DOWNLOAD PDF METADATA FOR THE LONG TERM PRESERVATION OF ELECTRONIC PUBLICATIONS

Chapter 1 : Best Practices for Digital Archiving: An Information Life Cycle Approach

The second main objective of the project is to address the issue of long-term digital preservation. Work in this area should provide better insight into the pro's and con's of different long-term preservation strategies as applied to digital deposit collections.

We are, to my mind, living in the midst of digital Dark Ages; consequently, much as monks of times past, it falls to librarians and archivists to hold to the tradition which reveres history and the published heritage of our times. However, digital information is fragile in ways that differ from traditional technologies, such as paper or microfilm. It is more easily corrupted or altered without recognition. Digital storage media have shorter life spans, and digital information requires access technologies that are changing at an ever-increasing pace. Some types of information, such as multimedia, are so closely linked to the software and hardware technologies that they cannot be used outside these proprietary environments [Kuny]. Because of the speed of technological advances, the time frame in which we must consider archiving becomes much shorter. The time between manufacture and preservation is shrinking. While there are traditions of stewardship and best practices that have become institutionalized in the print environment, many of these traditions are inadequate, inappropriate or not well known among the stakeholders in the digital environment. Originators are able to bypass the traditional publishing, dissemination and announcement processes that are part of the traditional path from creation to archiving and preservation. Groups and individuals who did not previously consider themselves to be archivists are now being drawn into the role, either because of the infrastructure and intellectual property issues involved or because user groups are demanding it. Librarians and archivists who traditionally managed the life cycle of print information from creation to long-term preservation and archiving, must now look to information managers from the computer science tradition to support the development of a system of stewardship in the new digital environment. There is a need to identify new best practices that satisfy the requirements and are practical for the various stakeholder groups involved. This study is the most recent in a series of efforts on the part of ICSTI to highlight the importance of digital archiving. Based on common interest in this topic, CENDI, an interagency working group of scientific and technical information managers in the U. Over 30 projects were identified, from which 18 were selected as the most "cutting edge". The highlighted projects covered six countries U. They came from a variety of sectors including government scientific and technical programs, national archives, national libraries, publishers, and research institutes. Project managers from the selected projects were asked a series of questions aimed at identifying emerging models and best practices for digital archiving. While technologies for storage and retrieval were discussed, technology was of secondary interest to the understanding of policy and practice. For purposes of the study, "digital archiving" was defined as the long-term storage, preservation and access to information that is "born digital" created and disseminated primarily in electronic form or for which the digital version is considered to be the primary archive. Primary attention was given to operational and prototype projects involving scientific and technical information at an international level. It included a variety of digital format types applicable to scientific and technical information, including data, text, images, audio, video and multimedia; and a variety of object types, such as electronic journals, monographs, satellite imagery, biological sequence data, and patents. The results, while not scientifically valid, identify emerging models and best practices for digital archives in an effort to support the development of a tradition of digital stewardship. The producer may be a human author or originator, or a piece of equipment such as a sensing device, satellite or laboratory instrument. Creation is viewed here in the broadest sense, as increasingly science is based on a variety of data types, products and originators. All project managers acknowledged that creation is where long-term archiving and preservation must start. Even in rigorously controlled situations, the digital information may be lost without the initial awareness on the part of the originator of the importance of archiving. Practices used when a digital object is created ultimately impact the ease with which the object can be digitally archived and

preserved. In addition, there are several key practices involving the creator that are evolving within the archiving projects. First, the creator may be involved in assessing the long-term value of the information. Secondly, the preservation and archiving process is made more efficient when attention is paid to issues of consistency, format, standardization and metadata description in the very beginning of the information life cycle. Limits are placed on both the software that can be used and on the format and layout of the documents in order to make short and long-term information management easier. However, only in the case of data objects is the metadata routinely collected at the point of creation. Many of the datasets are created by measurement or monitoring instruments, and the metadata is supplied along with the data stream. This may include location, instrument type, and other quality indicators concerning the context of the measurement. In some cases, this instrument-generated metadata is supplemented by information provided by the original researcher. For smaller datasets and other objects such as documents and images, much of the metadata continues to be created "by hand" and after-the-fact. Metadata creation is not sufficiently incorporated into the tools for the creation of these objects to rely solely on the creation process. As standards groups and vendors move to incorporate XML eXtensible Mark-up Language and RDF Resource Description Framework architectures in their word processing and database products, the creation of metadata as part of the origination of the object will be easier. The object must be known to the archive administration. There are two main aspects to the acquisition of digital objects -- collection policies and gathering procedures. Guidelines help to establish the boundaries in such an unregulated situation. It is also the case that there is just too much material that could be archived from the Internet, so guidelines are needed to tailor the general collection practices of the organization. The collection policies answer questions related to selecting what to archive, determining extent, archiving links, and refreshing site contents. In the network environment, any individual with access to the Internet can be a publisher, and the network publishing process does not always provide the initial screening and selection at the manuscript stage on which libraries have traditionally relied in the print environment Selection policies are, therefore, needed to ensure the collection of publications of lasting cultural and research value. These guidelines are key to the successful networking of the state libraries into the National Collection of Australian Electronic Publications, since they provide consistency across multiple acquisition activities. Scholarly publications of national significance and those of current and long term research value are archived comprehensively. Other items are archived on a selective basis "to provide a broad cultural snapshot of how Australians are using the Internet to disseminate information, express opinions, lobby, and publish their creative work. What is the extent or the boundary of a particular digital work? This is particularly an issue when selecting complex Web sites. Both higher and lower links on the site are explored to establish which components form a title that stands on its own for the purposes of preservation and cataloguing. However, sometimes the components of larger publications or sites do not stand well on their own but together do form a valuable source of information. In this case, if it fits the guidelines, the site should be selected for archiving as an entity. This issue has been addressed by the selected projects in a variety of ways. Most organizations archive the links the URLs or other identifiers but not the content of the linked objects. The American Institute of Physics archives the links embedded in the text and references of its electronic journal articles but not the text or content of any of these links, unless the linked item happens to be in its publication archive or in the supplemental material which it also archives. In a slightly different approach, the NLC has chosen to archive the text of the linked object only if it is on the same server as the object that is being archived. The NLC cites difficulties in tracking down hypertext links and acquiring the linked objects as the reason for its decision not to include the content of other links. The previous issue of the same periodical, accessed through a hypertext link, would be considered a part of the original publication. Another publication accessed through a hypertext link would not be considered part of the original publication. Only two of the reviewed projects archive the content of all links. Within a specific domain, the American Astronomical Society also maintains all links to both documents and supporting materials in other formats, based on extensive collaboration among the various international astronomical societies, researchers,

universities, and government agencies. Each organization archives its own publications, but links are maintained not only from references in the full text and cited references of the articles, but between and among the major international astronomical databases. Within this specific domain, the contents of all linked objects are available. A balance must be struck between the completeness and currency of the archive and the burden on the system resources. Obviously, the burden of refreshing the content increases as the number of sources stored in the archive increases. For example, NLA allocates a gathering schedule to each "publication" in its automatic harvesting program. The selection is dependent on the degree of change expected and the overall stability of the site. In the case of the NLA, the sites are reviewed and hand-selected. They are monitored for their persistence before being included in the archive. Alternatively, the Royal Library, the National Library of Sweden, acquires material by periodically running a robot to capture sites for its Kulturarw3 project without making value judgments [National Library of Sweden]. The harvester automatically captures sites from the. In addition, some material is obtained from foreign sites with material about Sweden, such as travel information or translations of Swedish literature. While the acquisition is automatic, the National Library gives priority to periodicals, static documents, and HTML pages. Conferences, usenet groups, ftp archives, and databases are considered lower priority. However, the guidelines from EVA address issues to be considered when using robots for harvesting. In order not to overload the servers being harvested, particularly those belonging to the public networks, the EVA guidelines establish time limits between visits to a single Web server and between capturing and recapturing a single URL. Even though this approach has allowed the EVA project to progress, developers at EVA consider this approach to be "very rough and not flexible enough for archiving purposes. The flexibility would require that the scheduler be a database application that can be modified by the librarian. The approaches to intellectual property vary based on the type of organization doing the archiving. In the case of data centers or corporate archives where there is a close tie between the center and the owner or funding source, there is little question about the intellectual property rights related to acquisition. However, in the case of national libraries, the approaches to intellectual property rights differ from country to country. The differences are based on variant national information policies or legal deposit laws. In many countries, the law has not yet caught up with the digital environment, and the libraries must make their own decisions. In the absence of digital deposit legislation, the PANDORA Project seeks permission from the copyright owner before copying the resource for the archive. In contrast, the Swedish and Finnish national library projects have an automated system and do not contact the owners. Both identification and cataloging allow the archiving organization to manage the digital objects over time. Identification provides a unique key for finding the object and linking that object to other related objects. Cataloging in the form of metadata supports organization, access and curation. Cataloging and identification practices are often related to what is being archived and the resources available for managing the archive. There are issues related to how the metadata is created, the metadata standards and content rules that are used, the level at which metadata is applied and where the metadata is stored. The majority of the projects created metadata in whole or part at the cataloging stage. However, there is increasing interest in automatic generation of metadata, since the manual creation of metadata is considered to be a major impediment to digital archiving. A project is underway at the U. Environmental Protection Agency to derive metadata at the data element level from legacy databases. Department of Defense is also investigating automated metadata generation. A variety of metadata formats are used by the selected projects, depending on the data type, discipline, resources available, and cataloging approaches used. Most national libraries use traditional library cataloging standards with some fields unable to be filled and others taking on new meaning. However, several newer abbreviated formats developed specifically for Web-based resources are also in use.

Chapter 2 : \u22Libraries, metadata and preservation of electronic resources \u22 - CORE

Note: Citations are based on reference standards. However, formatting rules can vary widely between applications and fields of interest or study. The specific requirements or preferences of your reviewing publisher, classroom teacher, institution or organization should be applied.

Eight national libraries in Europe, one national archive, two ICT organisations and three major publishers are participating in the project. The project addresses major technical issues confronting national deposit libraries that are in the process of extending their deposit, whether by legal or voluntary means, to digital works [ref. One important piece of work being carried out by the project is the functional specification and overall design of a DSEP. The main objective is to identify functional requirements that are common to all deposit libraries in order to arrive at a "generic" high-level design of a DSEP that can serve as a basis for local implementations by individual deposit libraries. A common workflow for handling deposited electronic publications was defined and helped to identify common functional requirements. Now work is being carried out to detail a DSEP process model and data model, based on the OAIS framework, applicable to all deposit libraries, and detailed enough to enable consistent implementation design and development work. The second main objective of the project is to address the issue of long-term digital preservation. The characteristics of electronic publications and other categories of digital deposit material and their associated preservation and authenticity requirements need to be defined. The NEDLIB partners recognise that many aspects, including cost-effectiveness, legal restrictions, agreements with publishers, and user access requirements, ultimately need to be taken into account when policy choices for preservation strategies are set. The Koninklijke Bibliotheek has taken a first tentative step to help define and test the technicalities of preservation mechanisms by starting an emulation experiment with Jeff Rothenberg. Besides work on abstract modelling and experimental preservation strategies, NEDLIB is very much geared towards producing pragmatic, ready-to-use results. Recommended standards and conventions for technical solutions are documented in order to provide deposit libraries with practical guidelines when implementing a DSEP. Practical experiences, technical infrastructures and organisational approaches taken by individual NEDLIB partners are gathered and compiled in such a way that these experiences can be of use to other libraries. The third and last main objective of the project is to build a demonstrator system, with tools and software already in use by project partners or developed by NEDLIB, covering all functional aspects of a DSEP. Software and tools are being developed, tested and integrated in functional building blocks of the demonstrator. Existing library systems, such as the online public access catalogue OPAC and the library acquisition and cataloguing systems, which are external to, but need to interact with a DSEP, will interface to the demonstrator. During the demonstration stage, the handling of electronic publications from acquisition to access will be demonstrated, with sample material provided by Elsevier Science, Kluwer Academic Publishers and Springer-Verlag. In this article I will expand a little on the first two work areas: It establishes a common framework for functional and information modelling concepts applicable to any archive. It is specifically applicable to organisations that have a responsibility to provide long-term access to digital information. As such, the OAIS model is relevant to deposit libraries. The prospect that such a model can provide a solid basis for standardisation within digital archives and promote greater vendor awareness and support of archival requirements, was decisive for the NEDLIB partners. Most of the functionality relating to the selection and description of digital works, the creation of finding aids such as bibliographies, catalogues, subject-guides and indexes , and the provision of user access, is part of the broader digital library configuration. Additionally, it is important to specify how a DSEP interfaces with the digital library system. Process model for a DSEP The workflow for handling electronic publications from selection for inclusion in the deposit collection to end-user access has been detailed into a prototype process of 13 steps. This process has been mapped to the OAIS set of functional entities. Figure 1 shows the result of this exercise. This interface may need to generate, if necessary, accompanying instructional data, in order for the

Ingest module of the DSEP to be able to process the publication properly. This "pre-processing" interface is needed because deposit libraries cannot dictate submission formats to publishers: Most development work at deposit libraries presently concentrates on this interface. For publishers, this interaction helps them to redesign their publishing process according to higher quality standards. In some cases a SIP may contain only metadata. The interface takes care of all processes needed to unpack a DIP and to make it fit for use by the library visitor. Through this interface, deposited material can be made available, taking account of all kinds of access variables of the digital library environment, such as user authorisation, user access rights, publisher license access conditions and other access controls. Presently, for example, deposit license agreements with publishers only permit installation of publications onsite, on a library workstation, and access by registered library users. This "post-processing" interface is needed because deposit libraries cannot anticipate all access modes and future variables. This interface also transfers, upon request, metadata from the DSEP through to other systems that need to process the data, either within the digital library system or external to it, such as systems from bibliographic utilities. Usually this concerns metadata uploads from DSEP to other systems. In some cases, it may also involve passing a whole publication through to a content indexing system, in order to generate, for example, a full-text index of the publication. The need for this additional module is clarified below. Ingest unpacks and verifies the publication, collects, generates and re-distributes data to other processes. Routines include integrity check of the medium, of the file formats and of the logical document structure. The process identifies the informational contents, the primary metadata, special access controls to be placed on the contents, abstracts, full-text indexes and other additional data accompanying the publication, technical data for installation and de-installation. The different data are copied to and processed in different environments for cataloguing, for access control, and for finding aids. In the process, the publication is installed and de-installed and its authenticity is established and recorded. This module consists of all procedures necessary for the secure storage of the electronic publication in the digital store, including storage management procedures, quality assurance, disaster recovery, etc. It also includes regular medium migration, in order to preserve the bit stream of a publication from decaying carriers. Data-Management Data-Management mainly stores and retrieves metadata. We distinguish between two types of metadata: The metadata associated with the publication may also be duplicated in other external systems. The cataloguing process, which creates a title-description of the electronic publication and also involves subject indexing, takes place in the cataloguing environment of the digital library system. It may re-use primary metadata provided by the publisher and return descriptive metadata to the DSEP system, through the Delivery and Capture interface. Access In the DSEP model the Access module is much more limited than in the OAIS model, because many related processes belong intrinsically to the digital library environment and not specifically to a DSEP, such as creating finding aids, registering library users, applying access controls, etc. This may entail extracting parts of the electronic publication, or adding a full-text index to it, or converting parts of the publication into appropriate formats for viewing, printing or downloading. It may involve providing a viewing configuration. It may even involve providing emulation software for displaying the publication. It regulates all the operations of the system and takes care of monitoring, quality control and auditing. It requests status reports from all processing modules and controls, regularly, if DSEP standards and policies set out by the deposit library management are applied throughout the system. Medium migration refreshing or copying a publication is a preservation procedure that takes place in Archival Storage. It should be associated with storage because the stored bits need to be preserved. But archival storage does not have and does not need to have any knowledge of the content of a publication. As formats become obsolete and the viewers needed to interpret and render these formats also become obsolete, it will be necessary to take measures to preserve the content of a publication and all related aspects such as data, layout, structure and functionality. To this end, several strategies may be followed, such as migration and emulation. In the OAIS model, digital migrations that require changes to the content are referred to as transformations. In all cases, transformation leads to a "new version" of the original publication. However, it is not clear where transformation processes take place in OAIS. We have added a dedicated

Preservation module to address this need. The module is configured according to the deposit library preservation policies. Both transformation and emulation approaches are worked out in some detail in the DSEP model. The resulting output is either a new version of a formerly deposited publication, in which case it is ingested anew in the system, or it is a set of specifications for building emulators that can render a whole generation of publications on a future unknown platform. In both cases, new preservation metadata will be generated and fed into Data-Management. Such a package contains the following: The original bit stream of the digital publication Metadata This may be primary metadata as provided by the publisher title information, system requirements information, etc. Software This is the application software required to "render" the publication viewer, browser, search and retrieval software, etc. Packaging information This is data about the package being exchanged, such as package label, identifier, structure of content, etc. However, it should be noted that the OAIS information objects are logical objects. This is done because metadata updates will be more or less frequent, whilst the data bit stream of the publication content will not change over time. It is therefore not deemed sensible to store both types of data together in one physical container. All logical data entities that belong together and pertain to the same publication need to be linked together via interoperable identifier systems identifier of the publication, of the information package, of the metadata records, etc. Metadata for preservation The OAIS concepts of "Representation information" and "Preservation description information" allow for the correct interpretation of the data bit stream over an indefinite period of time. In a DSEP environment, the "Representation information" includes all technical characteristics of a publication, in particular: The format s of a publication, referring to the way in which the data is encoded file formats, character encoding, etc. The application software accompanying the publication, referring to the software that is required to "render" the publication viewer, browser, search and retrieval software, etc. The system requirements, referring to the hardware and systems software configurations that can run the publication. Examples of such system requirements are: In DSEP, the "Preservation description information" includes all recorded metadata giving information about the authenticity of a deposit copy and the preservation measures taken by the DSEP. Depending on the preservation strategy followed, both types of information need to change over time. Management of change and versioning of AIPs are central to the migration strategy: Assessing the digital original and authenticity control are central to the emulation strategy: The parts that need to be emulated need to be specified in detail metadata in a high-level language and the user needs to be educated to "use" the digital original -- as future generations will not know how to interact with obsolete IT-based end user environments. It is, however, clear that there is not one ideal strategy. Many aspects are involved, such as deposit conditions agreed upon with the publishers, cost aspects, future user access requirements, legal constraints, etc. NEDLIB partners are in agreement that we need more practical, hands-on, experience with different preservation approaches, in order to be able to evaluate their adequacy for deposit libraries. Experimenting with emulation for preservation In May , the Koninklijke Bibliotheek and Jeff Rothenberg agreed upon a project proposal to perform emulation experiments for long-term preservation purposes. The overall purpose of the project is to test the viability of using hardware emulation as a means of preserving digital publications in a deposit library. The experiment will be designed to test and evaluate the hypothesis of Jeff Rothenberg, as publicised in in the Scientific American [ref. For the application area of deposit libraries, Rothenberg has formulated his hypothesis as follows: The first stage of the experiment, carried out during , performs a "base-case" iteration of the experiment. This entails developing an initial experimental environment, with a selection of materials, a set of preservation criteria and well-defined procedures for testing and validation, and consequently performing "null" and off-the shelf emulation experiments.

Chapter 3 : Metadata for the long term preservation of electronic publications | Search Results | IUCAT

Metadata for the long term preservation of electronic publications. Request This. Digital preservation. Electronic publications--Conservation and restoration.

Introduction It is now common knowledge that digital information is fragile in ways that differ from traditional technologies, such as paper or microfilm. The fact that information is increasingly stored in digital form, has led to an accelerated search for effective methods of managing electronic information resources. The huge and ever expanding multiple sources of information on the Web normally contain special formatting and are produced with a variety of software in different versions. If the original digital resource is not "born digital", it may be a digital representation or digital surrogate of the physical medium, e. The persistence of digital information resources is an important factor for any digital library development. Addressing the preservation and long-term access issues for digital resources is one of the key challenges facing libraries and information centers today. In order to make sense of the high heterogeneity that exists among digital resources, a growing body of research has attempted to deal with the problems associated with the volume and nature of information on the Web and to look into ways to achieve consensus on a standard. Metadata for Digital Resources Management Metadata is a set of attributes used to describe an object. In reviewing the library and information science literature of the past few years, there is no shortage of views of the significant role of metadata in meeting the most pressing needs and challenges of digital resource management. A number of researchers Moen, ; Waibel, ; Besser, ; Sutton, ; Zeng, ; among others , agreed that the underlying principle for metadata is to link and integrate heterogeneous, multi-platform, massive digital information collections that are contributed by different institutions into a single unified resource so these digital repositories are accessible by anyone, from anyplace, at anytime. A number of metadata initiatives provide detailed and descriptive information about a digital resource to facilitate discovery by users. Resource description is essentially about describing information resources using a standard framework or set of principles. But because of the specific nature of heterogeneous digital resources, describing digital resources in a consistence fashion may not be an easy task, and in some cases, it is a complex process. Those concerned with digital information management all regard metadata as an essential component of the evolving networked information environment, but each of these communities view metadata with notably different perspectives. Current Metadata Initiatives Metadata standards come from various professional community efforts to support many needs in the digital environment. The literature reveals that different communities view metadata in significantly different contexts. The recent report from Research Libraries Group RLG , comprised of key stakeholders from a variety of institutions, affirmed the fact that no single metadata standard can be expected to accommodate the needs of all communities. Although some projects, such as Dublin Core DC have tried to develop a coherent set of metadata schemes that can work for wide range of communities, they have not yet provided a complete description or solution for all types of digital information resources. There is a great diversity of perspectives on various aspects of metadata issues. However, what worked well for libraries may not work in other environments. Similarly, the basic metadata required for describing an image or work of art or non-text objects will bear a strong resemblance to the metadata that describes traditional print documents. However, some significantly different extra elements will be required for a complete description of non-text images and multi-media resources. In light of this, some formats of metadata have been developed specifically for use in certain fields of study or type of information source. Different communities have developed their own organizational and descriptive standards for accessing, arranging, and administering their specific digital collections, such as: A number of commentators e. Moen, ; Besser, ; and, Sutton, are optimistic that the core element set will be as minimal as possible. Thus, the core element set meanings will be easy to understand by most users and the element set will be flexible enough for description of diversified resources in a wide range of subject areas. Of course, the various previous efforts provide ways of describing digital resources to

facilitate interoperability among resource discovery tools. For instance, the most recent work at the Library of Congress LC - Metadata Encoding and Transmission Standard METS schema - provides a flexible mechanism for encoding descriptive, administrative, and structural metadata for a digital library object, and for expressing the complex links between these various forms of metadata. As indicated by Mullen , most metadata initiatives have focused on resource discovery and to make it easier for people to find all of the information they need. Although such standards and structures are the most important steps in the development of the Web to avoid a chaotic repository of information, they do not guarantee continual long-term access to digital resources. Preservation Metadata For years, information centers preserved important electronic resources by transferring the files at regular intervals to the latest new information carriers available. As described by Besser , refreshing a file involves periodically moving a file from one physical storage medium to another to avoid the physical decay or the obsolescence of that medium. Similarly, refreshing of files involves periodically moving files from one file encoding format to another that is usable in the current computing environment. But with multi-media digital resources unlike in print restoring in digital format may not be possible without the original software or hardware. Preserving digital resources is made difficult by the fact that digital resources can only be read by software. This would mean that in order to ensure long-term access to digital resources, we need to preserve all the software, hardware, and operating systems on which the software ran. However, with the current quick obsolescence of information technologies, such an approach may not be feasible. Furthermore, inadequate media longevity is one of the issues. For instance, optical disks are expected to have a physical lifetime of up to 30 years but even a life expectancy of 30 years for storage media far exceeds the lifespan of hardware and software. Considering the ever-growing global Internet traffic, another problem is the mass of data and the need to compress it for efficient storage and transmission. However, compression sometime causes loss of data. It is also likely that repeated transfers over years from one carrier to another may cause data loss. This raises a number of issues including copyright, authenticity, and reliability. Evidently, sustainable solutions to preserve digital resources are not yet available and are still being tested by various communities. As described by Chapman , if the objectives of digital preservation strategies were to preserve the artifact only, regardless of usability, longevity would be measured according to the lifespan of an object stored in a given environment. A number of researchers defined digital preservation in a variety of ways and present their views on how digital preservation might be achieved. According to the RLG report, the problem of preserving digital sources is compounded by the fact that most of the sources do not have proper descriptions. Similarly, Besser stated that such multi-format resources created by differing software require detail descriptions of the technical environment needed to view the digital resources. Despite the fact that most metadata research gives more emphasis on resource discovery, a small breakthrough has been achieved in the last couple of years for preservation issues. A growing number of efforts to perfect the digital preservation methods by various organizations and agencies include: These high-level preservation metadata initiatives provide much needed information required to manage the long-term preservation of digital resources. As indicated by Besser , preservation metadata is a strategy to provide sufficient technical information about the resources and to support the two primary strategies for preservation of digital resources, migration transfer of digital resources from one generation to a subsequent generation and emulation developing techniques for imitating obsolete systems on future generations of computer. Besser asserts that properly used metadata facilitate the long-term access of the digital resources by explaining the technical environment needed to view the work, including applications and version numbers needed, decompression schemes, other files that need to be linked to it, among others. As the name Cybercemetery indicates, the UNT Libraries collect the digital publications from "deceased" federal agencies and preserve them for current and future public access. Furthermore, various digitization undertakings at the Government Documents Department made available selected local resources such as the Texas Criminal Justice Statistical Reports. Cybercemetery Digital Project Page: For instance, Texas Register is one of the UNT libraries popular electronic resources that became available online in Since then, the user base has grown dramatically, with the current number of hits more

than sixteen times the hits in the early months. Usage grew from just over 30, in January, to over , in April Each digital information collection attracted a lot of users from all over the world. Preservation Metadata Requirement Analysis Metadata is a key factor for ensuring the long-term access of digital resources. There is a continuous need for extending the existing metadata element set to be able to describe all available digital resources. In addressing the issues of identifying specific metadata requirements, UNT Libraries attempted to assess the specific characteristics of the existing digital resources. In the preliminary needs assessment, the following issues, among others, were considered: Specific creation features and production life cycle of the digital information resources: Structural Type Text, image, Audio, Video etc. Documentation has always played a key role in preservation practice and there are many instances where documentation provided the only information about processes and changes that had been applied and might need to be corrected. In this regard, all available digital resources creation manuals, guidelines, and reports at the Government Documents Department were reviewed and modified accordingly. Sample documents, Texas Register digital collection creation processes report and ACIR procedure manual can be viewed at http: Those documents provide detailed information about the creation history and complete life cycle of the digital resources. The preliminary resource assessment and evaluations assisted us in identifying the specific characteristics and requirements of the available digital resources. Based on the thorough assessment of the available digital resources, attempts have been made to review current best practices and standards to represent a range of relevant fields. The review pays particular attention to the preservation and management metadata sets, which are needed to support various preservation approaches including migration and emulation. The work at NLA developed a practical model for dealing with the immediate threat of disappearing digital objects, and established a workable distributed archive. The Draft Metadata Architecture The extensive literature review revealed that effective metadata is our best way of minimizing the risk of digital resources becoming inaccessible. Metadata, to be most valuable, both for the users and owners, needs to be consistently maintained throughout the process. Creating documentation that governs and informs the metadata creation steps and procedures in a consistent and uniform manner is among the most important steps in metadata creation. The detailed workflow and user guide document provides procedural information required to create metadata with examples for different file formats. The following chart Figure 3 illustrates the basic structure of the UNT Libraries draft preservation metadata contents. A detailed description of the recommended preservation metadata elements can be found in Appendix II of this paper. The following table Table 1 describes the subheadings of each metadata elements.

Chapter 4 : Digital preservation and long-term access to the content of electronic serials

NEDLIB - LB D Metadata for long term preservation Issue: 29/07/00 5 2. THE OAIS DATA MODEL The NEDLIB Deposit System for Electronic Publications (DSEP) is based on the OAIS information model.

Appraisal[edit] Archival appraisal or, alternatively, selection [6] refers to the process of identifying records and other materials to be preserved by determining their permanent value. Several factors are usually considered when making this decision. Appraisal is identified as A4. Archival appraisal may be performed once or at the various stages of acquisition and processing. Macro appraisal, [10] a functional analysis of records at a high level, may be performed even before the records have been acquired to determine which records to acquire. More detailed, iterative appraisal may be performed while the records are being processed. Appraisal is performed on all archival materials, not just digital. It has been proposed that, in the digital context, it might be desirable to retain more records than have traditionally been retained after appraisal of analog records, primarily due to a combination of the declining cost of storage and the availability of sophisticated discovery tools which will allow researchers to find value in records of low information density. However, the selection, appraisal, and prioritization of materials must be carefully considered in relation to the ability of an organization to responsibly manage the totality of these materials. Often libraries, and to a lesser extent, archives, are offered the same materials in several different digital or analog formats. They prefer to select the format that they feel has the greatest potential for long-term preservation of the content. The Library of Congress has created a set of recommended formats for long-term preservation. Identification identifiers and descriptive metadata [edit] In digital preservation and collection management, discovery and identification of objects is aided by the use of assigned identifiers and accurate descriptive metadata. An identifier is a unique label that is used to reference an object or record, usually manifested as a number or string of numbers and letters. As a crucial element of metadata to be included in a database record or inventory, it is used in tandem with other descriptive metadata to differentiate objects and their various instantiations. Another common type of file identification is the filename. Implementing a file naming protocol is essential to maintaining consistency and efficient discovery and retrieval of objects in a collection, and is especially applicable during digitization of analog media. Using a file naming convention, such as the 8. Integrity[edit] The cornerstone of digital preservation, " data integrity " refers to the assurance that the data is "complete and unaltered in all essential respects"; a program designed to maintain integrity aims to "ensure data is recorded exactly as intended, and upon later retrieval, ensure the data is the same as it was when it was originally recorded". However, digital preservation efforts may necessitate modifications to content or metadata through responsibly-developed procedures and by well-documented policies. Data integrity practices also apply to modified versions, as their state of capture must be maintained and resistant to unintentional modifications. Fixity[edit] File fixity is the property of a digital file being fixed, or unchanged. File fixity checking is the process of validating that a file has not changed or been altered from a previous state. While checksums are the primary mechanism for monitoring fixity at the individual file level, an important additional consideration for monitoring fixity is file attendance. Whereas checksums identify if a file has changed, file attendance identifies if a file in a designated collection is newly created, deleted, or moved. Tracking and reporting on file attendance is a fundamental component of digital collection management and fixity. Characterization[edit] Characterization of digital materials is the identification and description of what a file is and of its defining technical characteristics [18] often captured by technical metadata, which records its technical attributes like creation or production environment. Digital sustainability concentrates less on the solution and technology and more on building an infrastructure and approach that is flexible with an emphasis on interoperability , continued maintenance and continuous development. External dependencies can refer to hardware, software, or physical carriers. Format obsolescence[edit] File format obsolescence can occur when adoption of new encoding formats supersedes use of existing formats, or when associated presentation tools

are no longer readily available. Significant properties[edit] Significant properties refer to the "essential attributes of a digital object which affect its appearance, behavior, quality and usability" and which "must be preserved over time for the digital object to remain accessible and meaningful. It assists appraisal and selection, processes in which choices are made about which significant properties of digital objects are worth preserving; it helps the development of preservation metadata, the assessment of different preservation strategies and informs future work on developing common standards across the preservation community. Authenticity has been defined as ". The content and meaning of that inaccurate record will remain unchanged. Access[edit] Digital preservation efforts are largely to enable decision-making in the future. Should an archive or library choose a particular strategy to enact, the content and associated metadata must persist to allow for actions to be taken or not taken at the discretion of the controlling party. Preservation metadata[edit] Preservation metadata is a key component of digital preservation, and includes information that documents the preservation process. It supports collection management practices and allows organizations or individuals to understand the chain of custody. Intellectual foundations[edit] Preserving Digital Information [edit] The challenges of long-term preservation of digital information have been recognized by the archival community for years. The final report published by the Task Force Garrett, J. Report of the task force on archiving of digital information. The concepts and recommendations outlined in the report laid a foundation for subsequent research and digital preservation initiatives. In , they published "Trusted Digital Repositories: The Trusted Digital Repository Model outlines relationships among these attributes. The report also recommended the collaborative development of digital repository certifications, models for cooperative networks, and sharing of research and information on digital preservation with regard to intellectual property rights. Gladney proposed another approach to digital object preservation that called for the creation of "Trustworthy Digital Objects" TDOs. TDOs are digital objects that can speak to their own authenticity since they incorporate a record maintaining their use and change history, which allows the future users to verify that the contents of the object are valid. The research is being conducted by focus groups from various institutions in North America, Europe, Asia, and Australia, with an objective of developing theories and methodologies that provide the basis for strategies, standards, policies, and procedures necessary to ensure the trustworthiness, reliability, and accuracy of digital records over time. Its goal is to utilize theoretical and methodological knowledge generated by InterPARES and other preservation research projects for developing guidelines, action plans, and training programs on long-term preservation of authentic records for small and medium-sized archival organizations. Now a large quantity of information exists in digital forms, including emails, blogs, social networking websites, national elections websites, web photo albums, and sites which change their content over time. These environments keep evolving and changing at a rapid pace, threatening the continuity of access to the content. In the case of born-digital content e. Rapidly changing technologies can hinder digital preservationists work and techniques due to outdated and antiquated machines or technology. This has become a common problem and one that is a constant worry for a digital archivistâ€”how to prepare for the future. Digital content can also present challenges to preservation because of its complex and dynamic nature, e. Without the source code an adaption Porting on modern computing hardware or operating system is most often impossible, therefore the original hardware and software context needs to be emulated. Another potential challenge for software preservation can be the copyright which prohibits often the bypassing of copy protection mechanisms Digital Millennium Copyright Act in case software has become an orphaned work Abandonware. An exemption from the United States Digital Millennium Copyright Act to permit to bypass copy protection was approved in for a period of 3 years to the Internet Archive who created an archive of "vintage software", as a way to preserve them. The amount of digital information being created along with the "proliferation of format types" [2] makes creating trusted digital repositories with adequate and sustainable resources a challenge. The Web is only one example of what might be considered the "data deluge". Preservation programs require significant up front investment to create, along with ongoing costs for data ingest, data management, data storage, and staffing. One of the key strategic challenges to such programs is the fact that, while they require

significant current and ongoing funding, their benefits accrue largely to future generations. Assessing the risks for loss of content posed by technology variables such as commonly used proprietary file formats and software applications. Evaluating the digital content objects to determine what type and degree of format conversion or other preservation actions should be applied. Determining the appropriate metadata needed for each object type and how it is associated with the objects. Providing access to the content. Refreshing[edit] Refreshing is the transfer of data between two types of the same storage medium so there are no bitrot changes or alteration of data. This strategy may need to be combined with migration when the software or hardware required to read the data is no longer available or is unable to understand the format of the data. Refreshing will likely always be necessary due to the deterioration of physical media. Migration[edit] Migration is the transferring of data to newer system environments Garrett et al. This may include conversion of resources from one file format to another e. Two significant problems face migration as a plausible method of digital preservation in the long terms. Due to the fact that digital objects are subject to a state of near continuous change, migration may cause problems in relation to authenticity and migration has proven to be time-consuming and expensive for "large collections of heterogeneous objects, which would need constant monitoring and intervention. CDs, USB flash drives, and 3. These types of devices are generally not recommended for long-term use, and the data can become inaccessible due to media and hardware obsolescence or degradation. Data that exists as a single copy in only one location is highly vulnerable to software or hardware failure, intentional or accidental alteration, and environmental catastrophes like fire, flooding, etc. Digital data is more likely to survive if it is replicated in several locations. Replicated data may introduce difficulties in refreshing, migration, versioning, and access control since the data is located in multiple places. Understanding digital preservation means comprehending how digital information is produced and reproduced. Because digital information e. Emulation[edit] Emulation is the replicating of functionality of an obsolete system. According to van der Hoeven, "Emulation does not focus on the digital object, but on the hard- and software environment in which the object is rendered. It aims at re creating the environment in which the digital object was originally created. Emulators may be built for applications, operating systems, or hardware platforms. Emulation has been a popular strategy for retaining the functionality of old video game systems, such as with the MAME project. The feasibility of emulation as a catch-all solution has been debated in the academic community. Granger, Raymond A. Lorie has suggested a Universal Virtual Computer UVC could be used to run any software in the future on a yet unknown platform. The UVC strategy has not yet been widely adopted by the digital preservation community. Jeff Rothenberg, a major proponent of Emulation for digital preservation in libraries, working in partnership with Koninklijke Bibliotheek and National Archief of the Netherlands, developed a software program called Dioscuri, a modular emulator that succeeds in running MS-DOS, WordPerfect 5. Rushdie donated an outdated computer to the Emory University library, which was so old that the library was unable to extract papers from the harddrive. In order to procure the papers, the library emulated the old software system and was able to take the papers off his old computer. Metadata attachment[edit] Metadata is data on a digital file that includes information on creation, access rights, restrictions, preservation history, and rights management. ASCII is considered to be the most durable format for metadata [71] because it is widespread, backwards compatible when used with Unicode , and utilizes human-readable characters, not numeric codes. It retains information, but not the structure information it is presented in. Preservation repository assessment and certification[edit] A few of the major frameworks for digital preservation repository assessment and certification are described below. A more detailed list is maintained by the U. Center for Research Libraries. TRAC is based upon existing standards and best practices for trustworthy digital repositories and incorporates a set of 84 audit and certification criteria arranged in three sections:

Chapter 5 : Electronic Records Management Guidelines, Metadata

Archiving of electronic publications and providing long-term access to this information, in other words offering digital continuity, is one of the major challenges facing national libraries and.

This is a preprint version of an article published in: Haworth Information Press, , pp. Please refer to the print version in any citation. Abstract This paper is an investigation of digital preservation issues that are related to the development and use of scholarly online electronic serials e-serials. Firstly, some background topics are discussed. Digital preservation issues are then discussed in more detail, including an examination of the technical difficulties of preserving information in digital form. The paper then considers some preservation strategies that have been proposed, including technology emulation and data migration and the role of metadata. There follows a brief introduction to some important non-technical issues. Introduction The long-term preservation of information in digital form has been a subject of growing interest for the library and information professions since the late s. Initially, this was a response to the growing use of digital technologies in publishing and the suggestion by Lancaster and others that society was steadily moving towards the use of information systems that would be largely paperless [1]. This report provided a good summary of relevant issues and acted as a catalyst for further research and development. Naturally, digital preservation issues apply across the whole range of digitally based information resources that exist. This paper will be concerned with just one subset of these; i. In order to reflect current trends, this will include both the contents of those electronic serials that have been formally published by commercial and learned society publishers and also those papers or e-prints that have been "self-archived. Initially, however, it is important to define some terms. Even the concept of "digital preservation" can be difficult to define unambiguously. For example, the term is sometimes confused with the quite separate idea of digitising resources as part of a preservation strategy for non-digital objects. In the context of scholarly communication, back issues of serials are often digitised and made available over networks in order to improve access to the information contained in them and, on occasion, to aid the preservation of the original item [3]. Although it is acknowledged that digitised versions of the back issues of serials will also need to be preserved, they will not form the main focus of this paper. The concept of digital preservation can also get muddled with ideas of digital archiving or digital archives. Indeed, the terms sometimes appear to be used interchangeably. This can be a source of confusion. For example, members of the archives and records professions will have a very different understanding of the archives concept than computer scientists, who often use the word "archive" as a verb, meaning the creation of secure backup copies for a fixed period of time. The word is used in a similar way by the developers of e-print archives. Wherever possible, this paper will use the word "preservation" in an attempt to avoid any confusion, except in relation to e-print archives. Preservation itself is primarily concerned with the survival of information in a usable form for as long as it is required [5]. Preservation, therefore, is not just concerned with the conservation or restoration of physical artefacts, but includes all of the strategic and organisational considerations that relates to the survival of information over time. Hedstrom has usefully defined digital preservation as "the planning, resource allocation, and application of preservation methods and technologies to ensure that digital information of continuing value remains accessible and usable" [6]. A distinction is sometimes made between the preservation of the information embodied in a document the information content and the physical conservation of an information carrier [7]. This is especially relevant for digital information - including the content of e-serials - because most users tend not to be interested in which particular type of physical object is being used for storing the information. There may be a case for retaining physical objects, for example in a museum of technology, but for e-serials, the preservation of informational content is the main issue. Another important, often emotive, issue is how long this information should be kept. It is often assumed that preservation should be permanent, often defined with reference to loaded terms like "in perpetuity" and "indefinitely. In consequence, this paper will assume that preservation is normally linked to continuing access

and use and not with nebulous concepts of permanence. The electronic serial landscape Scholarly and scientific papers published in peer-reviewed printed serials have fulfilled a vitally important role in the scholarly communication process since the seventeenth century. Line has correctly described them as the "established medium of record and dissemination" [9]. Despite this, there has been a growing feeling that the traditional printed serial has outlived its historical role and that some kind of digitally based form of scholarly communication will soon replace it [10]. At the moment, there are two main models for the digital distribution of scholarly communication through the Internet, and both have developed relatively independently [11]. Firstly, the publishers of traditional printed serials have begun to place digital copies of each issue on Web sites a form of parallel publishing and have started to experiment with the production of digital-only journals. Secondly, some scholars and librarians have outlined "subversive proposals" that aim to ensure free access to the scholarly and scientific literature through the self-archiving of papers by their authors. Both the publishers of e-serials and the supporters of author self-archiving initiatives hope - in different ways - that developments in the digital distribution of scholarly communication might help offset the effects of the ongoing "serials crisis" [12]. The functions of serials The idea of electronic serials has been around for a number of years. Research projects in the s first proved that e-serials were technically feasible [13]. However, these projects did not develop into sustainable services at the time because computer networks were immature and not ubiquitously available, even within the research community [15]. Also, in some cases, the proposed user-interfaces were poor and the system developers did not always have a clear idea of the many different functions fulfilled by the printed scholarly serial. McKnight has noted that there is no incentive for either authors or readers to change unless electronic serials can do "at least the same things - and preferably more - with electronic journals as they do with paper" [16]. The printed peer-reviewed serial has proved to be an extremely successful part of scholarly communication since the seventeenth century because it has fulfilled a wide range of different functions. The following list is expanded from those functions that have been identified by Rowland [17]. Dissemination - publication of a paper in a peer-reviewed serial allows an author to disseminate important research findings to the wider research community and beyond. It is important to recognise, however, that the content of many published papers may also have been discussed informally, reported on at conferences or distributed as pre-prints. Quality control - consistently applied editorial processes can help to ensure a high written standard of papers, but the main quality control process is the peer-review of all submitted papers. Establishing priority - one of the most important functions of the printed serial, especially in the science, technology and medicine STM disciplines, is to be able to establish priority over a particular discovery or advance. The recognition of authors - authors value publication in refereed-serials as a means of raising their profile and as a means of gaining further research contracts or promotion. The creation of a public domain "archive" - once published, serial papers are in the public domain and research libraries collectively act as a distributed "archive," preserving the knowledge embodied in them for future scholars. These are the basic functions that will need to be fulfilled by any new form of digitally based scholarly communication. Electronic serials Since the late s, both commercial and not-for-profit publishers e. Another example of library-publisher co-operation was TULIP The University Licensing Program , in which selected serials from the Elsevier Science group were delivered to participating US research libraries in an attempt to investigate some of the technical, legal and economic issues associated with e-serials and user behaviour [19]. It was, however, the emergence of the Internet as a mass medium that finally persuaded both commercial and learned society publishers that the development of electronic serials was a viable option. Publishers have started to distribute the content of their scholarly serials through the Internet. At the present time, probably for economic reasons, most publishers are using digital technologies in an essentially conservative way [20]. Most are creating Web-based services that give access to the content of already published printed serials rather than introducing new titles that implement digital-specific features like embedded multimedia. Many publishers "bundle" collections of e-serials together and offer institutions and their libraries the ability to subscribe to their entire list as a single product [21]. Halliday and Oppenheim note

that this helps publishers to spread production costs across the whole subscriber base and can help reduce the scope of license negotiations [22]. At the same time, the organisations that subscribe to serials are also beginning to join together in order to collectively negotiate licenses with publishers. NESLI, a consortium led by the University of Manchester and Swets Blackwell, negotiates deals with publishers on behalf of the whole UK higher education community, and where possible, incorporates a clause for ensuring long-term access [23]. Some research libraries are beginning to support partnerships with learned society publishers in order to provide digital outlets for existing titles and help foster the publication of new electronic serials. For example, HighWire Press - a not-for-profit initiative of the Stanford University Libraries - gives access to a large number of serials published by learned societies. This organisation looks for partnerships with learned society publishers and seeks to underwrite new serials that might be able to provide competition for more established and expensive titles. It is important to realise that most commercial and learned society publishers have not yet moved to digital-only publication. Many tend to offer access to parallel digital versions of printed articles - usually in PDF Portable Document Format - and some bundle both printed and digital formats into the subscription price. On occasion, however, the digital version will offer more functionality than the printed version, e. The Web sites of some serials also become a focus of more informal communication and sometimes provide space for news items, scholarly debates, information about relevant events, links to external Internet resources, etc. The Web sites of general scientific serials like Nature and Science are good examples of this. It is important to recognise that the Web sites of serials themselves may also be good candidates for long-term preservation. Self-archiving initiatives At the same time as this mainly publisher-led activity has been unfolding, other actors in the scholarly communication chain have seen the potential for more wide ranging changes in scholarly communication, often suggesting that printed serials have no long-term future [25]. Some scholars and librarians have begun to ask why the status quo in paper serials should simply duplicate itself in the Internet age [26]. People have begun to support "subversive proposals" that suggest that the authors of scholarly and scientific papers should simply make them available for free by storing or archiving copies on Internet sites [27]. Proponents of the self-archiving idea argue from the premise that peer-reviewed serials form a peculiar type of publishing that has little in common with the payment or royalty based trade-publishing sector. So, for example, Harnad and Hemus argue that the authors of scholarly and scientific papers are not primarily interested in monetary reward, but in having their work read, used, built-upon and cited [28]. In the print world, the authors of papers had to perpetuate what Harnad has called a "Faustian bargain," whereby they traded the copyright of works to publishers in exchange for having them printed and distributed. He argues that this type of bargain made sense when publishing remained a exclusive and expensive domain, but that it has little or no relevance in the Internet age when scholars and scientists can self-archive their own papers at little or no personal cost [29]. Authors, then, are being encouraged to deposit digital copies of their papers in centralised e-print services or in e-print archives based at their own institutions, thus making them freely available to fellow scholars and scientists. Furthermore technologies have now been developed that would enable distributed e-print archives implemented in accordance with the standards developed as part of the OAI Open Archives Initiative to be combined into a single global virtual archive [30]. The most frequently cited model of the first stages of the "subversive-proposal" in action is the e-print archive first set up by Ginsparg at the Los Alamos National Laboratory in [31]. This service originally gave access to digital pre-prints in the domain of high-energy physics and it very quickly became the primary means of scholarly communication in this subject area. It has since expanded to cover other areas of physics, mathematics and computer science. A physicist was quoted in as saying that the service had completely changed the way people in his field exchanged information. He noted that the only time that he needed to look at published serials was to find articles that predated the Los Alamos physics databases [32]. The success of the Los Alamos e-print archive has led to the development of other Web-based e-print services. This has not been without controversy. For example, in , the US National Institutes of Health published a proposal for an service called PubMed Central that would give free online access to published material in the biomedical

sciences [33]. The original proposal suggested the creation of two separate services: The non-peer-reviewed section of the proposed service received considerable criticism, largely focussed on the need for the strict evaluation of clinical research, because of potential adverse impacts on public health and medical practice [34]. Consequently, the PubMed Central system, when it was officially launched in January , only contained the peer-reviewed part of the proposed service. To date, many scholars and scientists have been reluctant to commit themselves to self-archiving. Part of this is due to cultural differences between different subject areas. Valauskas, for example, notes how different the styles of communication and verification, debate and consensus can be amongst different academic disciplines [35]. It is rather simplistic to say that because self-archiving appears to be accepted by most physicists then it should also be adopted by biomedical researchers or ancient historians. Another possible reason why self-archiving has failed to take off is that scholars and scientists have been reluctant to stop publishing in established high-impact serials.

Chapter 6 : CiteSeerX â€” Citation Query Metadata for the Long Term Preservation of Electronic Publicatio

The challenges that have to be addressed are preservation of the electronic publications. In the past years national libraries have taken the lead in developing practical solutions for digital preservation.

Implementation Strategies PREMIS The need for metadata systems to support the preservation of digital resources has been addressed by an international working group of information specialists. The TEI metadata schema thus allows digital texts to be searchable, both internally and in relation to other texts. The schema is flexible and highly customizable, and has been widely adopted internationally by all kinds of repositories and other information-managing organizations. The TEI Web site provides complete documentation on the schema, as well as links to implementation tools and examples. Transcripts are encoded with subject index tags, including implementation of authority lists for personal and corporate names. This has included markup with tags not only for subject searching but also for synchronizing transcripts of interviews with audio recordings. Further work is underway to incorporate georeferencing and geotagging GIS metadata to relate oral history interviews to maps and other geographical representations of context. Popular Web sites have for some time facilitated the addition of tags and comments to user-generated content. From this user-created metadata emerge so-called folksonomies, which can be included in metadata schemas that describe, analyze, and relate digital materials. An advantage of such user-created metadata is its intimate relevance to user communities, which should be of special interest to oral history curators. The disadvantages are the lack of quality control, where there is no screening or moderation to ensure relevance and adherence to community standards of appropriateness, and relative inefficiency, as multiple terms distinguishable only in minor variations of spelling or grammatical form usually proliferate and make search and retrievable less precise. It seems likely that the advantages will outweigh the disadvantages, as search tools become increasingly sophisticated. Professionally designed systems for tagging and indexing audio and video recordings of oral history interviews have been developed, using proprietary and open-source software tools. A similar system for segmenting and annotating audio and video recordings has been developed by the Concordia Digital History Lab, a part of the Centre for Oral History and Digital Storytelling at Concordia University in Montreal. Vertov is a free, open source Firefox 2. A section of the Zotero Web site addresses aspects of metadata exposure for harvesting and importing in various bibliographic metadata systems. This obligation to interview participants, their communities, and researchers, truly begins at the inception of every oral history project. By the time an oral history is handed over to a repository for access and preservation, documentation must already exist, in trustworthy form and complete detail. This suggests that there is need in oral history practice not only for best practices in metadata management for curatorial purposes, but also for best practices in oral history project management and record-keeping, as well as systems to support such management. Scalable, flexible, adaptable, and interoperable project management systems for oral history have yet to emerge. When they do, they will provide oral history practitioners with the means to do their work and accomplish their mission with long-term access and preservation in mind from the beginning. These systems will strike a balanceâ€”adjustable for differing levels of resources and supportâ€”between usability and sophistication, and between exhaustiveness and cost. They will also include measures of quality, tools by which their usability, their suitability for particular circumstances, their anticipated longevity and capacity for technological upgrade, and their cost can be assessed. In short, true metadata management for oral history requires a systems view, an overall perspective on oral history practice, from start to finish, with cognizance of the fact that for historical resources such as oral history, there should never be a finish: Rules for Archival Cataloging of Sound Recordings, revised edition. Available for purchase at http: AES standard for audio metadata â€” Audio object structures for preservation and restoration. AES standard for audio metadataâ€”core audio metadata. Gilliland, Maureen Whalen, and Mary S. Online edition, version 3. Getty Research Institute, Metadata and Its Applications in the Digital Library: Left Coast Press, Available online at

http: Oral History Cataloging Manual. Society of American Archivists, Metadata Encoding and Transmission Standard: Primer and Reference Manual. Digital Library Federation, Metadata for Digital Collections: Haworth Information Press, Introduction to Cataloging and Classification. An oral history interview is a trusted body of information, in this sense, if its associated objectsâ€"audio recording on cassette tape, bound transcript volume, audio WAV file on a preservation server, MP3 audio accessible on a Web site, to name a fewâ€"are verified as representations of the interview with the specified participants, conducted on the specified date, in the specified place, and for the specified purpose.

Chapter 7 : Metadata â€“ Oral History in the Digital Age

, Metadata for the long term preservation of electronic publications / Catherine Lupovici, Julien MasanÃ¨s Koninklijke Bibliotheek The Hague Wikipedia Citation Please see Wikipedia's template documentation for further citation fields that may be required.

Chapter 8 : Electronic Records Management Guidelines, Long-Term Preservation

BibTeX @MISC{Lupovici00titlemetadata, author = {Catherine Lupovici and Julien MasanÃ¨s and Catherine Lupovici and Julien MasanÃ¨s}, title = {Title Metadata for the Long Term Preservation of Electronic Publications}, year = {}}.

Chapter 9 : CiteSeerX â€" Title Metadata for the Long Term Preservation of Electronic Publications

Recent developments have suggested that, regardless of which particular digital preservation strategy is chosen, the successful long-term preservation of digital resources will depend upon the creation and maintenance of metadata that will be able to record some part of a digital object's functionality and context [71].