Revista Digital de Biblioteconomia e Ciência da Informação
DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646326
ARTIGO

IMPLEMENTING DIGITAL PRESERVATION IN REPOSITORIES: KNOWLEDGE AND PRACTICES

IMPLEMENTAÇÃO DA PRESERVAÇÃO DIGITAL EM REPOSITÓRIOS: CONHECIMENTO E PRÁTICAS

IMPLEMENTACIÓN DE LA PRESERVACIÓN DIGITAL EN REPOSITORIOS CONOCIMIENTOS Y PRÁCTICAS

Caterina Groposo Pavão[1]
Sônia Elisa Caregnato[2]
Rafael Port da Rocha[3]

ABSTRACT
Digital preservation has to be undertaken by institutional repositories, which are responsible for the preservation of the scientific output from academic institutions. However, due to the constant evolution of the field, to gain domain knowledge and recognise best practices is a complex task for people responsible for digital preservation in those institutions. Digital preservation research, practices and solutions address specific problems, such as formats, curation, reference models, authenticity, policies and preservation plans, tools, etc., while stakeholders need an integrated, contextualized and applicable overview. This paper focuses on the implementation of digital preservation in repositories, from the perspective of the team responsible for the project, regarding the necessary knowledge and best practices. Initially, it defines and contextualizes digital preservation repositories. The following section presents a conceptual model of digital preservation, synthesized from conceptual models developed in influential projects in the field, which allows us to identify the domain knowledge in digital preservation. Finally, aspects represented in the model are discussed in the light of the performance of teams implementing digital preservation repositories. It provides recommendations, guides and examples that may be useful for the implementation of digital preservation. It points to the need to strengthen the relationship between domain knowledge in digital preservation repositories with practices developed in numerous projects developed worldwide.

KEYWORDS: Digital preservation. Digital repositories. Digital preservation practices.

RESUMO
A preservação digital é uma necessidade real a ser atendida pelos repositórios institucionais, responsáveis pela preservação da produção científica de instituições de ensino e pesquisa. Entretanto, devido à constante evolução da área, a busca pelo domínio do conhecimento e das melhores práticas por parte dos responsáveis pela preservação digital é uma tarefa complexa. As pesquisas, práticas e soluções em preservação digital abordam problemas pontuais, como formatos, curadoria, modelos de referência, autenticidade, políticas e planos de preservação, ferramentas, etc., enquanto que os implementadores necessitam de uma visão integrada, geral, contextualizada e aplicável. Este artigo enfoca a implantação da preservação digital em repositórios, sob a perspectiva da equipe responsável pelo projeto, no que diz respeito aos conhecimentos e práticas necessárias. Inicialmente define e contextualiza a preservação digital em repositórios. A seguir, apresenta um modelo conceitual de preservação digital, sintetizado a partir de modelos conceituais e de fundamentos que foram desenvolvidos em projetos influentes na área, o qual permite identificar, de forma geral, os principais domínios

[1] Doutora e mestre em Comunicação e Informação (PPGCOM/UFRGS). Bibliotecária de sistemas do Centro de Processamento de Dados da UFRGS. Porto Alegre, RS. Email: caterina@cpd.ufrgs.br. ORCID: http://orcid.org/0000-0003-3712-7200.
[2] Doutorado em Estudos da Informação pela University of Sheffield. Professora do Ensino Superior (UFRGS). Porto Alegre, RS. Email: sonia.caregnato@ufrgs.br. RCID: http://orcid.org/0000-0002-5676-2763.
[3]Doutor em Computação (UFRGS). Professor Adjunto (UFRGS). Porto Alegre, RS. Email: rafael.rocha@ufrgs.br. ORCID: http://orcid.org/0000-0003-4131-0534.
Submitted: 03/08/2016 - **Accepted**: 15/09/2016.

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646326

de conhecimento em preservação digital. Finalmente, aspectos representados no modelo são discutidos no contexto da atuação das equipes na implantação da preservação digital em repositórios, oferecendo recomendações, guias ou exemplos que podem ser úteis para implantação da preservação digital. Aponta para a necessidade de se fortalecer a relação entre o domínio de conhecimentos em preservação digital em repositórios com as práticas dos inúmeros projetos desenvolvidos globalmente.

PALAVRAS-CHAVE: Preservação digital. Repositórios digitais. Práticas de preservação digital.

RESUMEN

La preservación digital es una necesidad real que deben cumplir los repositorios institucionales, responsable por la preservación de la producción científica de las instituciones de enseñanza e investigación. Sin embargo, debido a la constante evolución del área, la búsqueda por el dominio de conocimiento y las mejores prácticas por parte de los responsables de la preservación digital es una tarea compleja. A investigación, prácticas y soluciones en preservación digital enfoca problemas específicos, tales como formatos, tutela, modelos de referencia, autenticidad, políticas y planes de preservación, herramientas, etc., mientras que los ejecutores necesitan una visión integrada, general, contextualizada y aplicable. Este artículo se centra en la implementación de preservación digital me repositorios, desde la perspectiva del equipo responsable por el proyecto, a lo que respecta al conocimiento y las prácticas necesarias. Inicialmente define y contextualiza los repositorios de preservación digital. A continuación se presenta un modelo conceptual de preservación digital, sintetizado a partir de modelos conceptuales y fundamentos que se han desarrollado en los proyectos influyentes de esta área de estudio, lo que nos permite identificar, en general, los principales dominios de conocimiento en preservación digital. Por último, los aspectos representados en el modelo se discuten en el contexto de la actuación de los equipos en la implementación de preservación digital em repositorios, proporcionando recomendaciones, guías o ejemplos que pueden ser útiles para la aplicación de la preservación digital. Direcciona a la necesidad de fortalecer la relación entre el dominio de conocimiento sobre preservación digital en repositorios con las prácticas de incontables proyectos desarrollados a nivel mundial.

PALABRAS CLAVE: Preservación digital. Repositorios digitales. Prácticas de preservación digital.

1 INTRODUCTION

Digital preservation is crucial for institutional repositories in order to safeguard the scientific production of educational and research institutions. In this context, a central task for teams responsible for digital preservation is to acquire the necessary domain knowledge and implement best practices in an area that is in constant evolution.

Digital preservation is a set of management and technical actions, required to overcome technological changes and fragility of supports, ensuring access and interpretation of digital documents for as long as necessary (CONARQ, 2014, p.7). It consists of actions that guarantee access to digital materials beyond media failures or technological fluctuations. Thus, it should be a concern for information producers and holders, publishers, and heads of large digital collections and information centers (MÁRDERO ARELLANO, 2004).

Digital preservation in a repository can be implemented in the context of digital curation. According to the Digital Curation Centre (DCC)[4], digital curation is focused on actions related to the maintaining, preserving and adding value to research data throughout its

[4] http://www.dcc.ac.uk/digital-curation/what-digital-curation

lifecycle, and involves the active management and preservation of digital resources, for as long as the academic and scientific world find it appropriate. It is an interdisciplinary practice that reflects a holistic approach to managing digital objects, and includes activities for the entire lifecycle of these objects (SIEBRA, et al. 2013).

Digital curation is a broader concept than digital preservation. It involves activities of data management, starting with the planning of its creation, and including good practices in scanning, format selection, documentation, and ensuring that data are always available and suitable to be discovered and reused now and in the future (ABBOTT, 2008). According to Walter and Skinner (2011),

> Digital curation refers to the actions people take to maintain and add value to digital information over its lifecycle, including the processes used when creating digital content. Digital preservation focuses on the "series of managed activities necessary to ensure continued access to digital materials for as long as necessary".

In order to safeguard the authenticity of electronic documents during its whole life cycle, the preservation plan requires a preservation framework that is a combination of technology and organizational procedures. A preservation plan should ensure digital documents to be: discoverable and available for access at the appropriate time; interpretable (considering availability, presentation, representation and visualization); retrievable (including their appropriate metadata); protected against the loss of rights, as intellectual property and confidentiality; available for access at the time that is needed by authorized people; and supervised, regarding to the quality of access (availability, delivery, usage history).

For the implementation of digital preservation in an organization, personnel need an integrated, general, contextualized and applicable view. Research, practices and solutions on digital preservation address specific problems, such as formats, curatorship, reference models, authenticity, policies and preservation plans, tools, etc.

More specifically, higher education institutions should, according Gracio (2012), adopt digital preservation solutions based on planning and well-defined strategies for storage and use of digital objects for as long as necessary, ensuring the continuity of the digital preservation process, and their recovery over time. The author also points out that all digital preservation strategies must be dynamic and regularly reviewed in order to keep up with the constant changes and technological advances.

To overcome this apparent gap, this article addresses the implementation of digital preservation in repositories, focusing on the team responsible to do it and its concerns of necessary knowledge and practices, developed over time. The next section presents a conceptual model of digital preservation. This model have been synthesized from conceptual models and fundamentals that have been developed by influential projects or are present in reference models and metadata standards in the area of digital preservation/curation. In the

following section, some aspects of conceptual model are discussed in the context of the team responsible for the implementation of digital preservation in a repository. The next section provides recommendations, guidelines and examples that could be useful for the implementation of digital preservation

2 CONCEPTUAL MODEL OF DIGITAL PRESERVATION AND CURATION

A common understanding of digital preservation is a key requirement for institutions and professionals who want to implement it. To this end, it is crucial to define clearly concepts used.

This section presents a conceptual model for digital preservation/curation. The model is a compilation of specifications from different initiatives. These are: the digital preservation conceptual model developed by the PLANETS Project (DAPPERT, FARQUHAR, 2009); the PREMIS metadata standard for digital preservation (PREMIS EDITORIAL COMMITTEE, 2015); the digital curation lifecycle model developed by the digital curation Centre (DCC) (HIGGINS, 2008); the fundamentals of digital preservation developed by the University of California Curation Center (U3C) (CALIFORNIA DIGITAL LIBRARY, 2010; ABRAMS, KUNZE, LOY, 2010; ABRAMS, CRUSE, KUNZE, 2009); and the reference model for digital archives and repositories, Open Archival Information System (OAIS) (CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS, 2012).

The conceptual model developed by Dappert and Farquhar (2009) and its terminology are particularly relevant to the model presented in this article. Their model is part of the PLANETS Project, a major undertaking in digital preservation and in digital preservation planning.

Conceptual model is a term employed in this paper according to Mylopoulos (1990):

> Conceptual modelling is the activity of formally describing some aspects of the physical and social world around us for purposes of understanding and communication. Such descriptions, often referred to as conceptual schemata, require the adoption of a formal notation, a conceptual model in our terminology

Figure 1 depicts a conceptual model of digital preservation and digital curation. Digital curation concepts are mainly present in the specifications of services and life cycle, while digital preservation concepts are related to the specifications of the preservation resources, which are used for performing services (such as policies, strategies, plans, object ,record and environment). PLANETS Project and PREMIS metadata schema are the main sources for the specification of digital preservation concepts. The main sources for the specification of the services are: the fundamentals of digital curation from U3C, digital curation lifecycle from DCC, and the OAIS reference model.

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646326

The main notions of the conceptual model are: services, preservation policy, preservation strategy, preservation plan, preservation object, environment, events and agents of preservation (figure 1). The service class is specialized thought subclasses, such as planning, ingestion and storage; and the environment class has sublcasses, as software, hardware, community and producer. In the figure 1, the relationship of a subclass (such as planning) with its superclass (such as services) is represented by black solid lines, with arrow pointing to the superclass.

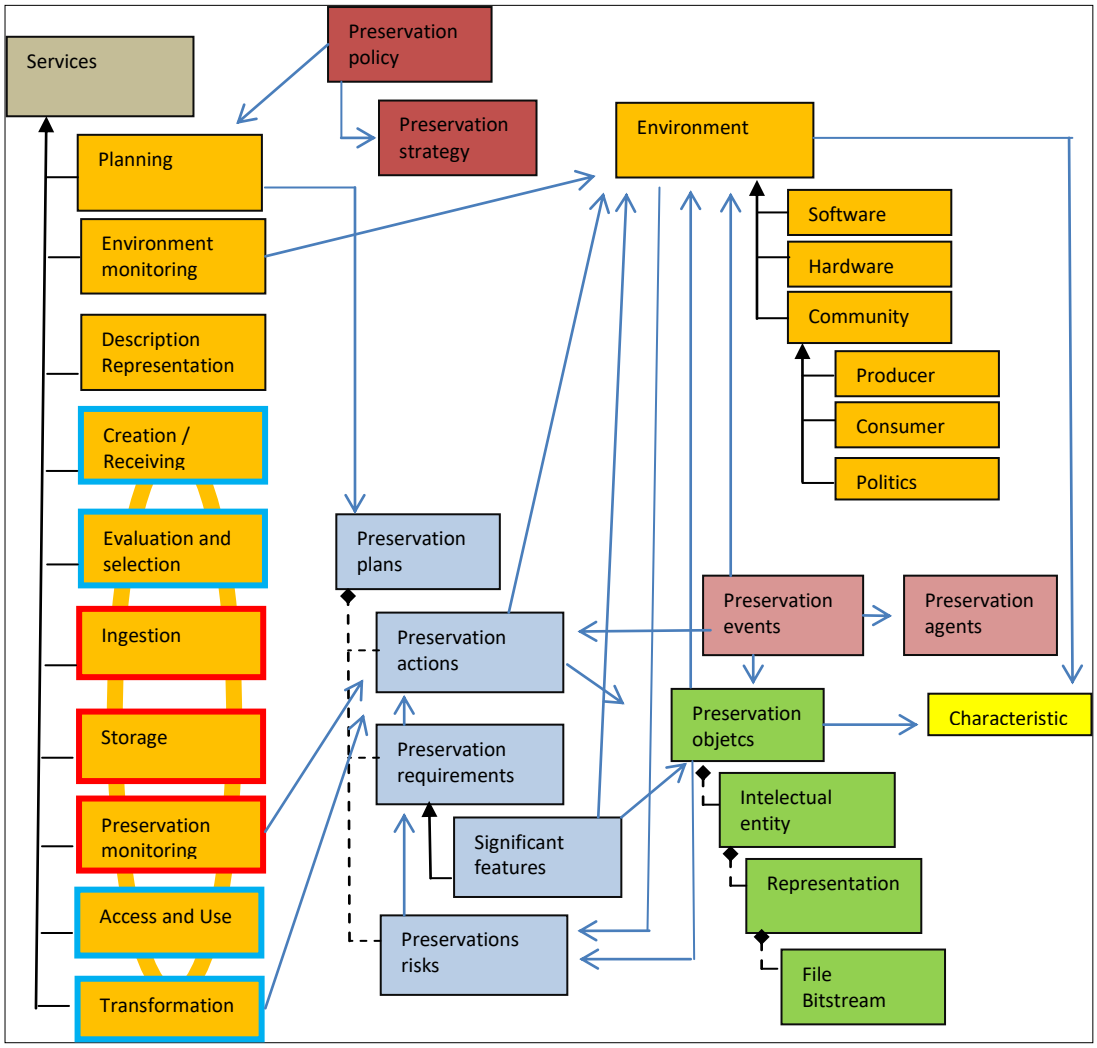

Figure 1. Conceptual model of digital preservation: key concepts and their relationships
Source: Prepared by the authors based on various sources consulted and cited.

In the model, there are also specifications of how plans and preservation of objects are composed. Compositions are represented by dashed lines, and the rhombuses at the end indicate the whole (Figure 1). For example, a preservation plan (whole) consists of preservation action, preservation requirements and preservation risks (parts). Blue arrows represent cause/effect relationships between concepts, indicating, for example, that a planning service has, as effect, a preservation plan, and that preservation policies specify preservation strategies. The digital curation life cycle is based on DDC life cycle, and the

cycle is graphically presented (Figure 1) as an ellipse that connects services, indicating that these services run sequentially. Services which are typically related to digital curation are represented by blue borders, and services more related to digital preservation are depicted by red border. The following text details the concepts. Each concept is highlighted in bold.

Digital preservation policy defines abstract, high-level policy concerns, providing a framework for concrete planning (BECKER et al, 2009). A policy is a formal statement of direction or guidance as to how an organization will carry out its preservation mandate, functions or activities, motivated by determined interests or programs (DAPPERT et al , 2009).

A preservation plan is viewed in more specific and concrete level than a preservation policy. It involves specifying preservation actions to be taken by a responsible institution due to an identified risk for a given set of digital objects or records (called collection) (BECKER et al, 2009). A preservation plan should contain the following: identification, state and triggers to preservation actions, instructional description, description of the collection, preservation requirements, evidence of decisions for a preservation strategy, costs, roles and responsibilities, and action plan preservation (BECKER et al, 2009). A preservation strategy is a set of actions to preserve a collection of digital objects, which focus only technical aspects (DAPPERT et al., 2009).

Dappert and Farquhar's (2009) conceptual model of digital preservation characterizes and relates the preservation object, its environment, preservation requirements, the risks of preservation, and preservation actions to be taken to mitigate these risks.

According to this model, an object to be preserved (preservation object) is any object that is at risk, directly or indirectly, and needs to be digitally preserved. A preservation object is associated with environments, such as operating system, hardware, application system, community, external factors, and even the legal system. An environment is a set of factors which constrains a preservation object or preservation action and that are necessary to interpret it (DAPPERT; FARQUHAR, 2009).

At a logical layer, a preservation object is viewed as an intellectual entity, i.e. as an "intellectual or artistic distinct creation", (DAPPERT; FARQUHAR, 2009), such as a book or a paper. An intellectual entity may have multiple representations, such as one in PDF, and a version composed by TIFF images. At the physical layer, a representation of an intellectual entity consists of a set of physical objects, called bitstreams (such as computer files). These bitstreams, combined, allow the interpretation of the intellectual object (viewing or hearing). For example, a representation of an article (intellectual entity) consists of five interconnected text files and three picture files, with texts represented in HTML and pictures in JPG.

The description of a digital object involves its intellectual entity, representations, files (bitsreams) and environment. PREMIS is a metadata schema used for this purpose.

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646326

The Dappert and Farquhar's conceptual model (2009) addresses preservation planning considering the following concepts: preservation actions, preservation requirements and preservation risks. Preservation requirements involve the specification of types of preservation actions that are desirable for a preservation object. The preservation requirements can define preservations actions which are independent of the object or environment characteristics (for instance, do not accept a particular file format, or perform a verification process on the subject each year), and, at the same time, dependent on the significant characteristics of the object. The preservation actions that are dependent on the significant characteristics address characteristics of the object or environment that should be maintained or provided to ensure continued accessibility, usability, and meaning of objects, and their capacity to be accepted as evidence of what they purport to record (DAPPERT and FARQUHAR, 2009).

Significant characteristics are central for preservation planning. They address the degree of reliability which is acceptable/desirable for the stakeholders (producers and consumers) for the new representations of the original document that are generated by preservation actions (such as migration) (DAPPERT; FARQUHAR, 2009). During the object life cycle, preservation actions may generate object representations different from the original. In order to deal with these changes, one of the requirements for digital preservation is to define the limits of the object to be preserved, that is, define the object preservation characteristics that must be maintained when performed digital preservation actions that will generate new representations of this object. Significant characteristics specify these limits. The project Investigating the Significant Properties of Electronic Content Over Time (INSPECT) studies and provides resources to identify significant characteristics of objects, for digital preservation propose (HOCKX-YU; KNIGHT, 2008).

Preservation actions are concrete actions to be taken to keep digital object collections active and accessible over time (BECKER et al, 2009). According to Dappert e Farquhar's conceptual model (2009), preservation actions are activated to mitigate risks and to ensure the achievement of the preservation requirements.

Changes in the preservation object or in the environment can lead to preservation risks. Examples of preservation risks are: the emergence of new file versions, media damage, hardware components that are no longer available, file formats that have become obsolete, changes in legislation, lack of support, or even the fact that users need to access the objects via new devices, such as tablets. In the conceptual model of Dappert and Farquhar (2009), preservation risks are specified through preservation risk requirements. When a characteristic of a preservation object or its environment has its value changed to a value that is not specified in the preservation requirements, the preservation object is at risk. Whenever a preservation risk requirement is violated, the digital preservation monitoring process should activate the better preservation action to mitigate risks, in the preservation object. The execution of a preservation action involves the preservation object and its environment, and

may create new copies of preservation of the object (representations) and /or its environments.

Dappert and Farquhar (2009) specify some preservation risks, such as: new version, loss or deterioration, obsolescence or lack of support, etc. Preservation actions can be, for example: change the file system, hardware replacement, support refreshing, format migration, re-orientations in the community, adjust the legal requirements, etc.

The implementation of a preservation action is an event, and the executors of this action are the agents. The PREMIS digital preservation model specifies events and agents. According to PREMIS, an event corresponds to an action that involves or affects at least one object or agent associated with or known by the preservation repository; an agent is a person, organization, or software program/system associated with events in the life of an object, or with rights attached to an object (PREMIS EDITORIAL COMMITTEE, 2015).

Community is another key concept in digital preservation. The OAIS model introduces the concept of designated community, comprising "an identified group of potential consumers able to understand a particular set of information" (CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS, 2012). The OAIS model highlights the importance of identifying and defining the designated community of a digital repository to be preserved, since the planning of preservation is focused on the characteristics of this community. The designated community should be monitored, since changes in their characteristics can affect both the preservation actions and planning.

For Dappert and Farquhar (2009), community is a kind of environment, as hardware and software also are. A characteristic of a community can change, and this may take the preservation at risk, determining the execution of preservation actions. The authors specify producers, consumers and political factors also as environment. They present political factors such as: legislation, standards, reference models and competitors.

A preservation service is a core service that supports a preservation object, as, for example, the preservation planning and the accomplishment of preservation actions. Preservation services are performed manually or by software tools, and are associated with hardware and other environments (DAPPERT; FARQUHAR 2009).

Preservation services are addressed by the DCC curation lifecycle model, the OAIS reference model, and the fundamentals of digital curation of U3C. In the OAIS, services are functional components of a repository. In the DCC life cycle model, services are actions related to the life cycle of digital preservation and curation. In the U3C fundamentals of digital curation, digital curation and preservation services are broken down into micro-services (ABRAMS, KUNZE, LOY, 2010; ABRAMS, CRUSE, KUNZE, 2009).

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646326

In the model depicted in the Figure 1, the services are compiled from the three model (DCC, OAIS and U3C), and the DCC lifecycle model is used as a base for classifying and to place services in a sequence (cycle). The life cycle of DCC classifies services as sequential services, occasional services, and services performed all time. The sequential services are further classified into curation services and digital preservation services. In Figure 1, services connected by an ellipse are sequentially performed. The curation services (depicted with blue border in the figure 1) are: access and reuse, creation or receipt, evaluation and selection, and transformation. Preservation services (with borders in red in figure 1) are ingestion, execution of preservation actions, and storage.

According to the DCC, creating or receiving service involves the creation of data, including administrative, structural, technical and descriptive metadata, and receiving data according to the registered collection policies. Evaluation and selection services mean validate data and select them for preservation and curation, adhering to guidelines, policies and legal requirements. Transformation service aims to create new data from the original, the migration of formats and the creation of subsets for the production of new results. Access and reuse guarantee that data will be accessible to users and re-users (HIGGINS, 2008).

Ingestion, storage and access are nuclear services (functional components) in the OAIS model, since they determine the functioning of the submission, storage and access of information in a repository.

Ingestion means transfer data to a file, archive, data center or curators, adhering to guidelines, policies and legal restrictions (HIGGINS, 2008). For the OAIS (Consultative Committee for Space Data Systems, 2012), ingestion involves legally accepting the submission from the producer, ensuring the quality of the received data, adding information representation (if necessary), and preparing the material for storage, according to standards. The storage service includes the safe guard of deposited items. Preservation monitoring means the preservation actions taken to ensure long-term preservation and retention of the nature of the data (HIGGINS, 2008).

According to the DCC life cycle, description and representation services mean assigning descriptive, technical, structural and preservation metadata (using appropriate standards), and collecting and assigning information representation needed to understand and interpret the digital object and its associated metadata (HIGGINS, 2008).

Environment monitoring service is, according to DCC lifecycle, monitoring community activities, and participating in the development of standards, tools and shared software (HIGGINS, 2008). The OAIS model highlights actions to monitor the designated community and technology, which provide information for the preparation of preservation plans and involve, respectively, interacting with consumers and producers (to track changes in technology services requirements and available products) and tracking technologies, standards and emerging computing platforms (CONSULTATIVE COMMITTEE FOR

SPACE DATA SYSTEMS, 2012).

3 KNOWLEDGE AND PRACTICES IN DIGITAL PRESERVATION AND CURATION

This section discusses aspects of the conceptual model in the context of team performance during the implementation of digital preservation in a repository. It aims to support professionals in the implementation of digital preservation. The section also presents useful resources, such as specifications, guides and examples for the implementation of digital preservation. The issues discussed in the section are:

a) developing and managing digital preservation policies and plans;
b) representing and describing objects of preservation;
c) providing automated services for digital preservation;
d) developing and managing a trusted digital repository

3.1 Developing and managing digital preservation policies and plans

The implementation of digital preservation requires the development of policies and plans. This section identify and characterizes policies and plans, offering examples in order to help developing these instruments.

Preservation policies for libraries and archives, developed by renowned institutions and available on websites, can serve as an example for those who are writing their policies. Some of these policies are enumerated below:

a) National Library of Australia, Digital Preservation Policy (http://www.nla.gov.au/policy---and---planning/digital---preservation---policy)
b) British Library Digital Preservation Strategy (http://www.bl.uk/aboutus/stratpolprog/digi/digitalpreservation/index.html)
c) National Library of Wales, Digital Preservation policy and strategy (http://www.llgc.org.uk/fileadmin/fileadmin/docs_gwefan/amdanom_ni/dogfennaeth_gorfforaethol/dog_gorff_strat_cad_dig_12_15S.pdf)
d) Cornell University Library Digital, Preservation Policy Framework (http://hdl.handle.net/1813/11230);
e) Columbia University Libraries, Policy for Preservation of Digital Resources (http://library.columbia.edu/content/libraryweb/services/preservation/dlpolicy.html)
f) Parliamentary Archives Houses of Parliament London, A Digital Preservation Policy for Parliament (http://www.parliament.uk/documents/upload/digitalpreservationpolicy1.0. pdf)

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646326

g) HathiTrust Digital Library, Digital Preservation Policy
http://www.hathitrust.org/preservation

Preservation policies guide the development of digital preservation plans. According to the guide created by the ISO TC Working Group 46 / SC 11 in 2010 and entitled "*Preservación de documentos digitales: guía "Cómo empezar"*, a preservation plan involves the preservation of a specific collection or a part of a collection of digital objects, considering the preservation policy, the legislation, limitations of the organization, technical limitations, user needs and the purposes of the preservation.

According to the guide (ISO / TC 46 / SC 11, 2010), the preservation plan contains a well-documented specification of actions that ensure access and availability of a collection in the long term. The preservation plan should include information such as: the preservation context, the strategy selected for preservation, the results and the decision of the evaluation of the preservation alternatives, the roles and responsibilities for the preservation plan and its supervision, the actions that trigger the execution of the plan, and the estimated cost for the execution of the preservation plan (preparation and implementation of the plan).

To ensure the authenticity of electronic documents during its lifetime, preservation plans must combine technology and organizational procedures. Plans should ensure that digital documents can be traceable and available for access, open to interpretation, recoverable, protected against abusive use, available for access by authorized people when needed, and supervised with regard to quality of access. The preservation plan must include all aspects, including initial planning, preservation activities, supervision, and continuing review. It also should include all processes and procedures used for the preservation of electronic documents. This should make easy to implementat reliable digital preservation processes, aligned with the general policy of the organization for records management (ISO / TC 46 / SC 11, 2010).

3.2 Representing and describing objects of preservation

The implementation of digital preservation involves defining structures to represent the preservation object and to record information related to this object. This information is metadata. Metadata are structured data that describe and allow us to search, manage and preserve documents over time. They are crucial to the information management in the digital environment. Metadata can be classified as descriptive, technical, structural, and for digital preservation.

The descriptive metadata address the intellectual or creative aspects of the object. Currently, there are many descriptive metadata standards schemes, which address many kinds of resources and communities necessities, such as EAD (for archival description), MODS (for

bibliographic records), VRA (for visual resources), LIDO (for cultural heritage) and Dublin Core (for the web). In many cases, repositories use more than one metadata schema to describe its resources, aiming supply different communities. Aspects of adaptation and combination of metadata schemas are discussed by Heery and Patel (2000) and Chan and Zeng (2006, 2006a).

Basically, technical metadata describe the digital object and its files, considering the technological environment the digital object was produced. The main use of this kind of metadata is to support the object to be viewed, processed and migrated. Usually, most technical metadata are automatically generated at the time of the ingestion of the object, and personnel have free tools for automatic identification of file formats, and for the automatic extraction of format characteristics from a file (called characterization tools). Van der Knijff and Wilson (2014) present a comparative analysis of the most used characterization tools: DROID, Fido, Unix File Tool, FITS and JHOVE2. Their analysis is part of the SCAPE project, which aims to provide scalable services for planning and implementing preservation strategies. For images, technical metadata are usually generated in the MIX and EXIF standards.

Structural metadata describe the representations of digital objects. It involves the description of the logical structure of the document, indicating, for example, the files (from various representations) that compose the pages of the document, and the organization of the document pages into sections (chapters). Structural metadata describe the relationships between the elements of a digital object, organizing, for example, the files that form a complex digital object such a book. METS is the structural metadata schema most used in digital repositories. METS aggregates descriptive, administrative and structural metadata, and can be used, for example, to reconstruct a document from multiple files (as image files of a scanned book), or to identify the relationship between the elements of a collection of documents. Europeana Regia (BACHER, R; FABIAN, C .; Ikas, W., SCHREIBER, C., 2011) is an exemple of project that uses the standard METS to structure the files that make up rare scanned works.

Digital preservation metadata are information required to maintain the viability (permanence and integrity), representation (ability to be represented and displayed by applications) and comprehensibility (ability to be interpreted and understood by the user) of digital objects over time. Preservation metadata are descriptive, structural and administrative metadata that enable long-term preservation of digital material. The PREMIS is a standard for representing digital preservation metadata. It consists of entities, semantic units and relations, and includes the description of the following entities: preservation of object (including intellectual entity, representations and bitsreams), environment, events, agents, rights. All preservation actions carried out on the preservation of an object are recorded as PREMIS event metadada, including the description of the agents who participated in the event.

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646326

Today, METS and PREMIS are used in an integrated way to make packages for storing preservation objects. The digitization project of the Spanish National Library (NATIONAL LIBRARY OF SPAIN, 2015), is an example of preservation planning that combines METS, PREMIS and MIX.

Preservation packages are typically stored in file systems. The BagIt File Packaging Format is a specification of a structure for packing digital objects in file system, developed by the Library of Congress (BOYKO et al, 2012). The project Chronicles in Preservation (SCHULTS et al, 2013) is an example of package specification that combines Bagit, PREMIS and METS. Bagit, METS and PREMIS combination is used in digital preservation systems, such as Archivematica (VAN GARDEREN, 2010).

3.3 Providing automated services for digital preservation

The implementation of digital preservation involves automating services and one of the challenges for the professionals who develop digital preservation is to implement an infrastructure for these services. Automated services that operate on digital preservation, seen individually, are called preservation of micro-services (CALIFORNIA DIGITAL LIBRARY, 2010 and ABRAMS, KUNZE, LOY, 2010). The micro-services are: **ingestion**, which involves virus scanning and document integrity, format validation, package integrity verification and notification of acceptance to the supplier; **description**, comprising extraction and production of technical metadata and records preservation actions carried out; **storage**, among which are file replication and media replacement; **monitoring of object integrity**, in order to verify the integrity of the documents; **access and use**, for searching and generating dissemination packages; and, finally, **transformation**, which involves format migration and the generation of different versions.

According to Van Garderen (2010) and Faria (2009), Archivematica and WHEEL are environments based on micro-services preservation, which follow the functional model for files/repositories from OAIS. RODA is an environment that adds digital preservation services to FEDORA software. This environment was developed for archival collections and is based on archival standards (EAD). Archivematica is a micro-environment service for ingestion, storage and monitoring of preservation. Therefore, Arquivematica is a solution that must be combined with other environments that offer access services. Archivematica was developed to work with a digital repositories management software like ICA-ATOM and DSpace, among others, in an architecture in which ICA-ATOM or DSpace are used only for service access and use. Master files are stored and managed by Archivematica and access versions are deposited in DSpace or ICA-ATOM (VAN GARDEREN, MUMMA, 2013).

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646326

3.4 Developing and managing a trusted digital repository

The implementation of digital preservation in a repository also requires demonstrating that it is reliable, by using assessment and certification tools. It is necessary to establish a climate of confidence around the repository and the information that it safeguards by demonstrating in an objective way its reliability. Transparent procedures must be adopted to show, based on evidence, that all processes and procedures are followed. Therefore, confidence, adoption of the OAIS reference model and certification are key points to be addressed.

Trust develops at different levels. In the case of trusted digital repositories at least three levels of confidence are applicable: that producers are sending the correct information, that consumers are getting the correct information and, finally, that suppliers are providing appropriate services

A trusted digital repository must be able to maintain digital materials as genuine, to preserve them and provide access to them as long as necessary. The main instruments for certification and auditing of a trusted digital repository are the Trustworthy Repositories Audit and Certification (TRAC[5]), DRAMBORA - Digital Repository Audit Method Based[6] and Data Seal of Approval (DAS[7]).

TRAC lists a number of requirements aimed at ensuring confidence around a repository, ranging from organizational management to support infrastructure. TRAC aims to: a) provide a tool that allows to audit, evaluate and potentially certificate digital repositories, b) establish the necessary documentation for an audit, c) outline the certification process, and) establish appropriate methodologies for determining the robustness and sustainability of a digital repository (FERREIRA, HAIL and RODRIGUES, 2012).

This instrument consists of evaluation criteria, which address aspects related to the **organizational structure** (governance and organizational viability, organizational and personnel structure, transfer, financial sustainability, contracts and licenses), **digital object management** (capture, generation of archiving package, preservation planning, storage, information management and access) and **technology, technical infrastructure and security**. The criteria for the digital object management are categorized according to the functional components of the OAIS model. This certification method resulted in the ISO 16363 standard.

DRAMBORA is a document and an interactive tool that provides de means for a self evaluation process. It focuses primary on aspects related to strategic management and organization, and secondarily on technical aspects related to the repository and its technology

[5] https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf
[6] http://www.repositoryaudit.eu
[7] http://www.datasealofapproval.org/en/

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646326

platform. It aims to encourage managers to develop an organizational profile, documenting its deposit policy, objectives, responsibilities, activities and guarded materials; identify and assess the risks that could prevent the achievement of its mission and threaten the protection of their materials; effectively manage the risks and mitigate their probability of occurrence; establish effective contingency plans to minimize the effects caused by risks that could not be avoided and report the results of the self assessment process.

DAS certifies that a repository is able to preserve scientific data for future use and processing without incurring high costs or investments to custodial organizations.

4 CONCLUDING REMARKS

Institutions are generating more and more documents and thus need to adopt good practices to manage them. Digital documents have some unique features that require the adoption of specific measures to ensure that maintain their value as long as necessary.

Curation and digital preservation are subject of debates, research and publications, but also demand practical solutions on the part of those responsible for digital repositories in educational and research institutions, which find themselves compelled to ensure continued access to digital objects in their custody in the long term.

In this sense, this article focused on the implementation of digital preservation repositories from the perspective of the team responsible for the project, stressing the necessary knowledge and its relationship with the practices adopted by the community. The conceptual model of curation and digital preservation presented and discussed in this article, which was synthesized from previous work, has shown the main aspects of the domain and pointed to the issues that must be observed in practice.

The paper points to the need to further strengthen the relationship between domain knowledge in curation and digital preservation repositories, with the practices of many projects developed globally.

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646326

REFERENCES

ABBOUT, Daisy. What is digital curation? Edinburgh, UK: Digital Curation Centre, 2008. Available at: < http://www.era.lib.ed.ac.uk/ bitstream/1842/3362/3/Abbott%20What%20 is%20digital%20curation_%20_%20Digital%20 Curation%20Centre.doc>. Accessed 20 Jun. 2016.

ABRAMS, Stephen; KUNZE, John; LOY, David. An emergent micro-services approach to digital curation infrastructure. **International Journal of Digital Curation**, v. 5, n. 1, p. 172-186, 2010. Available at: <http://www.ijdc.net/index.php/ijdc/article/viewFile/154/217>. Accessed 06 Jun. 2016.

ABRAMS, Stephen; CRUSE, Patricia; KUNZE, John. Preservation is not a place. **International Journal of Digital Curation**, v. 4, n. 1, p. 8-21, 2009. Available at: <http://www.ijdc.net/index.php/ijdc/article/viewFile/98/73>. Accessed 06 Jun. 2016.

ASSOCIATION FOR LIBRARY COLLECTIONS & TECHNICAL SERVICES. **Definitions of digital preservation**. ALA Annual Conference, Washington, D.C., June 24, 2007. Available at: <http://www.ala.org/alcts/resources/preserv/defdigpres0408>. Accessed 07 Jul. 2016.

BACHER, Rahe et al. State of the Art in Image Processing. Europeana regia, 2013. Available at http://www.europeanaregia.eu/sites/www.europeanaregia.eu/files/europeana-regia-state-of-the-art-in-image-processing.pdf

BECKER, Christoph et al. Systematic planning for digital preservation: evaluating potential strategies and building preservation plans. **International Journal on Digital Libraries**, v. 10, n. 4, p. 133-157, 2009. Available at: <http://www.ifs.tuwien.ac.at/~becker/pubs/becker-ijdl2009.pdf>. Accessed 30 Jun. 2016

BIBLIOTECA NACIONAL DE ESPAÑA. Proceso de digitalización en la Biblioteca Nacional de España. 2015. Available at: <http://www.bne.es/webdocs/Catalogos/ProcesoDigitalizacionBNE.pdf>. Accessed 28 Jun. 2016.

BOYKO, A. et al. The BagIt File Packaging Format (V0.97), 2012. Available at: <http://www.digitalpreservation.gov/documents/bagitspec.pdf>. Accessed 12 Jul. 2016.

CALIFORNIA DIGITAL LIBRARY. **UC3 Curation Foundations**. Rev. 0.13, 2010. 29 p. Available at: <https://confluence.ucop.edu/download/attachments/13860983/Merritt-latest.pdf?version=1&modificationDate=1269561985000>. Accessed 04 Jul. 2016.

CHAN, Lois Mai; ZENG, Marcia Lei. Metadata interoperability and standardization: A study of methodology, Part I. **D-Lib Magazine**, v. 12, n. 6, p. 3, 2006.

CHAN, Lois Mai; ZENG, Marcia Lei. Metadata interoperability and standardization: A study of methodology, Part II. **D-Lib magazine**, v. 12, n. 6, p. 4, 2006.

CONSELHO NACIONAL DE ARQUIVOS. Diretrizes para a implementação de repositórios digitais confiáveis de documentos arquivísticos. Rio de Janeiro: Conarq, 2014. 25 p.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS, 2012, **Reference model for an Open Archival Information System (OAIS)**: recommended practice CCSDS 652.0-M-2, Washington, DC: CCSDS, 2012. Available at: <http://public.ccsds.org/publications/archive/650x0m2.pdf>. Accessed 30 Jun. 2016.

DAPPERT, Angela; FARQUHAR, Adam. Modelling organizational preservation goals to guide digital preservation. **International Journal of Digital Curation**, v. 4, n. 2, p. 119-134, 2009. Available at: <https://www.researchgate.net/publication/242084343_Modelling_Organizational_Preservation_Goals_to_Guide_Digital_Preservation>. Accessed 07 Jul. 2016.

DAPPERT, Angela et al. **Report on policy and strategy models for libraries, archives and data centres**: PLANETS report PP2-D3. 2009. Available at: <http://www.planets-project.eu/docs/reports/Planets_PP2_D3_ReportOnPolicyAndStrategyModelsM36_Ext.pdf>. Accessed 04 Jul. 2016.

FARIA, L. et al. RODA: a service-oriented repository to preserve authentic digital objects. In INTERNATIONAL CONFERENCE ON OPEN REPOSITORIES,4, Atlanta, USA, 2009. **International Conference on Open Repositories**. 2009.

FERREIRA, Miguel; SARAIVA, Ricardo; RODRIGUES, Eloy. Estado da arte em preservação digital. 2012. Available at: <http://hdl.handle.net/1822/17049>. Accessed 28 Jul. 2016.

GRACIO, José Carlos Abbud. **Preservação digital na gestão da informação**: um modelo processual para as instituições de ensino superior. São Paulo: Cultura Acadêmica, 2012. 218 p. Available at: <http://www.culturaacademica.com.br/_img/arquivos/Preservacao_digital_na_gestao_da_informa cao-WEB_v2.pdf>.

HEERY, Rachel; PATEL, Manjula. Application profiles: mixing and matching metadata schemas. **Ariadne**, n. 25, 2000.

HIGGINS, Sarah. The DCC curation lifecycle model. **International Journal of Digital Curation**, v. 3, n. 1, p. 134-140, 2008. Available at: http://www.ijdc.net/index.php/ijdc/article/viewFile/69/48. Accessed 06 Jun. 2016.

HOCKX-YU, Helen; KNIGHT, Gareth. What to preserve?: significant properties of digital objects. **International Journal of Digital Curation**, v. 3, n. 1, p. 141-153, 2008. Available at: <http://www.ijdc.net/index.php/ijdc/article/view/70/49>. Accessed 30 Jun. 2016.

MÁRDERO ARELLANO, Miguel Angel. Preservação de documentos digitais. **Ciência da Informação**, Brasília, v. 33, n. 2, p. 15-27, 2004. Available at: <http://www.scielo.br/pdf/ci/v33n2/a02v33n2.pdf>. Accessed 4 Jan. 2007. DOI: 10.1590/S0100-19652004000200002.

MYLOPOULOS, J. Conceptual Modeling and Telos. In: Loucopoulos, Peri and Zicari, Roberto. **Conceptual Modeling, database, and case**: an integrated view of information systems development. New York: John Wiley & Sons, 1992. p. 49-68. Available at: <http://dl.acm.org/citation.cfm?id=573368>. Accessed 07 May 2016.

PREMIS EDITORIAL COMMITTEE. **PREMIS**: Data Dictionary for Preservation Metadata version 3.0, 2015. Available at: <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>. Accessed 04 Jul. 2016.

PRESERVACIÓN de documentos digitales: guía "Cómo empezar": ISO/TC 46/SC 11. 2010. Available at: <http://www.arxivers.com/index.php/documents/documentacio-1/normativa-tecnica-1/514-tec-isotc46sc11-wg7-preserv-docs-digit-es-1/file>. Accessed 04 Jul. 2016.

SCHULTZ, Matt et al. Building Institutional Capacity in Digital Preservation. In: **IPRES Annual Conference**, 2013. Available at: <http://purl.pt/24107/1/iPres2013_PDF/Building%20Institutional%20Capacity%20in%20Digital%20Preservation.pdf>. Accessed 12 Jul.2016.

SIEBRA, S. A. et al. Curadoria digital: além da questão da preservação digital. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14, 2013, Florianópolis. **Anais.**.. Florianópolis: UFSC, 2013. Available at: <http://www.egov.ufsc.br/portal/sites/default/files/curadoria_digital_0.pdf>. Accessed 22 Jul. 2016.

VAN DER KNIJFF, Johan; WILSON, Carl. Evaluation of characterisation tools. SCAPE Project. 2014. Available at: <http://scape-project.eu/wp-content/uploads/2014/08/SCAPE_PC_WP1_identification21092011.pdf>. Accessed 12 Jul. 2016.

VAN GARDEREN, Peter. Archivematica: Using micro-services and open-source software to deliver a comprehensive digital curation solution. In: PROCEEDINGS OF THE 7TH INTERNATIONAL CONFERENCE ON PRESERVATION OF DIGITAL OBJECTS**,** Vienna, Austria. 2010. p. 145-149

VAN GARDEREN, Peter; MUMMA, Courtney C. Realizing the Archivematica vision: delivering a comprehensive and free OAIS implementation. In: 10TH INTERNATIONAL CONFERENCE ON PRESERVATION OF DIGITAL OBJECTS. 2013. p. 84.

WALTERS, Tyler; SKINNER, Katherine. **New Roles for New Times**: Digital Curation for Preservation. Washington, DC: Association of Research Libraries, 2011. 76 p. Available at: https://vtechworks.lib.vt.edu/bitstream/handle/10919/10183/nrnt_digital_curation17mar11.pdf?sequence=1&isAllowed=y. Accessed 15 Jun. 2016.

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646326