InterPARES 3 Project

International Research on Permanent Authentic Records in Electronic Systems

Title: Case Study 04 – Library Network of the State University of Campinas (UNICAMP) – Digital Theses and Dissertations: Final Report.

Status: Final (public).

Version: 1.2

Date Submitted: December 2011.

Last Revised: February 2012.

Author: The InterPARES 3 Project, TEAM Brazil.

Writer(s): Luiz Atílio Vicentini
Regina Aparecida Blanco Vicentini
Danielle Thiago Ferreira
UNICAMP's Library Network

Fábio Rodrigo Pinheiro da Silva
Neire do Rossio Martins
UNICAMP's Central Archives

Brenda Couto de Brito Rocco
National Archives of Brazil

Project Component: Research.

Document Control

Version history			
Version	Date	By	Version notes
1.0	2012/1/27	Neire Martins Janaína dos Santos Thiago Sbrici	Discussion and making of report.
1.1	2012/2/3	Brenda Rocco	Revision of report.
1.2	2012/2/9	Fábio R. P. Silva Andressa Piconi	Making of final draft.

Table of Contents

Report of a case study

A. Outline

Case Study 04, "Library Network of the State University of Campinas (UNICAMP) – Digital Theses and Dissertations," targeted the proposal and implementation of an action plan for the preservation of the digital theses and dissertations produced by postgraduate students on their way to get Doctor's and Master's degrees, as well as accessioned, managed and made available for access by the university's Digital Library, so that they remain reliable, authentic and accessible over time, given their importance as proof that the student completed the post graduation course program.

To fulfill that purpose, a workgroup was created at UNICAMP with members from the Central Archives of the university's Archives System and from the Library Network, under the guidance of the National Archives of Brazil.

B. Methodology

The case study followed the methodology proposed by the InterPARES Project. UNICAMP researchers collected data based on specific guidelines defined by the project, aiming to gather information about the context of creation and characteristics of the digital entity. This data collection was carried out through interviews held with the personnel in charge of the maintenance of the theses and dissertations and of the Digital Library. After data collection, the researchers conducted the diplomatic analysis of the record and the action plan.

C. Description of context

Provenancial context

The Library Network is part of UNICAMP, a public institution of autarchic nature under special rule of São Paulo state's government administration. It was established in 2003, as part of the university's General Coordination, having as goals: to support educational, research and outreach programs; to define the development policy to be adopted for the various collections of which the university libraries are comprised; to enable access to the university and scientific communities to the information generated and

stored at UNICAMP; and to promote exchange of experiences and collections. Its mission is to provide access to and the retrieval of information, supporting the teaching, research and outreach programs; thus UNICAMP's Digital Library was created with the main objective of having the scientific production developed by the university available on the Web.

Juridical-administrative context

The Digital Library is part of the Library Network's organizational structure and it is regulated by a set of normative acts that supports its role, such as its inauguration act[1], the national legislation on copyright[2] and UNICAMP's rules on the availability, via Internet, of theses and dissertations produced at the university[3]. The arrangement of the records, as well as the maintenance of their integrity,[4] follow an act of UNICAMP's Council that provides guidelines for records management.

Procedural context

After the conclusion of the thesis/dissertation by the student and their teacher (advisor), it must undergo defense and approval before an examining committee. Then, the student prepares the printouts and digital versions for ratification by the designated authority; after that procedure, the Library Network receives the thesis and dissertation, both in print and in digital form, for cataloguing, insertion into the database and legal deposit.

Documentary context

"Theses and dissertations" represents a documentary series under the function "Postgraduate *Stricto Sensu* Education Management," a sub function of "Implementation and Monitoring of Postgraduate *Stricto Sensu* Courses" and an activity "Master's and Doctor's Degree Awarding," which pertains to the fonds "Library Network – State University of Campinas."

_____

[1] UNICAMP, GR no. 85, of Nov. 8, 2001.
[2] Law no. 9610, of Feb. 19, 1998.
[3] UNICAMP, Attorney General's Report no. 3325/02; Determination of the Central Post Graduation Committee no. 51/06.
[4] UNICAMP, CONSU A-8/1995.

Technological context

The Digital Library uses the software "NouRau," developed by its own staff, in which digital records are stored in PDF format and retrieved from a database whose descriptors use MARC standards. Open-source Linux operating system (GNU / Linux-Debian Woody) serves as its platform, using free tools, such as: PHP 4.3.5, Perl 5.8, Apache 1.3.29, PostgreSQL 7.4.2. In addition to that software, the platform is complemented with: FILE, HTDIG, ANTIWORD, DVI2TTY, PSTOTEXT, RECODE, XLHTML and XPDF.

D. Narrative answers to the applicable set of questions for researchers

The theses and dissertations are presented by postgraduate students at UNICAMP as a requirement to get Master's and Doctor's degrees. The original paper is stored in a legal deposit at the Central Library of the Library Network, a digital version is also sent to the Library Network for insertion into the Digital Library, allowing public access to the full texts via Internet.

The software "NouRau" associates the Digital Library's metadata with the theses and dissertations. It captures that metadata by interoperating with the "Base Acervus" (another UNICAMP cataloguing system, under the responsibility of the Library Network's Central Library). The description of the records is manually supplemented with the following metadata: titles in Portuguese and in English, author(s), keywords in Portuguese and English, name of the area of concentration, degree (Master's or Doctor's), names of examining committee members, abstracts in Portuguese and in English, date of defense, and code (identifier/cataloguing database with the Library Network). Other information is entered automatically in the system, such as: language, "owner" (responsible for the capture in the Digital Library), format, file name and size, estimated download time, date and time of creation, number of visits and downloads, and the identifier code of the record in the Digital Library.

Such metadata is recorded in tables in the PostgreSQL database, which are components of the digital theses/dissertations, along with the PDF file of the thesis/dissertation itself. For

each digital file inserted, a unique and persistent identifier called "bibid" is generated by the cataloguing software of the bibliographic database ("Base Acervus").

In order to ensure the authenticity of the data, there are security policies that allow access to the database only to the Digital Library's IT technicians, while responsible operations for the creation and publication of digital thesis and dissertations are restricted to some members of Electronic Information Access Program and the UNICAMP Department of Information Processing System Library. In addition, there are procedures that allow the tracking of events for audit purposes. The storage of these documents on the server itself, with back-up system and access through the interface of the Nou-Rau, aims to ensure their reliability.

Changes in thesis and dissertations occur through formal requests by authors and their advisors who are registered and kept in files.

The digital thesis and dissertations are related to their original paper and the student "Process of School Life" – it brings together the student documents dossier, since their enrollment at the University until the assessment of the thesis / dissertation, and which is guarded by the UNICAMP File System. However, during the study, it was found that these organic relations were not explained and, as a solution were proposed to insert the number of "School Life Process" in the original paper, by rubber stamp, as well as the inclusion of this number in the database, such as metadata.

E. Narrative answers to the project's applicable research questions

Which are the regulatory, auditing and policy making bodies that need to be sensitized to the importance of digital preservation, and what are the best ways of influencing them?

The organisms to be sensitized to it are those related to the strategic planning and the making of policies concerning archives, information and ICT, which are, in the context of a university, the higher deliberative bodies, such as the university council and colleges and institutes' boards. One way to influence them is to create, within the archival field, committees or groups composed of teachers, researchers, students and managers, in order to develop studies and proposals for digital preservation. Then awareness may reach the organization's areas horizontally and vertically, depending on their participation in these committees, which should include members of academic and administrative bodies

(including maintenance and supervision). Those groups may jointly propose, therefore, actions and procedures to the higher and decision-making bodies for the resolution of the management and preservation of records.

How and when should these archives or programs prepare themselves for digital preservation?

The institutional archives should prepare guidelines and best practices for preservation, enabling record creators to carry out, in their local archives, measures to ensure the reliability, authenticity and access of records, from the time of their creation and for as long as necessary.

What differentiates the preservation of digital records from that of any other digital entity for which the archives might be responsible?

The preservation of digital records, unlike the preservation of other digital entities, must be able to guarantee the basic characteristics of the archival document, particularly what comes to its fixed form, stable content and archival bond with the other documents that register the action.

What are the nature and the characteristics of the relationship that each of these archives or programs should establish with the creators of the records for which it is responsible?

There should be partnership between creators and archives in building solutions for digital preservation. Archives and creators must know each other and the former should provide guidance on the management and preservation of archival documents; the latter, in turn, should provide information about their operations and documents, allowing the identification of those which are archival, and be willing to incorporate archival knowledge in their work processes. That interaction is required for the joint development of solutions that enhance the creation and filing of records, in order to maintain their reliability, authenticity and access.

What action plans may be devised for the long-term preservation of these bodies of records?

In the case study of UNICAMP's Library Network, it was verified that the digital theses and dissertations are kept in the PDF file format and retrieved through metadata

embedded in the database NouRau, forming a group to be preserved. Despite displaying fixed form and stable content, the documents were potential records because their archival bond with other records of the same action was not explicit in the metadata. It was also found that the system and procedures used in the Digital Library needed to be strengthened when it came to the capture, security, management and preservation of the records.

For that purpose, a set of actions in the shape of a plan was established as the initial solution. Some of these actions have been implemented, such as the inclusion of the student's file number in the printed version of the thesis or dissertation, and the inclusion of its corresponding metadata in the database. However, other actions are still being implemented, for instance: the development of a new website for the Digital Library, to allow the upload of attachments that accompany the theses and dissertations; the monitoring and prevention of possible changes in the documentary form, due to digitization or conversion of records to the format required by the Digital Library; the implementation of audit trails in order to allow the tracking of events; the generation of an XML file for indexing theses/dissertations and its respective attachments to be archived together; the implementation of a mechanism to check the integrity of the file by installing antivirus and MD-5 keys, so as to verify the file's size and format.

> *Can the action plan chosen for a given body of records be valid for another body of records of the same type, produced and preserved by the same kind of organization, person, or community in the same country?*

Yes, it is possible, if the conditions are the same as the ones raised in this case study.

> *Can the action plan chosen for a certain type of record or system be valid independently of the creating or preserving organization and its context?*

It is possible that other institutions that have the same characteristics can adapt it to their realities.

> *How can records professionals keep their knowledge of digital preservation up-to-date in the face of ongoing and increasingly fast technological change?*

Records professionals can keep up by participating in open study groups (including networks), courses, conferences and meetings on the issues, or even by establishing

programs and forming study groups in their own organizations, in partnership with national and international projects.

F. Diplomatic analysis

The diplomartic analysis was held and led to the conclusion that the theses and dissertations should be kept in their stored form, i.e. all digital components that make up the document (the PDF file and the metadata).

The theses and dissertations included in the Digital Library of UNICAMP are manifested documents related to a set of digital components. The digital components include the PDF file with the thesis or dissertation and any attachments in various formats, as well as records from the database with the metadata of the document.

The theses and dissertations are potential records. They have fixed form and stable content, as well as identifiable contexts and people involved in the action they purport. However, their archival bond is not explicit, neither among the documents comprised in the activity, nor in the identification of their digital components, or even in the organization's classification plan.

Thus, preservation strategies must include records management procedures regarding archival bonding: PDF files, the attachments in various formats and tables in the database with related metadata, as well as a review.

G. Conclusion

The case study proved to be possible to apply the body of concepts, principles and methods of the InterPARES Project at the university, even when dealing with libraries. The Digital Library of UNICAMP's Library Network is a preserver of digital records. After the identification of the digital record and its contexts (creation, provenance, documentary, legal, administrative and technological), as well as its diplomatic analysis, it was concluded that the theses and dissertations are potential digital records because they present fixed form and stable content, in addition to identifiable people and contexts. However, the archival bond is not explicit among the records comprised in the activities listed in the institution's file plan.

In addition, the workgroup identified the need to plan improvements concerning capture, system security and maintenance, as well as the monitoring of reformatting processes and of the system itself, in order to increase reliability, for example. To this end, some further improvements were regarded as needed so as to: clarify the archival bond between the digital entity and other digital records through metadata; monitor PDF and the various formats of the attachment files when it comes to updates and the implementation of migration procedures; monitor and prevent possible changes in documentary form resulting from digitization and conversion of the original file to the format required by the Digital Library; and carry out an audit trail to track events.

Some definitions and solutions were implemented: the document is now filed in its manifested form, i.e. maintained in the form rendered by the Digital Library, accompanied by a set of context metadata; the database will be preserved, as well as its relationships with digital objects (theses and dissertations); the archival bond will be kept by including metadata (classification code of the record, number of the postgraduate student's dossier and the description of other records of the creator, generated during the management of the Digital Library); security requirements should be strengthened through the identification of agents and the implementation of audit trails; and the monitoring of file formats and of the system itself will be subject to applications to be developed from digital preservation requirements duly published by reputable institutions.

Another important issue raised was the need to develop a plan for the preservation of digital records under the Library's custody, also including, among its more specific guidelines, the need to form a group to continue the study on digital preservation at UNICAMP, which could serve other libraries within the Library Network and the partnership with the university's Archives System for the development of joint institutional solutions.