object or its parts.[12]

Currently, few metadata specifications contributing to digital assets' long-term preservation are sanctioned by national or international standards bodies. Some, like PREMIS or METS, have the status of *de facto* standards with well-defined community processes for maintaining and updating them. While communities have a strong desire for long-lasting, stable metadata standards, practices continue to evolve as the number of repository implementations and applications grows. Experience remains too limited to set a preservation metadata standard in stone (Dappert, Enders, 2010).

The OAIS information model continues to influence metadata initiatives, especially in its detailed requirements for comprehensive representation information. It also provides the theoretical basis for projects that aim to capture representation information in terms of file formats.

File Format Description

Format is a fundamental characteristic of a digital object that governs its ability to be used effectively. A number of phases of digital curation—appraisal, selection, acquisition, ingest, preservation, and access—include file format considerations. While preservation planning is a much broader activity that involves many other factors, monitoring for incipient obsolescence is a key activity, especially given that numerous preservation projects have reported difficulties with obtaining complete and reliable file format specifications and documentation (Lawrence, et al., 2000; Representation and Rendering Project, 2003). This has led to calls for the creation of sustainable format repositories to manage representation information about formats so that information will be available for future curation and preservation practitioners (Representation and Rendering Project, 2003; Christensen, 2004b). Such file format registries have emerged to serve the whole preservation community and codify the information that is required about formats for digital curation (cf. Planets, 2008a). However, the present lack of test corpora of digital objects for evaluating file format identification tools

[12] See METS schema: http://www.loc.gov/standards/rights/METSRights.xsd (last accessed 03-21-2012).

demonstrates suggests that the benchmarking of these tools has not yet been treated as a high priority.

The National Archives (TNA) of the UK has developed the PRONOM format registry[13] to provide a service for both human and machine clients. The PRONOM information model manages the relationships between the technical properties of formats, including classification; signatures; software, hardware, and media dependencies; and external entities such as actors, documentation, intellectual property rights, and identifiers which relate to these properties (Brown, 2005).

The development of a similar service, the Global Digital Format Registry (GDFR[14]), was initiated at the Harvard University Library with participation from OCLC. Its goal was to provide sustainable distributed services to store, discover, and deliver representation information about digital formats.

In 2009, the PRONOM and GDFR joined forces under a new name—the Unified Digital Format Registry (UDFR).[15] The UDFR aims to support the requirements and use cases compiled for GDFR and is seeded with PRONOM's software and formats database.

Information on file formats and their use for preservation has also been collected and published by the US Library of Congress as part of the National Digital Information Infrastructure and Preservation Program (NDIIPP).[16] A Registry/Repository of Representation Information has been created by the UK Digital Curation Centre, and focuses upon the representation information listed in the OAIS reference model.[17] The Open Planets Foundation (OPF) has proposed a new concept for representation information registries in digital preservation called "registry

[13] See the PRONOM registry: http://www.nationalarchives.gov.uk/PRONOM/Default.aspx (last accessed 03-21-2012).
[14] See Global Digital Format Registry: http://www.gdfr.info/index.html (last accessed 03-21-2012).
[15] See Unified Digital Format Registry: http://www.udfr.org/ (last accessed 03-21-2012).
[16] See the Library of Congress' Sustainability of Digital Formats: http://www.digitalpreservation.gov/formats/fdd/descriptions.shtml (last accessed 03-21-2012).
[17] See the Registry Web GUI, allowing browsing of the CASPAR/DCC Representation Information Repository: http://registry.dcc.ac.uk/ (last accessed 03-21-2012).

ecosystem." It is based on interlinking various sources of information to create interconnected "registry collections" using Linked Data, rather than creating and maintaining a single registry (OPF, 2011).

The data model of these file format registries has developed over the years to include new aspects of representation information that are required in preservation planning processes. The continuing work on significant properties of digital objects (e.g. the Investigating the Significant Properties of Electronic Content Over Time (InSPECT) project[18]), characterization languages (e.g. the Planets (2008b) project's Extensible Characterization Description language), and development of preservation planning tools (e.g., the Planets project's Plato[19] tool) have further advanced the standardization of information about file formats.

Other Pertinent Standardization

Digital preservation professionals must address numerous dependencies upon systems and processes that were developed by entities not specifically focused on preservation. Successful digital preservation thus relies heavily on standards developed for a variety of purposes.

Many of the existing standards that pertain to archival collections and digital preservation have served primarily to advance work within specific streams of activities, rather than spanning multiple professions. For example, before the recent "recognition that digital preservation poses issues and challenges shared by organizations of all descriptions" and the emerging prominence of the OAIS as a common framework, work on preservation metadata by several organizations was "conducted largely in isolation, lacking any substantial degree of cross-organizational coordination" (OCLC/RLG Working Group on Preservation Metadata, 2002, p.1).

Storage Media

One core set of issues in digital preservation involves the physical medium. The bits stored on an optical or magnetic medium degrade over time and are subject to damage from

[18] See Investigating the Significant Properties of Electronic Content Over Time (InSPECT): http://www.significantproperties.org.uk/ (last accessed 03-21-2012).
[19] See the Plato preservation planning tool: http://www.ifs.tuwien.ac.at/dp/plato/intro.html (last accessed 03-21-2012).

environmental factors. One area in need of standardization was thus the physical storage media and storage conditions (Carneal, 1977). This is the area of digital preservation that has seen the most active standardization and consensus. Standards have been developed and adopted by the Preservation Committee of the Audio Engineering Society (AES), United Nations Educational, Scientific, and Cultural Organization (UNESCO), National Institute of Standards and Technology (NIST) (formerly the National Bureau of Standards), Institute of Electrical and Electronics Engineers (IEEE), American National Standards Institute (ANSI), and the International Organization for Standardization (ISO).

Memory institutions have a long tradition of using their own domain-specific standards for storage, some of which continue to be relevant for handling digital storage media. For example, ISO 11799 *Document storage requirements for archive and library materials* (ISO 11799, 2003) with its national predecessor BS 5454 *Recommendations for storage and exhibition of archival documents* (BS 5454, 2000), set general requirements for storage rooms. Similarly, ISO 18925 (2008), ISO 18938 (2008), ISO/TR 10255 (2009) or BS 4783-8 (1994) all discuss storage requirements of specific storage media types, including magnetic tapes and optical discs.

Data Description, Data Management and Recordkeeping

Standards for descriptive metadata of archival materials have also developed along several distinct paths, based on the boundaries between different types of institution or document. For example, the archival profession developed MAchine Readable Cataloging, Archives and Manuscript Collections (MARC-AMC) (Smiraglia, 1990); Archives, Personal Papers and Manuscripts (APPM) (Hensen, 1989); Encoded Archival Description (EAD) (EAD Working Group, 1998); Rules for Archival Description (RAD) (Duff, 1999); and Describing Archives: A Content Standard (DACS) (2007) in order to develop access systems particular to its collections.

Several standards developed in the last two decades are intended to facilitate the design and management of "recordkeeping systems," which ensure the authenticity of electronic records as evidence. One of the most prominent standardization efforts in this area was a metadata schema for the Commonwealth of Australia (McKemmish et al, 1999). Design

Criteria Standard for Electronic Records Management Software Applications (DOD 5015.02 – STD) provides a set of requirements for the design and certification of applications used to manage electronic records (Assistant Secretary of Defense, 2007). A high-level international standard for records management was adopted in 2001 (ISO 15489, 2001). The body that was responsible for ISO 15489 (TC46/SC11) has subsequently worked on a variety of more specific standards. The Model Requirements for the Management of Electronic Records (MoReq) were released in 2001, updated and substantially revised as MoReq2 in 2008 (MoReq2, 2008) and MoReq2010 in 2011 (MoReq2010, 2011).

Social Science data archivists have also developed metadata standards that cater to the specific types of data residing in their collections, often for the purpose of exchanging data among collections of the same type. The American Council of Social Science Data Archives began discussing options for "study description schemes" at its annual meeting in 1967, and this conversation eventually resulted in a recommended unified scheme (Scheuch, 2003, p. 393). Several generations of proposed conventions for data exchange (De Vries, Van der Meer, 1992; Leighton, 2002; Rasmussen, 1978) and development of codebooks have followed. One important effort to this end is the Data Documentation Initiative (DDI).[20] The first public version of the DDI document type definition (DTD) was published in 2000 (DDI, 2000). Virtually every scientific domain has metadata standards for the description of its data, but not always created within a domain-specific archival context.

Other scientific communities have also followed relatively autonomous paths toward standardization related to their data. For example, the Consultative Committee for Space Data Systems (CCSDS) was formed in 1982, and it then served as an active forum for the development and promulgation of numerous standards for use by space agencies. As described previously, the CCSDS was the body responsible for development of the OAIS Reference Model, which took place between 1994 and 2002; its development involved a level of interaction with other disciplines that had not been the case for any of the CCSDS's previous activities. Space agencies have also developed and adopted several influential standards that have emerged outside of the CCSDS

[20] See Data Documentation Initiative (DDI): http://www.ddialliance.org/ (last accessed 03-21-2012).

process. For example, several separate efforts have attempted to address the need for device-independent data models and software for multidimensional data sets. Common Data Format (CDF) was developed in 1985 by the National Space Science Data Center (NSSDC); Network Common Data Form (NetCDF) was then developed at the Unidata Program Center managed by the University Corporation for Atmospheric Research in Boulder, Colorado; and the Hierarchical Data Format (HDF) was developed at National Center for Supercomputing Applications (NCSA) in 1988. Each initiative boasts a long list of private and public sector adopters. In 1993, NASA chose to adopt HDF for data in its Earth Observing System (EOS), resulting in its own flavor, known as HDF-EOS. Even with this customization of HDF, several actors within the EOS did not perceive it be appropriate to their needs and failed to adopt HDF-EOS (Duerr, et al., 2004).

Standards for File Formats

Using standard file formats that will remain accessible over time is a common digital preservation strategy. Formats that are stable and have been widely adopted are much more likely to be supported over a long period. File format standards that are open often are less likely to become obsolete in a short period, because there is a large user base willing to participate in ensuring that the standards are maintained (Harvey, 2010).

A number of file formats have been explicitly developed through a formal standardization process, such as PNG (ISO/IEC 15948, 2004), JPEG 2000 (ISO/IEC 15444), MPEG-4 (ISO/IEC 14496) and MPEG-7 (ISO/IEC 15938). Others, such as PDF/A (ISO 19005), TIFF/EP (ISO 12234-2, 2001), TIFF/IT (ISO 12639, 2004), Open Document Format (ODF) (ISO/IEC 26300, 2006), and Office Open XML (ISO/IEC 29500) were existing formats, or newly developed variants, that were subsequently promulgated through an accredited standards process.

In contrast to these *de jure* standards, many popular file formats fall into the category of *de facto* standardization on the basis of ubiquity. Although of potentially broad applicability, these standards are generally the result of parochial community interest (Abrams, 2007). For example Scalable Vector Graphics (SVG) (W3C, 2009) is an open file format standard which has not been formalized by a standards organization, and the Broadcast WAVE Format (BWF) was developed by the European Broadcasting

Union to simplify the interchange of broadcast media (EBU, 2011).

While standardization is obviously better than non-standardization, the mere existence of standards does not necessarily mean that they will be widely implemented. For example, the JPEG 2000 image format is an ISO standard, but it is not widely supported by the current generation of Web browsers, although less preservation-friendly formats such as GIF (Graphics Interchange Format), Joint Photographic Experts Group (JPEG), and Portable Network Graphics (PNG) are well supported (Abrams, 2007).

Regardless of their formal status, all of the formats discussed in this section share one important type of openness—whether proprietary or not, their specification documents are published. For most memory institutions, it will be nearly impossible to ensure that all acquired digital objects use file formats that are based on open standards, or indeed even that the formats a curator selects for archiving are all standard formats. The diversity of formats available in the digital domain is so vast and growing at such a rapid pace that the comparatively slow process of standardization does not cover the whole variety of content types and their *en vogue* formats. Thus far the impact of the preservation community on the development of format standards has been quite limited, although rare cases exist in which preservation specialists have been invited to contribute to the standards (e.g., PDF/A, MPEG-7). For practical reasons, many memory institutions have updated their policies by replacing lists of (limited) supported formats and standards, with more detailed criteria for the selection of preservation-friendly formats (e.g., Christensen, 2004b; Brown, 2008; The National Archives, 2011; Arms, et al., n.d.).

Tools for converting from one file format to another or exchanging information between systems remain pertinent for digital preservation.

Representation of Contextual Information

Preservation is a set of activities devoted to ensuring the conveyance of meaning over time. In the case of digital preservation, this involves ensuring that important characteristics and values of digital objects can be consistently reproduced over time within an acceptable range of variability. For a given target digital object, there are contextual information entities that play a role in conveying meaning. Consequently, digital preservation

metadata must convey information about whether or not these contextual information entities exist, and whether they have been altered. There are several types of contextual entities that can be important to describe in order to ensure meaningful use of digital objects over time, including objects, agents, occurrences, purposes, times, places, forms of expression, concepts/abstractions, and relationships (Lee, 2011). It would be neither feasible nor appropriate for digital preservation professionals to attempt to invent their own standards for representing information about such a rich diversity of contextual entities. Fortunately, no such effort is necessary. There are existing and emerging standards that can be applied to information about all of the contextual entity types. The International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR/CPF) (ICA Committee on Descriptive Standards, 2004) and Encoded Archival Context – Corporate bodies, Persons, and Families (EAC-CPF) (EAC Working Group, 2010) are two recent efforts specifically to formalize contextual information related to archival materials.

There are also numerous standards and conventions for representation of information about each of the nine types of contextual entities (Lee, 2011; for further detail, see the appendix to that paper). A few specific examples include:

- **Objects** - There is extensive guidance for generating information about physical objects, including the Global Trade Item (GTIN) for commercial products (GS1 US, 2006), the Categories for the Description of Works of Art (CDWA) (Baca and Harpring, 2009) for art and material culture, and the relatively institution-specific conventions for representing archaeological artifacts (Snow et al., 2006). There are also numerous standards for packing and representation of digital objects already discussed in this essay.

- **Agents** - Librarians and archivists have been working for some time on the elusive goal of uniquely identifying and describing agents over time. An Agents Working Group was formed in 1998, in order to address the agent information that was potentially embedded in (or missing from) the Dublin Core elements (Wilson and Clayphan, 2004). A project within the International Organization for Standardization (ISO) is developing the International Standard Name Identifier (ISNI) (ISO/CD 27729) to uniquely identify "public identities" across multiple areas of creative activity. The International Standard Archival Authority Record for Corporate Bodies, Persons and

Families (ISAAR(CPF)) and Encoded Archival Context: Corporate Bodies, Persons and Families (EAC-CPF) are two rich sources of guidance on the types of information one might hope to provide about agents. Resource Description and Access (RDA) (2011) provides detailed guidance for recording attributes of persons, families, and corporate bodies. In 2006, the Text Encoding Initiative also initiated the Personography Task Force, one product of which has been a report that describes and compares many existing schemes for marking up information about individuals (Wedervang-Jensen and Driscoll, 2006). METS and PREservation Metadata: Implementation Strategies (PREMIS) also provide simple taxonomies for identifying types of agents.

- **Occurrence** - There is a growing body of building blocks for the identification and encoding of occurrence information. Guidance for the detailed representation of processes includes the Process Specification Language (Bock and Gruninger, 2005); extension and application of the Unified Modeling Language (Penker and Eriksson, 2000); XML Process Definition Language (Workflow Management Coalition, 2008); and the Business Process Modeling Notation Specification (White, 2008). TimeML (Pustejovsky et al., 2003) and the Historical Event Mark-up and Linking (HEML) Project (Robertson, 2009) provide conventions for encoding and storage of event information.

- **Purpose** - Two sources of guidance for representing functional entities and their relationships from Australia are the Australian Governments' Interactive Functions Thesaurus (2007) and Keyword AAA (Robinson, 1997), and one from Canada is the Business Activity Structure Classification System (BASCS) Guidance (Library and Archives Canada, n.d.). The International Standard for Describing Functions (ICA Committee on Best Practices and Standards, 2008) has been designed to describe functions within archival information systems.

- **Time** - The most straightforward case of representing time is a precise time and date, as specified in ISO 8601 (2004). However, there is a myriad of other possible temporal units and expressions, which TIMEX2 attempts to accommodate (Ferro, et al., 2005). ISO 19108 (2002) provides detailed guidance for representing "temporal feature attributes, feature operations, and feature associations, and for defining the

temporal aspects of metadata about geographic information," though it is potentially applicable for describing other types of information. The Time Period Directory initiative aims to support translations between common language labels, such as the Civil War, and specific time spans (Petras et al., 2006). There are many other relevant specifications and research activities that fall within the arena of "temporal modeling," which attempt to address the deep connections between events (see above) and time (e.g. Grandi et al., 2005).

- **Place** - There are a number of detailed standards and guidance documents for encoding place information. The Alexandria Digital Library (ADL) project offers a "Guide to the ADL Gazetteer Content Standard" (2004). A well-established set of conventions for encoding locations as coordinates is available in the Department of Defense World Geodetic System 1984 (2000), which is supported by vCard and the geo microformat (Çelik, 2007). vCard also allows for specifying location based on time zone. The X.500 and Lightweight Directory Access Protocol (LDAP) families of standards identify ways to encode geographic and postal addresses. There are several detailed elaborations of places and types of places, including the Alexandria Digital Library Feature Type Thesaurus (2002), Geographic Names Information System, and the Getty Thesaurus of Geographic Names (TGN).

- **Form of Expression** - Many sources of guidance are available for encoding information related to form of expression or genre, with several of the most prominent ones listed in the Library of Congress "Genre/Form Code and Term Source Codes." MARC 21 (2010) also uses fixed-length fields for designating forms of material, has a field for Index Term—Genre/Form, and recently added several fields in the 300 range related to form of expression.

- **Concept or Abstraction** - For several centuries, librarians and other information professionals have been developing and refining systems to represent the concepts and abstractions associated with target information objects. The representation systems have often taken the form of nomenclatures, controlled subject headings, thesauri and, more recently, ontologies. The depth and diversity of standards—ranging from general subject headings for library cataloging to extremely specialized conventions for naming scientific entities—is far too extensive to address here.

- **Relationship** - Thesauri have traditionally expressed three primary types of relationships: equivalence, hierarchical and associative (ISO 2788, 1986). There are innumerable other types of relationships that can hold between entities (e.g. ancestral, emotional, logistical, causal, temporal, and polyhierarchical). Entity-relationship models have long been used to represent relationships of many types, which have generally been implemented using relational databases. Within computer science, the term "ontology" is used to describe data models that accommodate and define an arbitrarily complex set of relationships between entities, concepts, classes or elements. RDA (2008) provides detailed guidance on assigning various types of "relationship designators." In order to make effective use and sense of a digital object, it can be important to differentiate and provide separate information about: 1) the function (purpose), organization (high-level agent) or role responsible for its creation and use, and 2) "personal provenance," i.e., particular individuals involved (Hurley, 1995). Several detailed taxonomies are available for job roles and occupations, including the ERIC Thesaurus, North American Industry Classification System (2007), O*NET Content Model, O*NET-SOC Taxonomy (2009), and Standard Occupational Classification System (2000). METS, Interoperability of Data in E-commerce Systems (INDECS) (Rust & Bide, 2000), OAIS (ISO 14721, 2003), and InterPARES (Long-Term Preservation of Authentic Electronic Records, 2002) all elaborate roles of agents. MARC 21 (2010) includes numerous fields that can be used to identify relationships between the items being catalogued and other resources, as well as allowing for a relator term, which "describes the relationship between a name and a work;" the Library of Congress provides a detailed Relator and Role Code and Term Source Codes. In his investigation of collection relationships, Heaney (2000) also provides a list of "Types of Agent-Object Relationship." The Union List of Artist Names (ULAN) (Harpring et al, 2006) elaborates several dozen roles for use in a Person/Corporate Body record.

Information Security Standards

An emerging area in which memory institutions are subjected to external standards is information security. The international series of information standards (ISO 27000) started off as a British Standard (BS 7799) in the mid 1990's. The ISO 27000 has now

developed into a management system standard with a whole family of standards that support an industry of its own. There are also national versions of benchmarks and standards with which memory institutions may need to comply in their roles as public sector agencies, e.g., information security standards may be compulsory.

Information security standards are for preserving confidentiality, integrity, and the availability of information. However, in general there is a lack of provision for the long-term view in all information security measures that are being enforced on memory institutions except in terms of business continuity plans. The main aim of information security is to ensure protection of existing services and, hence, these standards do not replace preservation standards that the digital preservation community has produced, rather they augment existing standards. Awareness of information security requirements and standards is rising in memory institutions, but competence to apply them and to conceptualize the digital collection management tools in the same framework as other information systems and databases is not yet widespread. A number of procedures required for compliance with the ISO 16363 are closely related to ISO 27001 and could engender greater understanding and conformance to broad information security measures.

Security standards, like many other externally developed standards, suffer from the problem that many memory institutions, because of their lack of awareness of the complex information security requirements, can often take them at face value without consideration of the various contexts in which they can be implemented. Security standards often need significant contextualization to implement them for the needs and requirements of different organizations.

Summary

In summary, the range of standards available for use by and for the digital preservation community is huge and highly diverse in terms of subject and detail. The most significant standard, OAIS, while highly influential, is an informative standard that can be implemented in countless different ways and thus does not ensure interoperability. Standards and guidelines for the audit of organizations carrying out digital preservation have flourished in the environment post-OAIS, but have been hampered by the informative models on which they have been based. At the time of writing, there is increasing convergence on the application of these

standards. In terms of normative standards including those surrounding preservation metadata, the representation of contextual information, file format descriptions, and data and records management are in advanced stages of development, if not implementation.

Currently, there is no easily navigable map of all these standards. The digital preservation community should be in a position now to codify standards that relate to its activities. Digital preservation professionals should also be able to determine which standards are applicable and implement those which are beneficial to their individual organizations: every organization that carries out any digital curation activities should have a list of the appropriate standards which inform its practices, and be able at the least to report on its required level of conformity. We should also understand that standard adherence is more about continuous improvement than getting it right the first time, and it is important to keep a very good watch on what is happening in the future.

Challenges, Gaps, and Opportunities

As a community we often feel that on the one hand, we have too many standards and on the other, too few (of our own standards). The very large number of pertinent standards can fall into agreed categories, but sometimes the exact nature of these categories or classes is confusing. For example, should we treat legislation as a standard? Compliance to standards is voluntary, while legal acts must be adhered to, and in some jurisdictions legislation makes following a standard mandatory, even while not always defining that standard (EC, 2007). Often, what are considered the *de jure* standards are really just standards approved by standards organizations, which are often commercial organizations. How reliable are the *de facto* standards that are endorsed by popular acclaim? When is a standard an open standard? Proprietary standards can be public (e.g., PDF file format specification) but are not considered to be open. Which controlled vocabularies exist within a particular domain? Which standards for the representation of commonly used descriptive information (e.g., country code) is most appropriate for information interchange? What happens when present-day standards no longer address historical information? (For example, the ISO 3166-1 code of YU for Yugoslavia may still be relevant for historical material.) Navigating the standards' library requires

more than basic know-how and even a map, to help decide what is best for our organizations.

Many existing standards seem impractical, with unnecessary detail (e.g., the 16 page explanation of the representation of human sexes in ISO 5218 (2004)), or too technical to be applicable in everyday work of a memory institution (e.g., standards describing how files are written onto optical storage media (ISO/IEC 13490)), but there will be domains in which they are important, and may have applicability for both the preservation and the longer-term interpretation of the items curated. Other standards appear to be directly relevant to digital preservation yet are intended for a very specific purpose or domain (e.g., ISO/TR 15801 (2009) that discusses issues specific to document imaging). We firmly believe that standards should not be hoops to be jumped through—they must be useful and applicable and organizations should know when they are applicable and how to implement them in a pragmatic and consistent form.

In the following paragraphs we present some key issues and challenges facing the digital preservation community relating to standards and standardization.

Establishing Trust

Defining attributes of a trustworthy digital repository has been a discussion topic for more than a decade; several approaches to establishing trustworthiness exist and we are about to have *de jure* standards for measuring it. The European Commission has endorsed a three-tier system of assessment through which organizations can receive a basic certification of trust based on the Data Seal of Approval, an extended certification through adherence to the Data Seal of Approval principles and completion of a self-audit against one of the standards for trusted digital repository auditing (ISO 16363 or DIN 31644), or a formal certification through a formal external audit against one of these two standards.

What these approaches have in common is the underlying thinking that trust is something to be achieved by a standards-based approach to preservation planning. The standards that establish auditing criteria are generic in their nature and do not themselves address the needs of specific domains, cultures or nations (McHugh et al., 2008). This will require interpretation and conceptualization from the different types of digital repositories. The immediate challenge with these standards is how to embed

them within contractual requirements and to get support for implementing them.

Memory institutions satisfy the needs of society by safekeeping and providing access to information. Over time it has become clear that preservation practices—and in particular those practices relating to digital data—have proven more challenging than expected. Is society losing trust in the ability of memory institutions to fulfill their mission? Is this loss of trust factually based? Is there any proof that organizations have really "lost" digital collections? There is an increasing awareness that best practice requires a reliable process that can be verified by an independent body. The independence of verification, auditing, and certification requires a further level of trust in itself. Can one trust the auditing and certifying bodies?

It is important that the emerging audit and certification efforts do not simply create a business model for consultants, auditors, and certification bodies, and pile extraneous and unnecessary requirements onto a repository. Standardization with compliance and certification will hopefully satisfy large groups of stakeholders, but obtaining and retaining certification will require significant resources. Is formal certification the appropriate method of inculcating this trust? Similar to other standards, with their compliance and certification schemes, a trusted repository standard with a mandatory certification could increase operational overhead which might in turn increase the risk of losing digital objects and thus the long-term access to these objects. A proven best practice that evolves into a standard with certification will partially mitigate this risk. Simply implementing external standards that are not wholly relevant can be counter-productive.

One threat to trustworthy digital preservation relates to the continuous availability of highly trained and highly skilled human resources.

Conformity to Preservation Metadata Standards

While a great deal of progress has been made in defining preservation metadata requirements over the last several years, there are several important concerns about preservation metadata. First, there has been limited experience in the application of preservation strategies. This makes it difficult to know whether today's preservation metadata schemes will actually support the process of long-term preservation. Second, neither PREMIS nor LMER define format-specific technical metadata, which is crucial.

Only technical metadata for digital still images is formally standardized (NISO MIX, 2008); specifications for audio, video, text, vector graphics, and other formats are in various stages of development (or not). Third, it is important that the values of preservation metadata elements can be supplied and processed automatically, as many preservation projects will be very large in scale. Hand-entered, natural-language descriptions do not scale. However, there are few standard code lists or controlled vocabularies for the values of even the most important preservation metadata elements (Caplan, 2008).

Applying and Implementing Standards

The digital preservation community is not homogenous—memory institutions sit alongside research and government institutions, businesses, and service providers. Many of the standards that have been developed within this community deal with workflow control, but it is impossible (and undesirable) to completely homogenize preservation workflows across the whole community. This lack of homogeneity presents a paradoxical situation—if the success of a standard is dependent upon its community uptake, and a successful standard is one which gets the widest uptake, the most successful standards are ones which do not address the necessary differences within the community.

The type of standards that the digital preservation community has agreed upon—what could be called voluntary compliance standards—are mainly suitable for improving work processes (in their broadest sense). However, the uniform use of voluntary standards is difficult to coerce. More enforceable standards do exist but usually in areas such as quality control, security, safety, and environment controls, and are mostly implemented in sectors in which something has gone seriously wrong or there are serious threats. Digital preservation as an activity in its own right is, so far, not universally perceived as facing serious threats and, thus, it is, and will be in the future difficult to enforce uniform practices through standards. While we believe in the importance of the establishment of trust within (and between) repositories (see above) we do not believe that "enforceable" but highly homogenized standards will provide a panacea. Conformity in all matters is not necessarily desirable, especially if the primary purpose of the activity of digital preservation is to ensure that the materials being preserved can be used at some defined point in the future for some particular purpose, and that the defined point in the future and the particular purpose will differ across organizations.

Perhaps the biggest technology challenges for the digital preservation community are the *de facto* standards that the IT industry generates with every new format or device. To date it has been very difficult for anyone outside the computer industry and a few large government agencies to have any real impact on these standards. Hence, memory institutions are more "trend followers" than "trend setters" in technical standards, since they form a relatively small part of the market for computer systems and their influence is, correspondingly, relatively small. The digital preservation community can take action in one specific area: playing a more substantial role with standards bodies and working groups organized around formats. Digital preservation professionals could also attempt to detect relevant trends and actions sooner by undertaking active monitoring of technical developments in other domains.

Furthermore, as discussed above, successful digital preservation must be attentive to and draw from standards that were developed for other purposes. Any time a given community or industry has attempted to systematically share data or support interoperability across systems, there is the potential for digital preservation professionals to build upon those efforts rather than trying to invent entirely new or independent standards. This should also be understood as an opportunity for digital preservation specialists to continue to develop specific standards for the community.

Determining the Appropriate Scope for Digital Preservation Standards

A question frequently asked in standards-making is: "What is so special about this that it requires a standard?" The same can be asked about digital preservation—what about it necessitates standardization? Despite cutting across different domains and carrying out digital preservation for somewhat different purposes and in different organizational settings, there are some activities that are requirements for the whole community. However, as a community we do not yet have much experience with identifying situations in which standards are essential. Defining the appropriate level of granularity and detail required in standards remains a challenge—overly prescriptive or overly domain-specific standards will not be applicable across the numerous domains that must preserve digital objects.

Currently, the mainstream thinking on digital preservation is repository-centric—digital preservation is thought of as something that should happen within the digital repository environment. Consequently, most of the existing standards both from within and outside of the community are focused on repositories and how digital objects are managed within them. This is a clear demonstration of the maturity of preservation as the core business of the memory institutions that have successfully been doing it for at least two centuries.

At the moment, digital preservation standards are most successful in addressing issues that are not temporally dynamic. There are separate standards for compatibility, safety, commoditization, etc. However, digital preservation is not a static challenge and the future dependencies on present-day activities should not be underestimated. Will standardization, compliance testing, auditing, and certification provide the ability to address the issue of acting against a moving target? *De facto* technical standards appear and disappear at a rapid rate, and change is their only permanent characteristic. Digital preservation practitioners need to address this and learn to cope with continually changing external standards that represent a considerable organizational and management challenge, not least because withdrawn and obsolete standards are not always retained within any national standards organization registry.

The maturity curve of applying standards starts from testing and benchmarking then moves through risk management towards quality management, eventually reaching an apogee in an organization capable of learning. Standardization in digital preservation is still at the beginning of this curve, focusing primarily on benchmarking the performance of curation tasks and beginning to look at risk management of preservation. The existing "best practice" activities that have been formalized as standards can be tested through self-assessment and essentially peer-reviewing until formal certification bodies have been set up based on ISO 16363 and ISO 16919 auditing standards. An organization that is mature and able to adapt to changes is, however, looking more at efficiency of processes than controls over products (systems) and their interoperability. When moving from quality control to quality assurance to quality management, the management of people and skills becomes the biggest challenge, instead of technology and workflow. Reshaping an organization is more connected to its employees than the technologies it uses.

Demonstrating the benefits of using digital preservation standards remains a challenge. Positive use cases will help to improve user uptake of standards but it is also possible to bring benefits to users by including the concepts and requirements of digital preservation into other standards. The case of records management standards is a good example—aligning digital preservation standards with record-keeping standards can ensure that digital materials are created and maintained in such a way that allows memory institutions to preserve them for the long term. There is a lot of potential in stating the long-term retention requirements and including them in standards of other domains to achieve the aim of making sure that digital content remains accessible.

Next Steps: A Five-Year Forecast

We present four significant areas where we believe the digital preservation community will focus attention over the next five years.[21]

Interoperability Standards

Digital curation will increasingly be seen as an interoperability exercise along the whole chain of steps that form the lifecycle of an object—from its conception to its re-use through the process of preservation. Interoperability in turn requires adherence to standards. Interchange between software systems is currently known (and will increasingly be seen) as the main "at risk" point in the digital object lifecycle, since export-import functionality is generally not supported at a level that is required for legal or scientific requirements in commercial systems. Quality criteria for what must (or at least should) be transferred (i.e., digital objects with their metadata), how the transfer process should be documented, and how the success of this process can be validated, are all becoming urgent issues as larger volumes of content than ever before will be migrated between systems in the coming years. Part of this process is likely to include the education of content creators to help them understand the issues surrounding digital preservation (Van den Eynden et al., 2011).

[21] These recommendations were first outlined in a different form at the standards alignment panel at the ANADP 2011 conference. This discussion benefited greatly from the open discussion that followed the panel session.

Digital Preservation Requirements

Digital preservation standards will move away from a repository-centric world-view and become sets of requirements, that is, functional requirements that can be implemented in other information systems that manage digital assets for the short to medium term. Developing and setting technical and quality criteria/benchmarks for the successful management of digital information in systems that do not behave like collection management tools would be one way to demonstrate the value and feasibility of digital preservation.

Standards for Skill-Sets

Technical quality and success criteria should be accompanied by codes of practice that rely on clear requirements for skills and know-how. This will mean setting standards for education and training courses in digital preservation. Accreditation of digital preservation teaching programs and training courses based on quality criteria or competence standards are beginning to emerge, but should be pursued as an international alignment effort.

Engaging the Users of Standards

In order to better demonstrate the value of standards in digital preservation, the appropriate user communities should be engaged in the discussion of which standards are relevant in practice, which ones are still missing, and who should participate in creating new standards. It is likely that for many practitioners, the broad range of existing standards is fit for purpose, but remaining up to date is a non-trivial task. A standards-watch service providing up-to-date information would prove hugely useful.

Along with guidance and showcase examples, it will also be increasingly important for digital preservation practitioners to know how different sub-communities apply and use existing standards. Standards not only ensure standardization of processes, but also help the “customer” in the broadest sense make most use or take greatest value in the end products of processes.

It is likely that some digital preservation standards will need updating or re-standardization in the near future, because they tend to represent current best practices or current best thinking, rather than being extensively tested and generalized to apply in multiple domains.

Standards' Development

The development of standards is a time-consuming and often tediously bureaucratic process. The procedure followed by ISO is an eye-opener even to the most assiduous onlooker.[22] Tighter international collaboration between groups of experts and practitioners could possibly mitigate some of the delay. Other methods of streamlining or even circumventing the process may actually improve the implementation of standards in a pragmatic and organizationally specific manner.

Conclusions

Success in the arena of digital preservation standards requires identification of common priorities—across professional groups, (digital) heritage institutions, and all the end users, including what may eventually become the most "fickle" and changing group of all—the designated community. The development and implementation of standards should be guided by common priorities. Various parties can benefit from sharing schemas, tools, and methods, while acknowledging the importance and inevitability of different institutional limitations and strengths.

Standards bring about alignment or at least some commonality in thinking, but only if used appropriately and for the purposes that led to their creation. The digital preservation community has succeeded in developing a number of its own standards and often applies them quite successfully. The ease of use and universal applicability of these standards remains a challenge, as does the application of a multitude of external, mostly technical standards that characterize the materials to be preserved.

Just as memory institutions preserve materials that other organizations have created, they also apply many standards that have been created by other communities. Digital preservation is an inclusive domain and as far as standards are concerned cannot (and should not) rely on its own standards alone. Learning to piece together the jigsaw puzzle of standards from different domains is a skill that every digital curation specialist needs to have, alongside the skill of discriminating between what is or is not locally

[22] ISO's International harmonized stage codes are shown visually at: http://www.iso.org/iso/standards_development/processes_and_procedures/stages_description/stages_table.htm (last accessed 03-21-2012).

appropriate. These are the key areas in which the international alignment of efforts can be beneficial, but cannot provide a complete solution.

References

Abrams, Stephen. (2007). *File Formats*. DCC Digital Curation Manual Installment. http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/file-formats/file-formats.pdf (last accessed 08-03-2011).

Alexandria Digital Library Feature Type Thesaurus. (2002). University of California, Santa Barbara. http://www.alexandria.ucsb.edu/gazetteer/FeatureTypes/ver070302/ (last accessed 10-03-2011)

Ambacher, Bruce. (2007). "Government Archives and the Digital Repository Audit Checklist." *Journal of Digital Information* 8, no. 2. http://journals.tdl.org/jodi/article/view/190/171 (last accessed 08-04-2011).

Arms, C.R., Fleischhauer, C., Jones, J. *Sustainability of Digital Formats Planning for Library of Congress Collections*. NDIIPP. http://www.digitalpreservation.gov/formats/index.shtml (last accessed 08-05-2011).

Assistant Secretary of Defense for Networks and Information Integration. (2007). Electronic Records Management Software Applications Design Criteria Standard, DOD 5015.02 – STD. U.S. Department of Defense.

Australian Governments' Interactive Functions Thesaurus – AGIFT. (2007). 2nd ed., National Archives of Australia, Canberra.

Baca, M., Harpring, P. (eds.). (2009). *Categories for the description of works of art.* J. Paul Getty Trust, Los Angeles, CA. http://www.getty.edu/research/publications/electronic_publications/cdwa/index.html (last accessed 10-03-2011).

Bearman, David. (1994). Electronic Evidence: Strategies for Managing Records in Contemporary Organizations. Pittsburgh: Archives and Museum Informatics.

Bock, C., Gruninger, M. (2005). "PSL: a semantic domain for flow models." *Software and Systems Modeling* 4, (2): 209-31.

Bonino, Michal J., Michael B. Spring. (1999). "Standards as Change Agents in the Information Technology Market." *Computer Standards and Interfaces* 20, no. 4-5: 279-89.

Brown, Adrian. (2005). PRONOM 4 Information Model. TNA. http://www.nationalarchives.gov.uk/aboutapps/fileformat/pdf/pronom_4_info_model.pdf (last accessed 08-05-2011).

Brown, Adrian. (2008). *Digital Preservation Guidance Note 1: Selecting file formats for long-term preservation*. The National Archives. http://www.nationalarchives.gov.uk/documents/information-management/selecting-file-formats.pdf (last accessed 08-05-2011).

BS 4783-8 (1994). Storage, transportation and maintenance of media for use in data processing and information storage. Recommendations for 4 mm and 8 mm helical scan tape cartridges.

BS 5454 (2000). Recommendations for storage and exhibition of archival documents.

BS 7799-1 (1995). Information security management. Code of practice for information security management systems.

Byrne, Bernadette M., and Paul A. Golderb. (2002). "The Diffusion of Anticipatory Standards with Particular Reference to the ISO/IEC Information Resource Dictionary System Framework Standard." *Computer Standards and Interfaces* 24, no. 5: 369-79.

Caplan, Priscilla. (2006). *Preservation Metadata*. DCC Digital Curation Manual Installment. http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/preservation-metadata/preservation-metadata.pdf (last accessed 08-03-2011).

Caplan, Priscilla. (2008). *The Preservation of Digital Materials*.

Library Technology Reports 44, no. 2. Chicago: ALA TechSource

Cargill, Carl F. (1989). *Information Technology Standardization: Theory, Process, and Organizations*. Bedford, MA: Digital Press.

Cargill, Carl F. (1997). *Open Systems Standardization: A Business Approach*. Upper Saddle River, NJ: Prentice Hall.

Cargill, Carl F. (1999). "Consortia and the Evolution of Information Technology Standardization." *IEEE Conference on Standardisation and Innovation in Information Technology*, Aachen, Germany, September 15-17, 1999.

Cargill, Carl F. (2011). "Why Standardization Efforts Fail." *Journal of Electronic Publishing* 14, (1).

Carneal, Robert B. (1977). "Controlling Magnetic Tape for Archival Storage." *Phonographic Bulletin* 18: 11-14.

CCSDS. (2009) *Reference Model for an Archival Information System (OAIS). Draft Recommended Standard*. CCSDS 650.0-P-1.1. http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf (last accessed 08-01-2011).

Çelik, T. *Geo – Microformats*. http://microformats.org/wiki/geo (last accessed 08-01-2011).

Chandler, Alfred D., Jr. (1980). "The United States: Seedbed of Managerial Capitalism." In *Managerial Hierarchies: Comparative Perspectives on the Rise of the Modern Industrial Enterprise*, edited by Alfred Dupont Chandler and Herman Daems, 9-40. Cambridge, Mass.: Harvard University Press.

Christensen, Niels H. (2004a). *Towards format repositories for web archives*. 4th International Web Archiving Workshop. http://netarchive.dk/publikationer/FormatRepositories-2004.pdf (last accessed 08-05-2011).

Christensen, Steen S. (2004b). *Archival Data Format Requirements*. The Royal Library, Copenhagen; The State and University Library, Århus, Denmark.

http://netarkivet.dk/publikationer/Archival_format_requirements-2004.pdf (last accessed 08-05-2011).

Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age, National Academy of Sciences. (2009). *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age.* Washington: National Academies Press.

Dale, Robin L., Gore, Emily B. (2010). "Process Models and the Development of Trustworthy Digital Repositories." *Information Standards Quarterly* 22, (2).

DANS. (2009). Data Seal of Approval. http://www.datasealofapproval.org/ (last accessed 08-05-2011).

Dappert, Angela, Enders, Markus. (2010). "Digital Preservation Metadata Standards." *Information Standards Quarterly* 22, (2).

Data Documentation Initiative. DTD. Version 1. (2000). http://ddi-alliance.cvs.sourceforge.net/viewvc/ddi-alliance/ddi/dtd/Version1.dtd?view=log (last accessed 08-05-2011).

Day, Michael. (2005). *Metadata*. DCC Digital Curation Manual. http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/metadata/metadata.pdf (last accessed 08-03-2011).

DCC. (2009). *The DCC Curation Lifecycle Model*. http://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf (last accessed 06-03-2011).

DELOS. (2007). *A Reference Model for Digital Library Management Systems*. http://delos.info/index.php?option=com_content&task=view&id=345&Itemid (last accessed 06-03-2011).

Describing Archives: A Content Standard. (2007). Chicago, IL: Society of American Archivists.

Department of Defense World Geodetic System 1984: Its Definition and Relationships with Local Geodetic Systems, Third Edition, Amendment 1. (2000). NIMA Technical Report TR8350.2. National Imagery and Mapping Agency, St. Louis, MO.

de Vries, Repke, and Cor van der Meer. (1992). "Exchange of Scanned Documentation between Social Scientists and Data Archives: Establishing an Image File Format and Method of Transfer." *IASSIST Quarterly* 16, no. 1-2: 18-22.

DIN 31645. (2011). Information und Dokumentation - Leitfaden zur Informationsübernahme in digitale Langzeitarchive. [*Information and documentation - Guide to the transfer of information objects into digital long-term archives*]

DIN 31644. (2010). Information und Dokumentation - Kriterien für vertrauenswürdige digitale Langzeitarchive. [*Information and documentation - Criteria for trustworthy digital archives*]

DL.org. (2011). *DL.org Reference Model*. http://www.dlorg.eu/index.php/outcomes/reference-modeloutcomes/reference-model (last accessed 06-03-2011).

Dollar, Charles M. (1992). *Archival Theory and Information Technologies: The Impact of Information Technologies on Archival Principles and Methods*. Edited by Oddo Bucci. Vol. 1, Informatics and Documentation Series. Macerata, Italy: Università degli studi di Macerata.

Dollar, Charles M. (1999). *Authentic Electronic Records: Strategies for Long-Term Access*. Chicago, IL: Cohasset Associates.

Dollar, Charles M., and Thomas E. Weir, Jr. (1991). "Archival Administration, Records Management, and Computer Data Exchange Standards: An Intersection of Practice." In *A Sourcebook of Standards Information - Education, Access and Development*, edited by Steven M. Spivak and Keith A. Winsell, 191-211. Boston, MA: G. K. Hall & Co.

Duerr, R., Parsons, M.A., Marquis, M., Dichtl, R., Mullins, T. (2004). "Challenges in long-term data stewardship." In: *Proceedings of 21st IEEE Conference on Mass Storage Systems and Technologies*. NASA/CP-2004-212750.

Duff, Wendy. (1999). "The Acceptance and Implementation of the Rules for Archival Description by Canadian Archives: A Survey." *Archivaria* 47: 27-45.

EBU. (2011). Specification of the Broadcast Wave Format (BWF). A format for audio data files in broadcasting. Version 2.0. EBU-TECH 3285.

EC. (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF (last accessed 10-01-2011).

Edwards, Paul N. (2004). 'A Vast Machine': Standards as Social Technology. *Science* 304, no. 5672: 827-28.

Encoded Archival Context Working Group. (2010). *Encoded Archival Context—Corporate Bodies, Persons, and Families (EAC-CPF) Tag Library*. Encoded Archival Context Working Group of the Society of American Archivists and the Staatsbibliothek zu Berlin. http://www3.iath.virginia.edu/eac/cpf/tagLibrary/cpfTagLibrary.html (last accessed 08-03-2011).

Encoded Archival Description Working Group. (1998). EAD Tag Library for Version 1.0. Chicago, IL: Society of American Archivists and Library of Congress.

ERIC Thesaurus. U.S. Department of Education, Institute of Education Sciences. http://www.eric.ed.gov/ERICWebPortal/thesaurus/thesaurus.jsp (last accessed 08-03-2011).

Ferro, L., Gerber, L., Mani, I., Sundheim, B. and Wilson, G. (2005). *TIDES 2005 Standard for the Annotation of Temporal Expressions*. MITRE Corporation. http://timex2.mitre.org/annotation_guidelines/2005_timex2_standard_v1.1.pdf (last accessed 08-03-2011).

Genre/Form Code and Term Source Codes. Library of Congress, Washington, DC. http://www.loc.gov/standards/sourcelist/genre-form.html (last accessed 08-03-2011).

Geographic Names Information System. U.S. Board on Geographic Names. http://geonames.usgs.gov/ (last accessed 08-03-2011).

Getty Thesaurus of Geographic Names. J. Paul Getty Trust, Los Angeles, CA.

http://www.getty.edu/research/tools/vocabularies/tgn/ (last accessed 08-03-2011).

Gilliland-Swetland, Anne J. (1993). "From Education to Application and Back: Archival Literature and an Electronic Records Curriculum." *American Archivist* 56: 532-44.

Grandi, F., Mandreoli, F. and Tiberio, P. (2005). "Temporal modelling and management of normative documents in XML format." *Data & Knowledge Engineering* 54, (3): 327-354.

GS1 US. (2006). *An introduction to the global trade item number (GTIN)*. Lawrenceville, NJ. http://www.gs1us.org/library?EntryId=31&Command=Core_Download (last accessed 08-03-2011).

Guide to the ADL Gazetteer Content Standard. Version 3.2. (2004). Alexandria Digital Library Project, Santa Barbara, CA. http://www.alexandria.ucsb.edu/gazetteer/ContentStandard/version3.2/GCS3.2-guide.htm (last accessed 08-03-2011).

Harmsen, Henk. (2008). "Data seal of approval - assessment and review of the quality of operations for research data repositories." In: *Proceedings of the iPRES 2008 Conference*, British Library

Harpring, P., Beecroft, A., Johnson, R. and Ward, J. (Eds.). (2006). *Union List of Artist Names: Editorial Guidelines*. J.Paul Getty Trust, Los Angeles, CA.

Harvey, Ross. (2010). *Digital Curation: A How-to-do-it Manual*. Neal-Schuman Publishers.

Heaney, M. (2000). *An Analytical Model of Collections and their Catalogues.* Third Issue. UK Office for Library and Information Networking. http://www.ukoln.ac.uk/metadata/rslp/model/amcc-v31.pdf (last accessed 08-03-2011).

Hedstrom, Margaret L. (1993). "Teaching Archivists About Electronic Records and Automated Techniques: A Needs Assessment." *American Archivist* 56: 424-33.

Hensen, Steven L. (1989). Archives, Personal Papers, and

Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries. 2nd ed. Chicago: Society of American Archivists.

Hofman, Hans; McHugh, Andrew; Ross, Seamus; Ruusalepp, Raivo. (2007). *Digital Repository Audit Method Based on Risk Assessment (DRAMBORA).* DigitalPreservationEurope and the UK Digital Curation Centre

Hurley, C. (1995). Problems with Provenance. *Archives and Manuscripts* 23, (2): 234-259.

ICA Committee on Descriptive Standards. (2004). International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR/CPF). 2nd ed. International Council on Archives, Paris.

ICA Committee on Best Practices and Standards. (2008). *ISDF: International Standard for Describing Functions.* 1st ed. International Council on Archives, Paris.

InterPARES. (2007). *Chain of Preservation (COP) Model*. http://www.interpares.org/ip2/ip2_models.cfm# (last accessed 06-03-2011).

ISO 2788. (1986), Documentation – Guidelines for the establishment and development of monolingual thesauri.

ISO/IEC 5218. (2004). Information technology – Codes for the representation of human sexes.

ISO 8601. (2004). Data elements and interchange formats – Information interchange – Representation of dates and times.

ISO/TR 10255. (2009). Document management applications – Optical disk storage technology, management and standards.

ISO 11799. (2003). Information and documentation – Document storage requirements for archive and library materials.

ISO 12234-2. (2001). Electronic still-picture imaging – Removable memory – Part 2: TIFF/EP image data format.

ISO 12639. (2004). Graphic technology – Prepress digital data exchange – Tag image file format for image technology

(TIFF/IT).

ISO/TR 13028. (2010). Information and documentation – Implementation guidelines for digitization of records.

ISO/IEC 13490. Information technology – Volume and file structure of read-only and write-once compact disk media for information interchange. Parts 1-2.

ISO/IEC 14496. Information technology – Coding of audio-visual objects. Parts 1-27.

ISO 14721. (2003). Space data and information transfer systems – Open archival information system – Reference model.

ISO/IEC 15444. Information technology – JPEG 2000 image coding system. Parts 1-13.

ISO 15489. (2001). Information and Documentation – Records Management. Parts 1 and 2.

ISO/IEC 15938. Information technology – Multimedia content description interface. Parts 1-12.

ISO/IEC 15948. (2004). Information technology – Computer graphics and image processing – Portable Network Graphics (PNG): Functional specification.

ISO/TR 15801. (2009). Document management – Information stored electronically – Recommendations for trustworthiness and reliability.

ISO 16363. (2011). Space data and information transfer systems - Audit and certification of trustworthy digital repositories.

ISO 16919. (2011). Space data and information transfer systems - Requirements for bodies providing audit and certification of candidate trustworthy digital repositories.

ISO 18925. (2008). Imaging materials – Optical disc media – Storage practices.

ISO 18938. (2008). Imaging materials – Optical discs – Care and handling for extended storage.

ISO 19005. Document management – Electronic document file format for long-term preservation. Parts 1-2.

ISO 19108. (2002). Geographic information – Temporal schema.

ISO 20652. (2006). Space data and information transfer systems –

Producer-archive interface – Methodology abstract standard.

ISO/IEC 26300. (2006). Information technology – Open Document Format for Office Applications (OpenDocument) v1.0.

ISO/IEC 27000. (2009). Information technology – Security techniques – Information security management systems.

ISO/CD 27729. (2008). Information and Documentation – International Standard Name Identifier (ISNI).

ISO/IEC 29500. Information technology – Document description and processing languages – Office Open XML File Formats. Parts 1-4.

Jacobs, Richard. (1961). *On Records Preparation and Magnetic Tape*. Washington, DC: National Archives and Records Service.

Klump, J. (2011, January/February). "Criteria for the Trustworthiness of Data Centres." *D-Lib Magazine.* 17,(1/2). http://www.dlib.org/dlib/january11/klump/01klump.print.html (last accessed 08-02-2011).

Lawrence, G., Kehoe, W., Rieger, O., Walters, W., Kenney, A. (2000). *Risk Management of Digital Information: A File Format Investigation.* Council on Library and Information Resources. http://www.clir.org/pubs/reports/pub93/pub93.pdf (last accessed 08-02-2011).

Lee, Christopher A. (2009). "Open Archival Information System (OAIS) Reference Model." In *Encyclopedia of Library and Information Sciences, Third Edition*, edited by Marcia J. Bates and Mary Niles Maack, 4020-4030. Boca Raton, FL: CRC Press.

Lee, Christopher A. (2011). "A Framework for Contextual Information in Digital Collections." *Journal of Documentation* 67, no. 1: 95-143.

Leighton, Vernon. (2002). "Developing a New Data Archive in a Time of Maturing Standards." *IASSIST Quarterly* 26: 5-9.

Library and Archives Canada. (n.d.). *Business activity structure classification system (BASCS) guidance*. Retrieved October 3, 2011, from

http://www.collectionscanada.gc.ca/government/products-services/007002-2089-e.html

LMER:http://www.dnb.de/EN/Standardisierung/LMER/lmer_node.html (last accessed 03-21-2012).

Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project. (2002), Vancouver, Canada. http://www.interpares.org/book/index.cfm (last accessed 08-02-2011).

Ludäsher, B., Marciano, R., Moore, R. (2001). "Preservation of digital data with self-validating, self-instantiating knowledge-based archives." *SIGMOD Record* 30, (3): 54-63.

Lupovici, C., Masanès, J. (2000). *Metadata for the long term preservation of electronic publications*. NEDLIB Report series 2. The Hague: Koninklijke Bibliotheek. http://www.kb.nl/hrd/dd/dd_links_en_publicaties/nedlib/NEDLIBmetadata.pdf (last accessed 08-02-2011).

MARC 21 Format for Bibliographic Data. (2010). U.S. Library of Congress, Washington, DC. http://www.loc.gov/marc/bibliographic/ecbdhome.html (last accessed 08-03-2011).

McHugh, A., Ross, S., Innocenti, P., Ruusalepp, R., Hofman, H. (2008). "Bringing Self-assessment Home: Repository Profiling and Key Lines of Enquiry within DRAMBORA." *International Journal of Digital Curation.* 2, (3). http://www.ijdc.net/index.php/ijdc/article/view/93 (last accessed 08-06-2011).

McKemmish, Sue, Acland, Glenda, Ward, Nigel, Reed, Barbara. (1999). "Describing Records in Context in the Continuum: The Australian Recordkeeping Metadata Schema." *Archivaria* 48: 3-43.

nestor:http://www.langzeitarchivierung.de/Subsites/nestor/EN/Home/home_node.html (last accessed 03-21-2012).

NISO Metadata for Images in XML (2008). http://www.loc.gov/standards/mix/ (last accessed 01-12-2012).

Model Requirements for the Management of Electronic Records:

Update and Extension (Moreq2 Specification). (2008). Hampshire, United Kingdom: Serco Consulting.

MoReq2010. Modular Requirements for Records Systems. Volume 1. Core Services and Plug-in Modules. (2011). DLM Forum Foundation. http://moreq2010.eu/pdf/MoReq2010-Core+Plugin(v1-0).pdf (last accessed 08-11-2011).

The National Archives. (2011). *Evaluating Your File Formats*. Version 1.2. http://www.nationalarchives.gov.uk/documents/information-management/evaluating-file-formats.pdf (last accessed 08-05-2011).

National Library of Australia. (1999). *Preservation Metadata for Digital Collections. Exposure Draft*. Retrieved August 4, 2011, from http://www.nla.gov.au/preserve/pmeta.html

National Library of New Zealand. (2003). *Metadata Standards Framework - Preservation Metadata (Revised)*. Retrieved August 4, 2011, from http://www.natlib.govt.nz/downloads/metaschema-revised.pdf

nestor. (2006). *Criteria for Trusted Digital Long-Term Preservation Repositories - Version 1* (Request for Public Comment), edited by nestor - Network of Expertise in Long-Term Storage of Digital Resources and nestor Working Group on Trusted Repositories Certification, nestor materials 8. Retrieved on August 5, 2011, from http://nbn-resolving.de/urn:nbn:de:0008-2006060703

nestor. (2008). Kompetenznetzwerk Langzeitarchivierung / Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung: nestor-Kriterien, Kriterienkatalog vertrauenswürdige digitale Langzeitarchive. Version 2. Retrieved on August 5, 2011, from http://www.nbn-resolving.de?urn:nbn:de:0008-2008021802

North American Industry Classification System. (2007). U.S. Census Bureau. Washington, DC. http://www.census.gov/eos/www/naics/ (last accessed 10-02-2011).

OCLC/RLG Working Group on Preservation Metadata. (2002). "Preservation Metadata and the OAIS Information Model, A Metadata Framework to Support the Preservation of

Digital Objects." http://www.oclc.org/research/projects/pmwg/pm_framework.pdf (last accessed 09-29-2011).

*O*NET Content Model*. National O*NET Consortium. http://www.onetcenter.org/content.html (last accessed 10-03-2011).

*O*NET-SOC Taxonomy*. (2009). National O*NET Consortium. http://www.onetcenter.org/taxonomy.html (last accessed 10-03-2011).

Open Planets Foundation. (2011). A New Registry for Digital Preservation: Conceptual Overview. http://openplanetsfoundation.org/sites/default/files/OPF_A_New_Registry_Conceptual_Overview_v1%201.pdf (last accessed 08-05-2011).

Oßwald, A., Das Referenzmodell. "OAIS – Open Archival Information System." In: Neuroth, H., Oßwald, A., Scheffel, R., Strathmann, S., Huth, K. (2010) *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. *Version 2.3*. nestor. http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010071949 (last accessed 08-02-2011).

Penker, M., Eriksson, H.-E. (2000). *Business Modeling with UML: Business Patterns at Work.* John Wiley & Sons, New York, NY.

Petras, V., Larson, R.R., Buckland, M. (2006). "Time period directories: a metadata infrastructure for placing events in temporal and geographic context." In: *Opening Information Horizons*: 6th ACM/IEEE-CS Joint Conference on Digital Libraries: June 11-15, 2006, Chapel Hill, NC, USA: JCDL 2006, ACM Press, New York, NY, pp. 151-60.

Planets project. (2008a). *White Paper: Representation Information Registries*. http://www.planets-project.eu/docs/reports/Planets_PC3-D7_RepInformationRegistries.pdf (last accessed 08-05-2011).

Planets project. (2008b). *Final XCDL Specification*. http://www.planets-project.eu/docs/reports/Planets_PC2-D7_FinalXCDLSpec_Ext.pdf (last accessed 08-02-2011).

PREMIS Working Group. (2005). *Data Dictionary for Preservation Metadata*. Version 1.0. http://www.oclc.org/research/activities/past/orprojects/pmwg/premis-final.pdf (last accessed 08-02-2011).

PREMIS Editorial Committee. (2011). *PREMIS Data Dictionary for Preservation Metadata*. Version 2.1. http://www.loc.gov/standards/premis/v2/premis-2-1.pdf (last accessed 08-02-2011).

Pustejovsky, James, Castaño, José, Ingria, Robert, Saurí, Roser, Gaizauskas, Robert, Setzer, Andrea, Katz, Graham. "TimeML: robust specification of event and temporal expressions in text." In: Bunt, H., van der Sluis, I. and Morante, R. (Eds), *Proceedings of the IWCS-5 5th International Workshop on Computational Semantics*, Tilburg University, Computational Linguistics and AI Group, Tilburg, The Netherlands.

Rasmussen, Karsten Boye. (1978). "Technical Standards for Magnetic Tape Exchange between Data Organizations." *IASSIST Quarterly* 2, no. 3: 76-77.

RDA: Resource Description and Access. (2011). Chicago: American Library Association.

Relator and Role Code and Term Source Codes. U.S. Library of Congress. http://www.loc.gov/standards/sourcelist/relator-role.html (last accessed 10-03-2011).

Representation and Rendering Project. (2003). *Survey and assessment of sources of information on file formats and software documentation*. University of Leeds. http://www.jisc.ac.uk/media/documents/programmes/preservation/fileformatsreport.pdf (last accessed 08-05-2011).

Research Libraries Group and National Archives and Records Administration Digital Repository Certification Task Force. (2007). *Trustworthy Repository Audit and Certification (TRAC): Criteria and Checklist*. http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf (last accessed 08-05-2011).

Robertson, B. (2009). "Exploring historical RDF with HEML." *Digital Humanities Quarterly*. 3, (1)

Robinson, C. (1997). "Records control and disposal using functional analysis." *Archives and Manuscripts*. 25, (2):

288-303.

Rothenberg, Jeff. (1999). "Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation." Washington, DC: Council on Library and Information Resources.

Russell, K., Sergeant, D., Stone, A., Weinberger, E., Day, M. (2000). *Metadata for digital preservation: the Cedars project outline specification*. CEDARS, University of Leeds

Rust, G., Bide, M. (2000). *The <indecs> Metadata Framework: Principles, Model and Data Dictionary*. WP1a-006-2.0. http://www.doi.org/topics/indecs/indecs_framework_2000.pdf (last accessed 10-03-2011).

Scheuch, Erwin K. (2003). History and Visions in the Development of Data Services for the Social Sciences. *International Social Science Journal* 55, no. 3: 385-99.

Schumny, Harald. (2002). "Standards Developed Ahead of Technology." *Computer Standards and Interfaces* 24, no. 5: 363-67.

Smiraglia, Richard P., ed. (1990). *Describing Archival Materials: The Use of the MARC AMC Format*. New York: Haworth Press.

Snow, D.R., Gahegan, M., Giles, C.L., Hirth, K.G., Milner, G.R., Mitra, P., Wang, J.Z. (2006). Cybertools and archaeology. *Science.* 311, (5763): 958-9.

Spring, Michael B., Weiss, Martin B.H. (1994). Financing the Standards Development Process. In: *Standards Policy for Information Infrastructure*, edited by Brian Kahin and Janet Abbate, 289-320. Cambridge, MA: MIT Press.

Standard Occupational Classification System. (2000). U.S. Department of Labor, Bureau of Labor Statistics. Washington, DC. http://stats.bls.gov/soc/ (last accessed 10-03-2011).

Stielow, Frederick J. (1992). Archival Theory and the Preservation of Electronic Media: Opportunities and Standards Below the Cutting Edge. *American Archivist* 55: 332-43.

"*Taking a Byte out of History: the Archival Preservation of Federal Computer Records*." House Report No.101-978,

101st Cong., 2nd Sess., Washington, D.C.: U.S. Government Printing Office, 1990.

Task Force on Archiving of Digital Information. (1996). *Preserving Digital Information*. Report of the Task Force on Archiving of Digital Information commissioned by The Commission on Preservation and Access and The Research Libraries Group. Retrieved August 2, 2011, from http://www.clir.org/pubs/reports/pub63watersgarrett.pdf

Tiemann, Michael (2005). "An Objective Definition of Open Standards." *Computer Standards and Interfaces* 28, no. 5: 495-507.

Tolk, Andreas, Muguira, James A. (2003). *The Levels of Conceptual Interoperability Model*. Simulation Interoperability Workshop, Orlando, Florida.

Updegrove, Andrew. (1995). "Consortia and the Role of the Government in Standard Setting." In *Standards Policy for Information Infrastructure*, edited by Brian Kahin and Janet Abbate, 321-48. Cambridge, MA: MIT Press.

Van den Eynden, Veerle, Corti, Louise, Woollard, Matthew, Bishop, Libby, Horton, Laurence. (2011). *Managing and Sharing Data. Best Practice for Researchers.* Colchester, Essex: UK Data Archive.

W3C. (2011). *Scalable Vector Graphics (SVG) 1.1* (Second Edition). Retrieved October 1, 2011, from http://www.w3.org/TR/SVG11/ (last accessed 10-01-2011).

Walch, Victoria Irons. (1990). The Role of Standards in the Archival Management of Electronic Records. *American Archivist* 53: 30-43.

Walch, Victoria Irons. (1993). Automated Records and Techniques Curriculum Development Project: Committee on Automated Records and Techniques. *American Archivist* 56, no. 3: 468-505.

Wedervang-Jensen, E., Driscoll, M. (2006). *Report on XML mark-up of biographical and prosopographical data.* Text Encoding Initiative. http://www.tei-c.org/Activities/Workgroups/PERS/persw02.xml (last accessed 10-03-2011).

Weiss, Martin B.H., and Carl Cargill. (1992). "Consortia in the Standards Development Process." *Journal of the American Society for Information Science* 43, no. 8: 559-65.

White, S.A. (2008). *Business Process Modeling Notation.* version1.1. Object Management Group. Needham, MA.

Wilson, A., Clayphan, R. (2004). *Functional requirements for describing agents*. Draft 2. Dublin Core Metadata Initiative – Agents Working Group. http://dublincore.org/groups/agents/agentFRdraft2-2.html (last accessed 10-03-2011).

Workflow Management Coalition. (2008). *XML Process Definition Language*. Version 2.1. Workflow Management Coalition, Hingham, MA.

Yates, JoAnne. (1989). *Control through Communication*. Baltimore: Johns Hopkins University Press.

Zuboff, Shoshana. (1988). *In the Age of the Smart Machine: The Future of Work and Power.* New York: Basic Books.