Arley Soto

Digital preservation of scientific e-journals: Colombian case study

Master thesis
International Master in Digital Library Learning
2015

[ABSTRACT]

Largest publishers and biggest libraries around the world have treated the long-term preservation of scientific e-journals with high standard practices. There are international initiatives that provide both libraries and publishers a digital redundancy of digital works, and availability of the content if a trigger event happened, so they are considered good practices to assure long-term access for users. Some national libraries have also a key role in the preservation of scientific e-journals. However, these best practices are not a reality for small publishers or independent journals in Latin America.

The aim of the study is to have a better understanding about the digital preservation of scientific e-journals in Colombia. The work was done under the case study research method. Ten interviews were applied to high-level people from several organizations involved in the process of digital preservation of scientific e-journals. The analysis was a combination of deductive and inductive methods, from a qualitative approach, using the grounded theory and constant comparative analysis technique.

The results are that digital preservation of scientific e-journals is a multilevel practice, its main objective is to reduce the risk of losing digital content and the aim is to provide long term access to the users. Digital preservation is influenced by the cultural and organizational change, because people involved in the process are still living a transition from printed to digital-born issues. Digital preservation is part of a long-term planning at several levels; it requires a synergy between the actors involved, including policies at journal, publishers, research and national level. All of these policies required meeting standards and copyright requirements to became the preservation feasible. The policies should include not only roles and responsibilities but also the strategies and infrastructure required. As conclusion, Latin America countries have the responsibility and the opportunity to provide long-term solutions to scientific e-journals, but they need to undertake actions at journal, publisher, research and national level.

Keywords: Electronic publishing; Digital preservation; E-journals; Case study; Colombia; E-journal preservation; Scientific communication; Scientific E-journals; Scholarly publishing; Grounded theory; Constant comparative analysis

Master thesis
International Master in Digital Library Learning
2015

TABLE OF CONTENTS

LIST OF TABLES

LIST OF IMAGES

LIST OF ABBREVIATIONS

CLOCKSS: Controlled LOCKSS

CMS: Content Management Systems

COLCIENCIAS: Departamento Administrativo de Ciencia, Tecnología e Innovación

CRL: Center for Research Libraries

DRAMBORA: Digital Repository Audit Method Based On Risk Assessment

ERPANET: Electronic Resource Preservation and Network

ICT: Information & Communication Technologies

ISO: International Organization for Standardization

JISC: Joint Information Systems Committee

LOCKSS: Lots of Copies Keep Stuff Safe

NESTOR: Competence Network on Digital Preservation

OAIS: Open Archival Information Systems

OCLC: Online Computer Library Center

OJS: Open Journal Systems

OSUL: Ohio State University Libraries

PEPRS: Pilot an E-journal Preservation Registry Service

PKP: Public Knowledge Project

PLN: Private LOCKSS Network

PLANETS: Preservation and Long-term Access through Networked Services

PREMIS: Preservation Metadata: Implementation Strategies

TRAC: Trustworthy Repositories Audit & Certification

DECLARATION AND PLAGIARISM DISCLAIMER

"The opinions expressed in this dissertation are solely those of the author and acceptance of the dissertation as a contribution to the award of a degree cannot be regarded as constituting approval of all of its contents by the Faculty of Journalism, Library and Information Science".

I certify that all material in this dissertation which is not my own work has been identified and properly attributed.

Signed:

Date: October 30th, 2015

CHAPTER 1. INTRODUCTION

1.1. Background

According to the Colombian body of research there are 542 scientific journals in Colombia (Colciencias, 2015). Some of them are becoming electronic journals, has an electronic version or are completely digital-born journals. The majority of them are published by Universities (around the 88%) and the remaining 12% are published by other research organizations, associations or corporate bodies (Colciencias, 2012). These numbers give us a quick picture about the size of scientific journals available in Colombia. For the purpose of this study, a journal is considered "scientific" if the works submitted to the journal are approved after a peer-review process. There are not studies that analyze the risk that Colombian scientific e-journals have regarding long-term preservation, and there is not evidence of previous research work related with the subject in Colombia.

Not all the publishers are depositing electronic version of their journals to a third party organizations such as the National Library, PORTICO, CLOCKSS or other, so there is a clear risk that this content will be lost if no body is taking care of their digital archiving, mainly when a publisher decides to finish certain publication or stop producing more issues.

Considering the risk of e-journals to became unavailable in the future, we try to ask the question about what are the roles and responsibilities, and who is or should be in charge of

such responsibilities and what are the plans and obstacles for providing digital preservation solutions for Colombian scientific e-journals.

Chanod, et. al. (2010) highlighted that the ideal solution to the preservation challenge should be an (almost completely) automated environment for the creation of preservation metadata and the definition of a preservation workflow, however they also recognized the importance of policy and rule management for communicating service expectations and requirements between stakeholders involved in the digital preservation domain.

The research questions that guides this study are:

> *What are the roles and responsibilities involved in the digital preservation of scientific e-journals?*
>
> *What organizations should undertake these responsibilities?*
>
> *How Colombian e-journals are planning for long-term preservation?*
>
> *How deposit agencies are dealing with long-term preservation of e-journals?*
>
> *What are the plans and obstacles for providing digital preservation solutions to scientific e-journals?*

All of these pretend to contribute in some way to the development of the digital preservation subject and further research in the area.

1.2. Aims and objectives

The aim of the study is to provide a better understanding about the current status of long-term preservation for Colombian scientific e-journals, as well as identify stakeholders, roles, responsibilities and possible actions for assuring permanent and continuous access to the content published by Colombian scientific journals.

Objectives

- To analyze the perception and expectations about the roles and responsibilities regarding digital preservation for Colombian scientific e-journals.

- To identify the organizational needs and possible obstacles for providing digital preservation solutions for Colombian scientific electronic journals.

1.3. Scope

For the purpose of the research, we will consider digital preservation as the series of managed activities necessary to ensure continued access to digital materials for as long as necessary, including all the actions required to maintain access to digital materials beyond the limits of media failure or technological change. We will also consider digital preservation as long-term preservation, that means continued access to digital materials, or at least to the information contained in them, indefinitely (Jones & Beagrie, 2001). This research excludes the use of technology to preserve the original artifacts through digitization, mainly because the study will focus on e-journals that are available on line (born-digital materials). We have excluded journals published in physical forms like CD-ROMS, DVD-ROMS, USB, and others.

The study considers the key actors involved in the scientific publishing processes and digital preservation in Colombia, including the National Body of Research, Legal depositing agencies,

Publishers and Journals. When mentioning scientific journals I refer to all journals that have a peer review process made by experts to evaluate manuscripts for potential publication, that's mean the 542 journals included in Publindex - National Index of Journals.

1.4. Ethical considerations

I certify that I do not have any of real, evident or potential conflict of interest regarding the research, neither for personal or commercial interests.

I have followed the guidelines of the Statement of Ethical Practice for the British Sociological Association (British Sociological Association, 2012) with special care to the informed consent. All interviewed participants agreed and signed the Consent for participation in interview research (Appendix 01) and the confidentiality, privacy and anonymity of research participants will be guaranteed.

I provided all the information details about the research procedures, aims and methodology to the institutions and persons within the case studies. The participation on the research was entirely voluntary, and the participants were free to decline to answer any question or to end the interview. They were aware that they were being recorded and that the information contained in the transcription and notes may be used to the general public. All research participants got the Spanish translation of the Informed Consent for a better understanding about the terms and conditions to be part of the study (Appendix 02).

1.5. Limitations

The main restrictions for this study were lack of time and resources. The researcher and the participants were too busy in other projects and activities, so that trying to schedule an appointment was difficult in some cases. This lack of time prevented us to have further

interviews with the same participants, and because the lack of resources it was difficult to organize focus group meetings with all participants, which could have been a great opportunity to have richer data and open discussions.

1.6. Timetable

	2015								
	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.
Research design									
Designing methodological approach									
Identifying possible sources of evidence									
Designing interviews									
Presenting the proposal at Parma semester									
Reviewing the methodological approach									
Lighting talk at PKP Conference, Canada									
Data gathering									
Non-participant observation									
Make initial contact									
Conducting interviews									
Transcribing interviews									
Literature review									
Analysis									
Open coding									
Axial coding									
Selective coding									
Writing findings research									
Writing findings research									
Writing conclusions and implications									
Final editing and proofreading									

Table 1. Research project timetable

CHAPTER 2. METHODOLOGY

This study has been done under the interpretivist paradigm, considering that the reality of the phenomena is constructed among the interaction between the subject (the researcher) and the object (the case), so that the perspective will be entirely a qualitative approach. Consequently, the context is extremely important, and the transfer of findings depends on the specific characteristics of similar scientific publishing and digital preservation practices in other countries.

The aim of the study is to have a better understanding about the digital preservation of scientific e-journals in Colombia and the work was done under the case study research method. Eleven interviews were applied to high-level people from several organizations involved in the process of digital preservation of e-journals. The analysis was inductive and qualitative, using the grounded theory and constant comparative analysis technique.

The research question remains about the plans for long-term digital preservation for scientific e-journals in the Colombian context. For answering this question, the first step was a literature review about the scientific communication process and digital preservation foundations. Then we selected the case study as the most appropriate method to undertake the research. After that, I designed the instruments to gather the data, made the interviews, and analyzed the data using the inductive technique of the constant comparative analysis.

We have selected the case study method because it provides a way to make an in-depth analysis about the conditions, goals and obstacles to provide long-term digital preservation solutions for scientific e-journals in the context of the different organizations involved within digital preservation and scientific publishing in in Colombia. As stated by Yin (2013) the case study is an empirical inquiry that investigates a contemporary phenomenon within its real-life context; when the boundaries between the phenomenon and context are not clearly evident; and in which multiple source of evidence are used. This method assures to conduct a deep investigation about the broader context scientific publishing in Colombia but also about the specific problem of long term preservation of Colombian scientific e-journals.

The data has been gathered using multiple data collection techniques with multiple source of evidence in order to fulfill a triangulation that guarantee strong support for the results of the research. We have selected interviews and observation as techniques for data collection. For interviews, we have decided to ask similar questions to different people at the different organizations, in order to analyze different biases, opinions and facts. The purpose of this triangulation is to collect information from multiple sources but aimed at corroborating the same facts or phenomenon (Yin, 2013).

2.1. Research method

The research method used for carrying out the study was an exploratory case study because the main objective of the inquiry is to have a better understanding about the roles, responsibilities and the broader context of long term preservation for scientific e-journals. So that, we will have, as an outcome, a clear picture about the phenomena in Colombia. We will not pretend to explain causal links of the problem or describe it deeply; so it is not explanatory

neither descriptive. The case study has used the holistic design, examining the global nature of the problem.

For the purpose of this study we will use the definition of case study provided by Gar Thomas (2011):

> *Case studies are analyses of persons, events, decisions, periods, projects, policies, institutions or other systems, which are studied holistically by one or more methods. The case that is the subject of the inquiry will be an instance of a class of phenomena that provides an analytical frame – an object – within which the study is conducted and which the case illuminates and explicates.*

In addition, according to Yin (2013) the case study inquiry:

- *Copes with the technically distinctive situation in which there will be many more variables of interest than data points, and as one result.*
- *Relies on multiple source of evidence, with data needing to converge in a triangulation fashion, and as another result.*
- *Benefits from the prior development of theoretical propositions to guide data collection and analysis.*

The objects of analysis are the potential organizations involved in the digital preservation of scientific e-journals in Colombia. As the intrinsic case study is carried out for no other purpose than to give us a better understanding of the case (Pickard, 2013), the analysis will be focus on the digital preservation plans at these organizations and all the associated complexities. Case studies, like experiments, are generalizable to theoretical propositions and not to populations or universes (Yin, 2013). The goal of the case study is to expand and generalize theories (Analytic generalization) and not to enumerate frequencies (statistical generalization).

2.2. Sampling strategy

As it is mentioned above, the study has done under the interpretivist paradigm and has mainly a qualitative approach, so the criteria to select the sources of data has been the potential richness of information of each case. As Patton stated, the logic of purposeful sampling lies in selecting information rich-cases for study in depth. Information–rich cases are those from which one can learn a great deal about issues of central importance to the purpose of the research (Patton, 2005). According to that we selected 11 people from 8 institutions involved in the process of digital preservation or scientific publishing that could provide us enough information to analyze the phenomena in Colombia.

We have focused in the National Library because we consider it is a typical case of National Libraries in Latin America and its analysis could be informative about the situation of the long-term preservation of local scientific e-journals. Four people from the National Library were participants of the study. The reason for focusing in this institution is because they are the main deposit agency of publications produced in Colombia. We also analyze the context, including organizations such as the National Body of Research (Colciencias), and the National Archive (Archivo General de la Nación), as well as some editors and publishers. A total of 11 people from 8 institutions were interviewed, and all of them are sources of data from different institutions involved in the process of digital preservation or scientific publishing. Three of the participants did not agree to be recorded, but their comments were annotated as observations. The following table shows how the participants are enrolled with certain institution and how many are participating in each case:

	Institutions	Source of data
Scientific journals	2 institutions	An editor (1 participant) A journal expert (1 participant)
Publishers	2 institutions	A public University (1 participant) A private University (1 participant)
Legal depositing agencies	2 institutions	Biblioteca Nacional de Colombia Universidad Nacional de Colombia (4 participants)
National body of research	1 institution	Colciencias (2 participants)
National Archive	1 institution	Archivo General de la Nación (1 participant)

Table 2. Sampling strategy

2.3. Data collection techniques

The data for the case study was gathered trough multiple sources of evidence using in-depth interviews aimed to corroborate the same facts. The aim of collecting information in this way is to complement different opinions and to reduce biases (Miles and Hubberman, 1994).

2.3.1. In-depth interviews

We have selected in-depth interviews because it provides an excellent way to gather detailed information about the problem, the context and different elements regarding long-term preservation of Colombian scientific e-journals. As this specific topic has not been analyzed before in the scientific literature, it is worth to inquiry the stakeholders as much as possible, in order to have enough information for further analysis. Some of the advantages of in-depth interviews for qualitative research highlighted by Valles (1999) are that its "open" style allows to obtain richness information in the words and approaches from the interviewee; it provides to the researcher an opportunity for clarification and tracing of questions and answers, in an interactive framework more direct, personalized, flexible and spontaneous than the survey.

At the first stage of every study, in-depth interviews has the advantage of generating points of view, approaches, hypothesis and other orientations useful for further research. Others advantages of in-depth interviews are informative richness; inquiries about road maps; flexibility, agility, and saving; qualitative analysis of quantitative results; accessibility to information hard to observe; familiarity and comfort (Valles, 1999). An interview is an active listening, concentrating and focusing on the answers while assess the progress or the interview and you stay alert for cues about how to move the interview forward as necessary (Seidman, 2013).

Though qualitative case study design does not allow for a priori selection and design of individual data collection techniques and instruments (Pickard, 2013), I proposed the initial issue about “plans for digital preservation of e-journals” that guided the first interviews and then I were adding more categories of analysis with the following interviews.

For conducting each interview I followed the recommendations about the structure of in-depth phenomenological interviewing (Seidman, 2013). The first stage focused on life history, for example the current role in the organization, then I asked for details of experience and the

third stage I motivated the participants to express the opinions, and perception about the problem of digital preservation of Colombian scientific e-journals. All the interviews performed during the study had open-ended questions. One of the advantages of open-ended question is that they establishes the territory to be explored while allowing the participant to take any direction he or she wants (Seidman, 2013).

In-depth interviewing is not designed to test hypotheses, gather answers to questions, or corroborate opinions. Rather, it is designed to ask participants to reconstruct their experience and to explore their meaning. The questions most used in an in-depth interview follow from what the participant has said. The most important personal characteristic interviewers must have is a genuine interest in other people. (Seidman, 2013).

Some of the techniques proposed by Seidman (2013), and that I have followed during the interviews were:

- *The key to asking questions during in-depth interviewing is to let them follow, as much as possible, from what the participant is saying.*
- *It is in response to what participant says that the interviewers follow up, asks for clarification, seeks concrete details, and requests stories.*
- *Listen actively and to move the interview forward as much as possible by building on what the participant has begun to share.*
- *The interview question is cumulative; one interview establishes the context for the next.*
- *Ask questions when something are not understood (you do not understand). Ask for clarification if you do not understand the meaning of a word in a specific context.*
- *Ask for in-depth subjects, listen more, talk less, and ask real questions (one to which the interviewer does not already know or anticipate the response).*
- *Avoid leading questions (one that influences the direction the response will take)*

- *Follow up, but do not interrupt*
- *Ask participant to reconstruct not to remember what happened?*
- *Keep participants focused and ask for concrete details*
- *It is important to understand experiences in the context of time.*

The interviews were synchronous with some pre-elaborated questions. Each interview was designed in an informal semi structural way, avoiding close questions but motivating the participant to give all the information they want to share and that could be relevant to the research project.

The steps for conducting the interviews were:

1. To make initial contact with people from the selected institutions (Appendix 03)

2. To arrange a meeting by e-mail with each participant (Appendix 04). Once a participant agreed to be interviewed, I replied to each one enclosing the research project and the informed consent.

3. Once in the interview session, I explained the structure of the interview, the need of having the informed consent signed, and asked if they have questions before to start recording.

4. The interviews were recorded using the digital recorder SONY ICD-PX312F, then exported in MP3 format, and finally fully transcribed using NVivo for Mac (Version 11.0)

5. The transcriptions were copied and edited using Microsoft Word 2011 for Mac (Version 14.0) and then sent to each participant asking for modifications or corrections up to them (Appendix 05). The final versions of each interview were uploaded into NVivo, as internal sources of data. As result, there were a total of 64 pages interviews transcriptions, which means more than 24000 words of available data (Appendix 07).

2.3.2. Observation

Data was also gathered from a non-participant observation. According to Pickard (2013), this kind of observation requires the researcher to be as "interesting as wallpaper" and have nothing at all to do with the setting being observed. You watch, record, and later interpret what you have seen in the context of your research question.

The observations was done in short periods of time when performing informal conversations with editors and publishers from March to October, 2015. The observations were recorded in memos using the software Nvivo.

2.4. Analysis techniques

2.4.1. Grounded theory

I have selected Grounded theory as the strategy for handling data and analysis and it has influenced the design of the case study. Grounded theory is a general process of comparative analysis and a strategy for handling data in research, providing modes of conceptualization for describing and explaining (Glaser and Strauss, 1967).

Following this approach I have done the literature review after the analysis in order to facilitate the inductive method, to make it easier the coding process and to let the categories emerge from the data itself and not from assumptions coming from the literature. In pure Grounded theory the literature review very often comes after the analysis of empirical data (Charmaz, 2006).

There are two approaches to grounded theory research: Strauss offering the structural approach to analysis through the constant comparative technique and Glaser encouraging the less structured approach of theoretical sensitivity (Pickard, 2013). I have chosen the Strauss approach, as it provides a structured strategy for handling the data and analysis.

According to Strauss & Corbin (1998), Grounded Theory:

> "*...derives its theoretical underpinnings from Pragmatism and Symbolic Interactionism (...) and two important principles drawn from them are built into it. The first principle pertains to change. Since phenomena are not conceived of as static but as continually changing in response to evolving conditions, an important component of the method is to build change, through process, into the method. The second principle pertains to a clear stand on the issue of "determinism." Strict determinism is rejected, as is nondeterminism. Actors are seen as having, though not always utilizing, the means of controlling their destinies by their responses to conditions. They are able to make choices according to their perceptions, which are often accurate, about the options they encounter. Both Pragmatism and Symbolic Interactionism share this stance...*"

I have followed the above perspective for analyzing the data and for constructing the concepts emerging from the analysis. I have also applied the Grounded theory principles highlighted by Pickard (2013): the research is focused in discovery and analysis of the context; data collection and analysis have done simultaneously; the categories were constructed from empirical data; the development of theory has been interpretivist and has emerged after continuous written commentaries on the data or memos.

2.4.2. Constant comparative analysis

Constant comparative analysis involves taking one piece of data and comparing it with all others that may be similar or different in order to develop conceptualizations of the possible relations between various pieces of data; it also demands that that the creation of categories is driven by the raw data and not established a priori, although it is inevitable that prior research will have identified some of the salient issues (Pickard, 2013).

I have followed the ideas and procedures for coding included in the book *Basics of qualitative research: Techniques and procedures for developing grounded theory* (Strauss & Corbin, 1998). The microanalysis of data was particularly useful at the first stage of open coding. As Strauss & Corbin (1998) states:

> *Microanalysis is a form of coding that is open, detailed and exploratory. It is designed to focus on certain pieces of data and to explore their meaning in greater depth and develop concepts in terms of their properties and dimensions. Microanalysis is most likely to be used in the early exploratory stages of analysis when an analyst is trying to get some sense of the meaning of data and to find concepts that reflect that meaning.*

Along with the microanalysis I have followed some of their suggested strategies for open, axial and selective coding, including a diary of analysis, memos and diagrams. The software Nvivo was extremely useful not only for recording the emerged categories with their dimensions and properties, but also for identifying relations and writing memos. I have also kept in mind that analysis is a process that throughout the research, researchers are constantly updating and revising concepts, adding new concepts, identifying new properties and dimensions, and seeing new relationships between concepts (Strauss & Corbin, 1998).

2.4.2.1. Open coding

The first step with the analysis of the initial data gathered trough interviews was to start doing a microanalysis line by line. With this technique, some initial categories were emerging from the data. At the beginning these categories were vague and with high level of uncertainty. As soon I was progressing with the analysis I found some categories that became very relevant to the topic. The result of this process was a list of 53 categories, 183 text references and 4 memos about my first reflections about my understanding of the problem and the initial analysis about the core emerged categories. All of this work was done using the software Nvivo for Mac (See image 1).

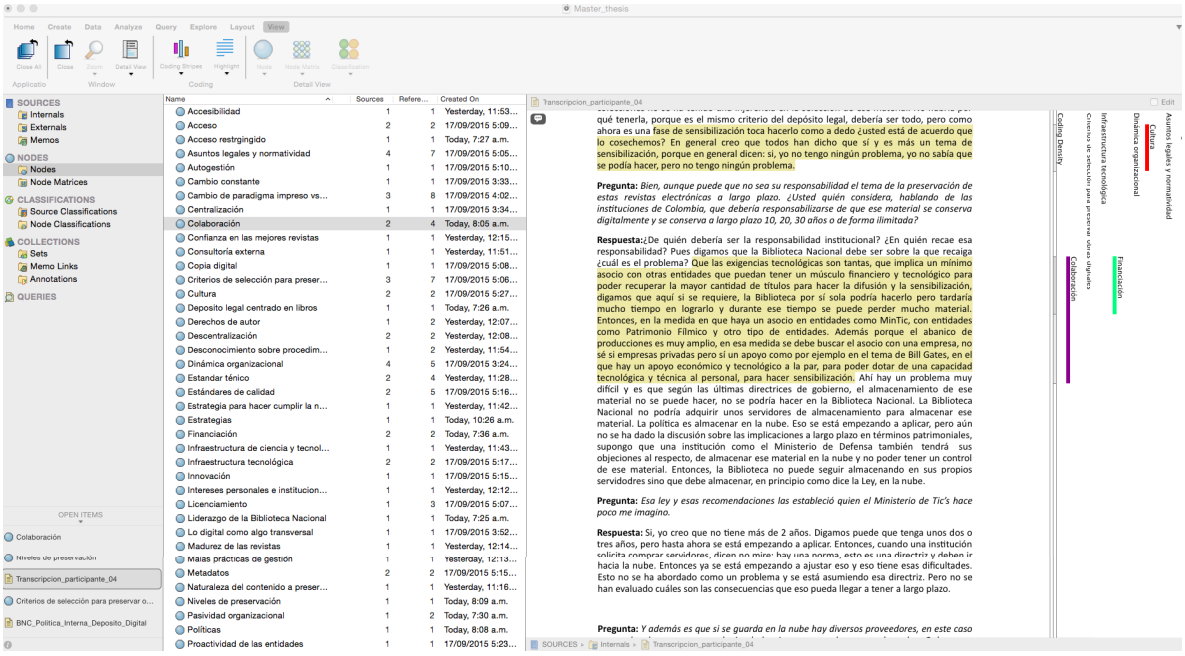

Image 1. Open coding: screenshot of NVivo for Mac

2.4.2.2. Axial coding

The axial coding is the process of relating categories to their subcategories, termed *axial* because coding occurs around the axis of a category, linking categories at the level of properties and dimensions (Strauss & Corbin, 1998).

The axial coding was much about reviewing the data and classification of categories. I focused on the data and then created some categories. Some of them were based in the categories identified in the open coding and others were new. As far as the analysis was deeper I started to clarify the dimensions and properties of each category and subcategory, finding the relation between the categories and also ridding of some of the initial categories. I performed some queries in NVivo with keywords that help to me to identify other possible hidden categories or relations (Appendix 06). I also wrote more memos in Nvivo about each category. The result of this process was a refinement of the categories and 15 memos (See image 2).

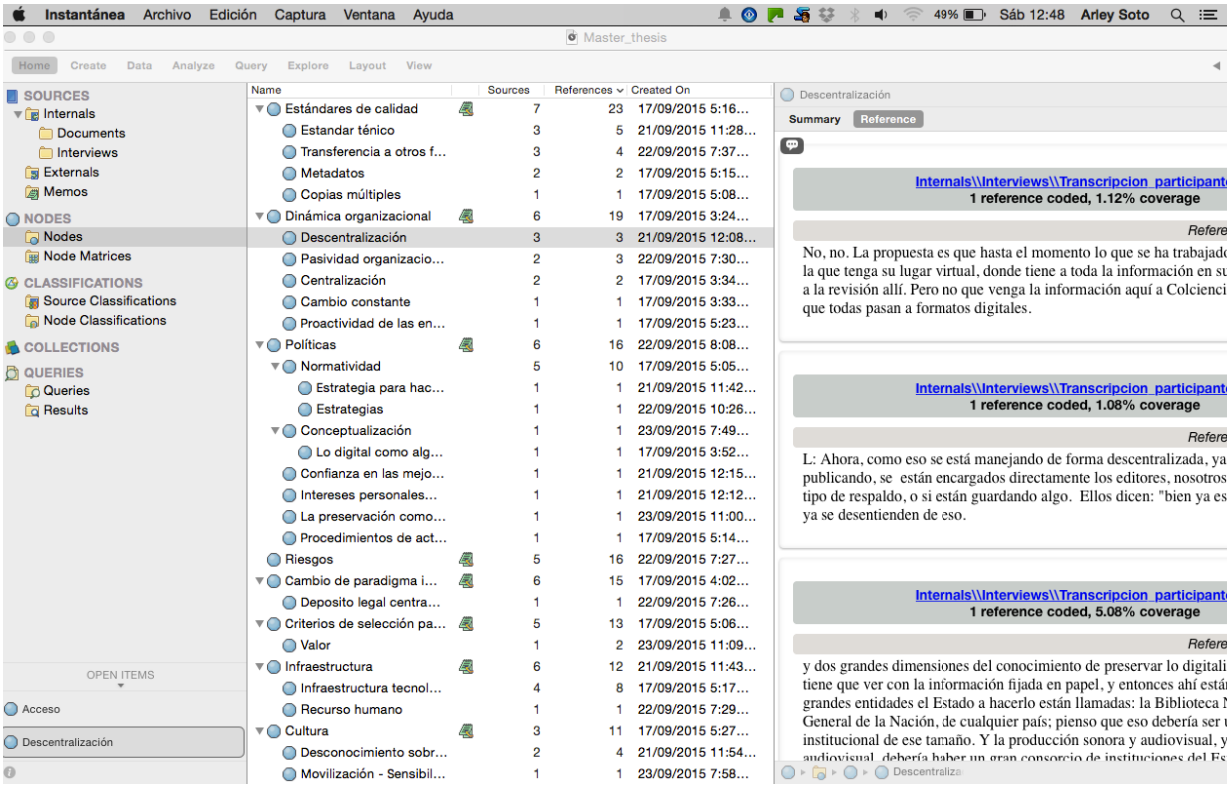

Image 2. Axial coding: screenshot of NVivo for Mac

2.4.2.3. Selective coding

Under the grounded theory approach for analyzing the data, the objective of selecting coding is to identify a central category and its relations with the other categories. At this stage, I could not use Nvivo for other thing than to see and review my categories and memos, instead I used a blank piece of paper and a pencil to organize my ideas and to draft how the categories were related to each other. I draw many diagrams and finally I was comfortable with one (See image 3); then I clarified the core category writing a narrative understanding about the emerging concepts. The description of the core category is described in chapter 4.

CHAPTER 3. LITERATURE REVIEW

Although the publishing business models have changed dramatically due to Internet and Open Access Policies, academic journal publisher role has remained unchanged: registering and date stamping new research findings across the disciplines; ensuring the highest quality through a rigorous system of peer review; disseminating material as broadly as possible; and creating a permanent archive as a legacy for future generations (Morgan, Campbell & Teleen, 2012).

Scientific communication takes place between researchers, mostly acting in some sort of capacity as authors and readers. The authors want priority of discovery in order to be read, but they are also heavy users of other´s creations (Roosendaal & Geurts, 1997). Publishers and libraries provide a bridge between them. Publishers add value to authors´ creations and Libraries add value to readers. Although some publishers act as information providers and some libraries support publishing, both roles are key actors of scientific communication. In this information space there are five main functions that guarantee an effective scholarly communication process: registration, awareness, certification, archive and rewarding.

Registration is the function to declare precedence for a scholarly finding. It includes what is the claim about, who does the claim, where and when. *Awareness* allows actors in the scholarly system to remain aware of new claims and findings. *Certification* establishes the validity of a registered scholarly claim. The peer-review process, conducted under the auspices of the journal publisher certifies the claims made in the manuscript. The *Archiving* function preserves

the scholarly record over time. *Rewarding* was traditionally based on the mere fact of publishing in a certain class of journals and on being referenced in articles by other scholars (Van de Sompel et. al., 2004).

The digital preservation of e-journals is under the Archiving function. It is interesting to note that in characterizing the notion of "digital preservation," we speak or write about ensuring "continuing access to digital collections." we acknowledge that, with reference to digital technology, preservation and access are fused, because preservation becomes the ability over the long term to retrieve and reproduce digital information (Besser, 2007).

For the purpose of the research, we considered digital preservation of e-journals as long-term preservation, which means continued access to the digital content published by e-journals indefinitely. Digital preservation is defined as all of the actions required to maintain access to digital materials beyond the limits of media failure or technological change. Those materials may be records created during the day-to-day business of an organization; "born-digital" materials created for a specific purpose (e.g. teaching resources); or the products of digitization projects (Jones & Beagrie, 2001).

3.1. Digital preservation standards

There are two international standards that underline the requirements for digital preservation systems: OAIS and PREMIS. The Open Archival Information Systems (OAIS), published as international standard ISO 14721, is the most common framework for digital preservation systems and it has proven to be a very useful high-level reference model, describing functional entities and the exchange of information between them (Becker, et.al., 2009). On the other hand, the PREMIS Data Dictionary for Preservation Metadata is the international standard for metadata to support the preservation of digital objects and ensure their long-term usability.

Developed by an international team of experts, PREMIS is implemented in digital preservation projects around the world, and support for PREMIS is incorporated into a number of commercial and open-source digital preservation tools and systems (Library of Congress, 2015).

3.2. Technical methods of digital preservation

Migration and emulation are the two primary strategies used for long-term preservation (Kirchhoff, 2008). Migration involves transforming digital content from its existing format to a different format that is usable and accessible on the technology in current use. Migration requires the repeated copying or conversion of digital objects from one technology to a more stable or current, be it hardware or software. Each migration incurs certain risks and preserves only a certain fraction of the characteristics of a digital object (Becker, et.al., 2009). Emulation operates on the environment of an object, trying to simulate the original environment that the object needs (Becker, et.al., 2009) and involves developing software that imitates earlier hardware and software (Kirchhoff, 2008). However, emulation is technically complex to achieve and hard to scale up to large amounts of data; specific characteristics of an object may be lost due to incomplete or faulty emulation, or due to the impossibility of emulating certain aspects. Furthermore, users may have difficulties in using old software environments, and some functionality of newer systems might not be available when relying on the original environment of an object (Becker, et.al., 2009).

3.3. Digital preservation policy frameworks

There are two commons frameworks for building digital preservation policies: ERPANET and JISC approach. ERPANET's Digital Preservation Policy Tool provides a framework to address the benefits of digital preservation, the scope and objectives of the policy, as well as requirements, roles and responsibilities, context, areas of coverage, costs, monitoring and review, and implementation of the policy; on the other hand the JISC's Digital Preservation Policies Study provides a matrix for policy clauses that include a principle statement, contextual links, preservation objectives, identification of content, procedural accountability, guidance and implementation, a glossary, and version control (Noonan, 2014). In the following table we presented a summary of the comparison, made by Noonan, between Erpanet, JISC and Ohio State University Libraries (OSUL) policy frameworks:

	OSUL Review	Erpanet	JISC
Introduction or purpose	✓	✓	✓
Mandate	✓	✓	✓
Objectives	✓	✓	✓
Scope	✓	✓	✓
Requirements	✓	✓	
Challenges	✓	✓	
Principles	✓	✓	
Roles and responsibilities	✓	✓	✓
Collaboration	✓	✓	
Selection and acquisition	✓	✓	✓
Access and use	✓	✓	
Statement of financial commitment		✓	
Training and education		✓	
References	✓		
Glossary	✓		✓
Review cycle	✓	✓	✓
Implementation	✓	✓	✓

Table 3. Components of policy models

3.4. Digital preservation planning tools

The most common digital preservation tools are the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA); the Preservation Planning approach (PLANETS) and the Trusted Digital Repositories and Audit Checklist (TRAC), the latter became now an international standard (ISO 16363), suitable for certification.

DRAMBORA is a method that supports the self-assessment of a digital repository by identifying assets, activities and associated risks in a structured way. It adapts standard risk assessment principles and tailors them to digital repository assessment. It can be applied to analyze and verify the risks that apply to preservation planning activities within an organization and can thus support the ongoing improvement and implementation within an organization (Becker, et. al., 2009)

PLANETS Preservation Planning approach is another tool for supporting preservation planning activities. It provides an approved way to make informed and accountable decisions on which solution to implement in order to optimally preserve digital objects for a given purpose. It therefore offers a standardized way of planning and evaluating preservation strategies based on a set of experiments (Strodl, et. al., 2007). The Planets Planning method implements the Develop Preservation Strategies and Standards and the Develop Packaging Designs and Migration Plans functions of the OAIS model (Becker, et. al., 2009).

The international standard ISO 16363:2012 "Audit and certification of trustworthy digital repositories" defines a recommended practice for assessing the trustworthiness of digital repositories. It is applicable to the entire range of digital repositories and can be used as a basis for certification. This standard was based upon the Trusted Digital Repositories and Audit Checklist (TRAC), a set of digital preservation best practice criteria that can be used to evaluate repositories, published originally on 2007, by the Online Computer Library Center

(OCLC) and the Center for Research Libraries (CRL). It is also well known the Network of Expertise in Long Term Storage of Digital Resources (NESTOR), a German catalogue of Criteria for Trusted Digital Repositories—a checklist similar to TRAC to assess the technical and organizational trustworthiness of a digital repository (Kirchhoff, 2008). TRAC categories are Organizational infrastructure; Digital Object Management and Technologies, Technical Infrastructure, Security.

3.5. Digital preservation solutions for scientific e-journals

The Keepers Registry (http://thekeepers.org/) is the product of the PEPRS project, which aimed to Pilot an E-journal Preservation Registry Service (PEPRS). The purpose is to give an account of an international endeavor to monitor the success of arrangements for long-term access to the world's serials literature. The Registry is a global facility that enables all concerned to discover what serial content is being preserved for the long term (Burnhill, 2013). According to Burnhill (2013), both libraries and publishers need to prompt archiving organizations to be alert to what is missing and summarized the following propositions:

> 1. Assign an identifier at the 'point of issue' for a stream of digital content. • If it is worth preserving for the long term then it should have an identifier.
> 2. Ensure that (digital) content is archived routinely. • Have others/peers do that for you too; lots of copies keep stuff safe(r).
> 3. Tell someone what you are doing (and how) and what you hold. • So all know what has been preserved and what is still at risk of loss.
> 4. Publish the terms of access for the archived content (now and when triggered as orphaned). • The purpose of preservation today is assurance of access tomorrow.

Currently, there are 10 archiving agencies running a program for the long-term archiving of e-journals and actives in the Keepers Registry: Library of Congress, Scholars Portal, Archaeology Data Service, British Library, HathiTrust, CLOCKSS Archive, Global LOCKSS Network, Portico, National Science Library, Chinese Academy of Sciences and e-Depot (National Library

of the Netherlands). Some of these initiatives are solutions for specific regions and the only two that could be used by Colombian scientific e-journals for now, which is CLOCKSS or PORTICO, because they are open for all kind of journals and provide an easy way to participate in it.

One of the most well known tools for preserving scholarly e-journals is LOCKSS (Lots of Copies Keep the Stuff Safe). LOCKSS collects and preserves all content in its original format, as delivered from the publisher, including the format metadata that enables a browser to render the content. Formats that are collected and preserved include: spreadsheets, XML, HTML, PDF, video, and sound. LOCKSS defines "obsolete content" as "when a reader's web browser does not display the content." The reader's browser determines this based on the preserved format metadata, LOCKSS detects it and invokes an "on the fly" process that creates an access copy by migrating the preserved content and format metadata so that it displays properly in the reader's web browser (2CUL LOCKSS, 2011). Software LOCKSS is available as open-source software, and it could be used as a Private LOCKSS Network (PLN). Recently, the Public Knowledge Project launched the PKP-PLN available for journals using Open Journal Systems (OJS).

LOCKSS also has the global LOCKSS network, a web-based electronic subscription, including e- journals, to support post cancellation access. Content preserved by libraries through LOCKSS becomes a part of their collections, and they have perpetual access to 100% of the titles preserved in their LOCKSS box. The box collects content from target web sites using a web crawler similar to those used by search engines and continually compares the content it has collected with the same content collected by other LOCKSS boxes (2CUL LOCKSS, 2011).

On the other hand, Controlled LOCKSS (CLOCKSS) is a kind of consortia based on the LOCKSS technology, with the aim of providing a long-term global archiving solution that will serve the joint library and publisher communities in the event of a long-term business interruption or in making orphaned or abandoned works readily available to the scholarly community (2CUL LOCKSS, 2011).

The other solution available for Colombian scientific e-journals is PORTICO, a centrally administered archive with content stored in multiple locations. Portico is a digital preservation service provided by ITHAKA, a not-for-profit organization with a mission to help the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways (Gaur & Tripathi, 2012). Portico ingests data files in whatever format the publisher uses. These could be database files, XML, HTML, or files in an unspecified proprietary format. Then Portico normalizes the files to a standard archival format, which it can subsequently manage over time. Portico's primary preservation methodology is migration, which involves transitioning content from one file format to another as technology evolves and file formats become obsolete.

However, these initiatives archives only a small portion of the total e-journals available worldwide. The study performed by Seadle (2011) shows that only 8% of the DOAJ titles are in LOCKSS/CLOCKSS and only 5% in Portico. Small and financially poor publishers do not fit the business model for Portico (Seadle, 2011) and the same apply to CLOCKSS where publishers have also to pay an annual membership and file ingesting charges.

CHAPTER 4. RESEARCH FINDINGS

The main finding of the research is that digital preservation of scientific e-journals is a multilevel practice, its main objective is to reduce the risk of losing digital content and the aim is to provide long-term public access to the users (See image 3). Digital preservation is influenced by the cultural and organizational change, because people involved in the process is living a transition from printed to digital-born issues. Digital preservation is part of long-term planning at several levels; it requires a synergy between the actors involved, including policies at journal, publishers, research and national level. All of these policies require meeting technical standards and to solve copyright issues to became the preservation feasible. The policies should include not only roles and responsibilities but also the strategies and infrastructure required. All of these concepts and relations will be explored and described in the following paragraphs.

The objective of preservation is to give guarantee of long-term permanent access to the content published by e-journals, considering technological advances and reducing possible risks of losing information. As the main objective is to "bring access", any solution should consider not only assuring that information remains over time but also that information is public and accessible to the people.

Digital preservation of Colombian scientific e-journals is a:

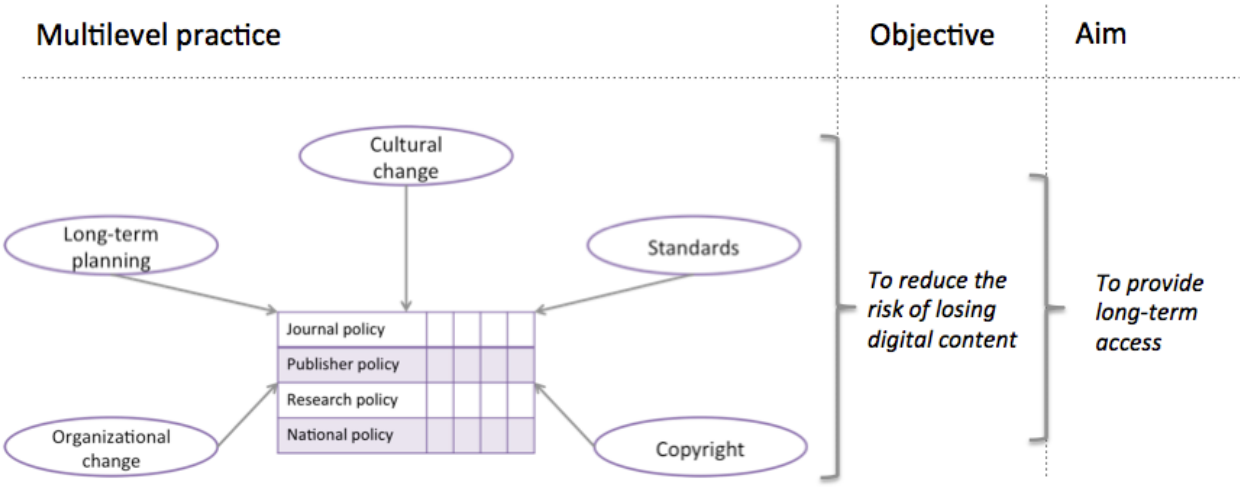

Image 3. Selective coding: central category diagram

4.1. Digital preservation of e-journals as a multi-level practice

Digital preservation of e-journals is a multi-level practice. That means that assuring long-term access depends on policies and actions at journal, publisher, research and national level. In Colombia the journal level corresponds to the Editor and Editorial team; at publisher level we talk about institutions like Universities and Associations; the research level is the position of national body of research (Colciencias), and finally in the national level is the National Library of Colombia as the major legal deposit agency. The roles and responsibilities are summarized in table 3.

Policy	Institution	Roles and responsibilities related to digital preservation of scientific e-journals
Journal level	Editor	To publish content under international standards and to have the legal requirement to preserve the submitted content To ask an authorization from authors to preserve and transform digital content To transfer digital copies to a third party for long term access To adopt best practices for preserving digital content.
Publisher level	Universities & Associations	To have an institutional policy about scientific publishing To have an institutional policy about digital preservation To have a long-term plan for scientific journals To guarantee long-term funding for journals.
Research level	Colciencias	To provide financing and support for the National Library of Colombia to create a national solution for digital preservation of scientific e-journals To contribute for digital preservation capability building at all levels To include long-term digital preservation as one quality criteria for scientific journals To recommend to journals and publishers to deposit copy of the materials in a third party organization, at national (ideally) or international level.
National level	National Library of Colombia	To offer a national solution for preserving and granting long-term access to digital content published by scientific e-journals To maintain and develop new features for the long-term digital preservation solutions To offer technical and copyright guidelines to journals to guarantee long-term preservation To produce recommendations to journals about digital preservation. To promote cultural change about transition to digital formats To promote collaboration between institutions To have a long-term plan for digital preservation, including financing sources

Table 4. Multilevel practice of digital preservation of Colombian scientific e-journals

4.1.1. Role of Editors

Editors should publish digital content under international standards. A participant from one legal depositing agency has recognized its importance: "Something that we have realized in the library is that meeting standards make extremely easy the process of preservation. For the specific case of journals that have a system designed for online journals, such as the Open Journal Systems (OJS), the process is much easier than in those using a content management system (CMS) that is not specialized for journals". Another participant have added "we can assure that the storage (of the journal) data is under minimum standards that allows, sometime, to migrate this information to other formats that could be readable by other software, this is going to happened".

In contrast, there are some journals that are not applying and are not aware of the role of standards for long-term preservation and access and this has a direct relation with the quality classification of the journal. In general terms, top quality journals (for instance, journals classified in category A1 or A2 by Publindex) are better prepared for meeting these standards than others that are in the low categories, or unclassified. A participant from the national body of research have underlined that "... from the 542 journals, there are some that are playing with international standards, others not". Another participant suggested, "I dare to say that the most important thing (for being able to preserve) is to fulfill standards (…) I think the universities are doing so in a very rudimentary way, without meeting standards".

Editors are the first responsibles of applying international standards and good practices for e-journals digital preservation. For doing that, they have to include the digital preservation as an issue of the long-term planning, but editors faced some obstacles for implementing it. Firstly, they do not have enough time to invest in the journal management. Editors in Colombia are mainly faculty staff that has to share the role with other activities of teaching and research. They usually do not get paid for their editorial time, though some universities are starting to assign hours where they can spend in the journal. This is relatively new and it does not happened in all institutions.

However, it is the hand of the Editors where the digital preservation starts. They have the responsibility to prepare long-term plans for the journal, including how the journal will be accessible in the future, even if the journal or the institutions closed. As editors of a journal changed over time, they have also to assure that their successors know its responsibility, and take the required actions for giving the guarantee that the content will be preserved forever. The editors are focused on the editorial process of peer reviewing and they do not make so much in terms of journal management and long-term planning.

Another role that is critical for editors is to propose and use a license that allows to third parties to make a digital copy or transformation of formats for preservation purposes. The National Library has started the digital deposit only with journals that have this kind of licenses.

4.1.2. Role of Universities and Associations

Nevertheless journals are not a property of the editors, they are from the institutions. Even if a journal was an individual initiative, it remains as a journal of certain institution, usually a University, but just a few acts as a publisher. I mean, they appear as the publisher of the journals, but not all have the organization, policies and infrastructure required for being a journal publisher. Actually this is another obstacle for long-term preservation, because if the institution do not provide continuous support for journals it is difficult to guarantee long-term planning, including digital preservation. A participant expressed their perception of the problem in the following words: "…finally Vice-Chancellors and Deans are the people who take the decisions in Universities, and if there is not an alignment with the people who are in the head, journals could lose a lot of the effort done. We (the editors) know the needs and resources required but sometimes people who sign is not attuned with this…"

The journal expert interviewed said that "most of Colombian journals do not born from associations but from universities, this produces an organizational effect which means that almost each department has its own journal and this practice do not promote the identity of a discipline". However, apart from the problems that could cause this to the discipline and to the creation of science communities, in terms of digital preservation the problem is that not all departments manage the journals in the same way, for example, some journals at the same university could retain some rights for doing future transformation of formats for digital preservation purposes, others not. A solution for digital preservation should consider not only

the publisher level but also at journal level because not all publishers have actual control over the journals. A publisher should have clear digital preservation and copyright policies.

It seems like we do have Colombian high quality journals despite we do not have strong publisher organizations of scientific journals. Apparently publishers do not have even a clear institutional policy about scientific publishing, science dissemination or digital preservation. They follow the policies stated by the national body of research (Colciencias), but do not all have their own institutional policy that assume a particular position about it. Institutions playing the role of publishers do not have long-term plans for scientific journals that include continuous and permanent financing and long-term preservation. This is another cause of poor digital preservation practices of scientific e-journals.

4.1.3. Role of National Library of Colombia

Despite journals and publishers have a big responsibility about long-term planning and preservation, it is in the National Library of Colombia where people hope this content will be available in the future. All of interviewees agreed that the National Library should be doing something to provide a solution for long-term preservation of both printed and e-journals. They do not mention the others two legal depositing agencies, so the leadership of the National Library is clear and remarkable. This generalized assumption give to the library an "ipso-facto" responsibility even if there is not an explicit law to make the digital deposits on it.

At the same time, the National Library of Colombia is aware of its responsibility and has implemented a voluntary on-line deposit for digital works, including e-journals. They started a pilot project using the tool archive-it.org for archiving the web of journals from public universities. However archive-it.org is not recognized as a valid archiving agency for e-journals by the Keepers registry (http://thekeepers.org/), so that it could be a partial solution for the National Library needs but it is not for scientific e-journals.

The pilot project using archive-it.org or a cloud-based service involve certain risks or obstacles, some of them recognized by the participants, for instance "It has not discussed yet the consequences of having that content in the cloud, mainly because this digital content is part of the cultural heritage of Colombia, and we would not have control over it". Another participant said: "It does not matter where the content is archived, in the library or outside, the important thing is to have total control and to give guarantee of long-term preservation. The cloud solutions do not assure long-term preservation activities, such as migration, monitoring or distributed redundancy".

Apparently the risk is in both sides. First, the risk of having the content in a third party such as the cloud, is that the library would not have control over whatever that happens with this third party, for example technical incidents, legal issues or agreement limitations. On the other side, the risk of having the content in the National Library is that the Library does not have the technological infrastructure to support and to give the guarantee that the content will be safe, reliable and accessible over time. Meanwhile, journals are depositing their content voluntarily and the content will be increasing each day. So, what the library should do is to reduce the risk in both ways, first, diminishing the dependency of third parties and at the same time increasing its capacity and improving their digital infrastructure to provide long-term preservation and access. It implies having the digital content preserved in distributed servers of different institutions at separate locations with high digital preservation standards such as OAIS and PREMIS. But to achieve this, there are some barriers to reduce: lack of financing and poor collaboration.

The National Library does not have enough budgets to invest in the infrastructure required for digital preservation, not only for e-journals but overall. One of the participants at the National Library of Colombia has said: "These projects are expensive, so it should be a priority to increase the budget of the library considerably. In my opinion institutions with the dimensions

of a nation, should have budgets with the dimensions of a nation, to include needs and complexities as dense as the preservation of our written and audiovisual memory". It is clear that to create the suit of solutions for providing long-term digital preservation, it is absolutely necessary to invest resources according to the scope and expected results of each preservation project, not only for the implementation but also for its maintenance and continuous development.

Others responsibilities and opportunities for the National Library of Colombia are to develop an institutional policy and long-term plan for digital preservation, for doing that it could use some of the tools available, for example ERPANET's Digital Preservation Policy Tool or the JISC's Digital Preservation Policies for Policies, and DRAMBORA, PLANETS and ISO 16363 for planning. The Library also has the responsibility to offer a national solution for preserving and granting long-term access to digital content published by scientific e-journals; to maintain and develop new features for the long-term digital preservation solutions; to offer technical and copyright guidelines to journals to guarantee long-term preservation and public access; and to promote cultural change for facing the challenges of the transition to digital formats. In addition, it is important to consider that for improving the capability of the National Library, they need to improve its infrastructure, including hardware, software, financing, methodologies and the implementation of international standards in their practices.

4.1.4. Role of Colciencias

Colciencias is the national body of research of Colombia and between others, has the responsibility of building the Colombian capability for science, technology and innovation as well as to promote the production, circulation and use of knowledge. It also defines the strategic programs for visibility, use and appropriation of the knowledge created by Colombian community of researchers and innovators. In terms of scientific journals, their main strategy is

the National Index of Scientific Journals, known as Publindex. This index is an evaluation service for journals based on a set of quality criteria that defines if a journal is classified in one of the four categories established: A1, A2, B, C. Journals categorized as "A" are the top quality journals and those in "C" meet the minimum criteria. Journals "B" are in the middle. Colciencias is the leader institution in terms of Science, Technology and Innovation promotion in Colombia and they have a strong level of influence over the scientific journals and its publishers.

In terms of digital preservation of e-journals Colciencias do not have an explicit role or responsibility, neither a plan nor a program for running an archiving solution. However, as they develop the national policy for scientific journals, they could play a strong role to promote digital preservation in both ways: firstly, funding a national project for developing a suit of solutions for digital preservation of scientific journals and secondly to adding digital preservation as one of the quality criteria for scientific journals in Publindex. The logic of this strategy is that they should not include archiving as another criteria until the creation of the national archiving solution for scientific e-journals. The solution should include not only the technological infrastructure at the National Library of Colombia, with several nodes of digital redundancy but also recommendations and technical guides to journals and publishers for licensing and submitting content in that solution.

4.2. Transition from printed to digital-born journals

One of main obstacles identified in all the levels described in the previous section for long-term digital preservation is the cultural barrier. So many people are still in the paper paradigm and they do not understand the implications of the digital content. Some expressions that participants have used to described this transition are “Most of the journals are paper-based. The last five years journals are transforming its content to digital format”; “…many journals are migrating to digital, and mainly from universities, in fact, some of them have reduced the

printed copies to a minimum or has converted in only digital journals without a printed version". All participants mention this transition, and we found experiences in two ways. For one hand, there are initiatives carefully designed and planned, considering many of the implications of the digital environment before to start and throughout the project. On the other hand, there are experiences that seems like the context have made pressure over people to migrate to the digital without a strong understanding and planning about their implications.

Some of the people making the decisions have said "… there is some fear about the digital, we do not know what is going to happened, the digital is still something novelty…". It seems like there are still a focus on the problems that the printed material bring to us, and they are not realizing that the amount of digital information is huge and that we need to start to think and implement solutions for future access of this content. There are some organizations that are aware of this challenge, for example a participant has said: "… we are changing to be more prepared for dynamics of the digital issues, because we were focused on the printed material and the digital was laggard".

Journals are still migrating from printed to digital-born journals in a very slightly way, they feel like they are in an unexplored land giving slow steps to test the new digital environment and then taking the next steps carefully; almost everything is empirically, without a rigorous process of planning. They could have the digital target clear, but instead of drawing a map and watch the path with all the conditions and the difficulties they will find, they are already on their way and responding to all difficulties as they are coming up, without considering the budget needed, copyright implications, dynamic of organizations, between others. All of these factors prevent editors for acting faster and to be more prepared for the digital challenge.

Some editors have underestimate the importance of the digital version of the journal, for example they would not tolerate a typographic error in the printed version like a wrong

comma, a capital letter or other wrong character, but once the printed version has finished, they do not care so much about similar problems in the digital form, for example typographic errors in the metadata. They still think that the printed journal is an object that finishes when printing but do not recognize that there are new requirements for the digital environments that demands too much work after the issue is ready than it was when only printed.

The transition from printed to digital has brought also some misunderstandings about economic issues. Some people though that publishing in digital is cheaper than traditional printing, but this is a fallacy. The costs of producing a digital journal with high standards, could be as expensive as publishing in paper, even more expensive if you consider include multimedia objects such as video, sound or web interaction. A publisher has recognized this using the following words: "We have realized that digital publishing is expensive. What happened is that there is a popular belief that everything that is published in Internet is cheaper than the printed". This is another obstacle for digital preservation, because people who makes the decisions could assume that producing an e-journal is cheaper and do not allocate enough budget for the new requirements of the digital environment.

This transition implies also changes with the traditional way of managing copyright issues. For example, the authors should send the works for the journals with a license to allow the journal or the publisher to make copies of the work in a third party or to do transformations of the content for digital preservation purposes. This could make digital preservation easier and faster. We have found situations where a journal does not have a unique license; it has changed along the time. Journals could have asked authors to give the journal all the patrimonial rights, then it could have changed to an authorization to publish in printed and finally they could have a license to retain the rights on the authors, but allowing to makes copies for digital preservation. In this case, we can not treat all works of the journals in the same way, we should review under which conditions each article were submitted and check if there is any restriction to put a copy

in a third party digital archive. National Library of Colombia cannot give access to the content preserved without a license or authorization. So if the journals want the library to offer access if a trigger event happens the publishers first need to offer a proper authorization to the Library.

Lack of acknowledgement about the implications of digital preservation is part of the problem. Some editors and publishers are not aware about solutions available, so for this reason they do not use this kind of solutions. Some of them do know the solutions but they do not understand its purpose. They have never been worried about long-term access, they are worried about the following issue or about current and day by day problems, and they do not have time to worry about what would happened in 20 or 30 years. Journals follow the recommendations of Colciencias but as the digital preservation it is not part of the criteria, they are not aware of it.

4.3. Digital preservation as a long-term management problem

Another reason why digital preservation is not a common practice of e-journals in Colombia is because journals do not have long-term management plans and publishers do not have digital preservation or scientific publishing policies. One of the participants has recognized this as a possible cause that a journal have ceased its activities: "When a journals loses its north, without options for re-discovery or re-thinking and without looking for alternative for continuing improvement, there is management problem". So we have a two ways problem: first, if the e-journal closes, there is a considerable risk of losing the content published during its history, if no body is taking care of their content. On the other side, if a journal did not have long-term planning they probably do not know how those content will be preserved, available and accessible in the future, after the journal or the publisher decide do not publish more issues. In addition, not all publishers have digital preservation policies.

But the management problem is not only from journals and publishers. Institutions at all mentioned levels have some managerial limitations; one is their instability in terms of administration and organizations. Organizations are dynamic, and it is affecting the digital preservation plans and programs. It has been identified that dramatic changes in organizations are critical for digital preservation. For some of the participants the reorganization of the institutions influences how to make decisions and how to establish priorities. For instance, meanwhile for a certain lapse of time the digital preservation is a relevant and urgent issue, for others not. Some of these priorities depend on the vision of the people who is temporary as the institution manager. A digital preservation practice could be affected if there is not a permanent alignment with the organizational strategic objectives of all levels involved.

Sometimes, institutions have good intentions but they could be just initiatives that never transformed as a project. Some of them just finish as a diagnostic, but nothing more. Actions and plans included as recommendations not always becomes a reality, because they need financial resources that are difficult to get. On the other hand, some institutions are worried about the short-term results and not about long-term issues of digital preservation.

Poor collaboration is another of the issues identified. Even if the nation increases the budget for the National Library of Colombia, never will be enough for all the projects required to assure the digital preservation of the Colombian heritage. Collaboration between organizations is important not only for getting more monetary resources, but also for understanding the problem of each sector and providing solutions that satisfies the needs of the stakeholders. A participant has said this in a very clear way: "The technological needs are so many that implies a minimum association with others entities that can have the financing muscle to preserve the broader quantity of titles possible… the library of its own could do that but it would take so long and in the mean time a lot of material could be missed." Others participants have agreed on the importance of this collaboration with opinions like "… to make an inter-institutional

board and let a proposal to the Congress…" or "… it is possible if other organizations like the ICT Ministry or other kind of entities are involved". Collaboration is one of the challenges of the digital preservation. One strategy that appears in all the interviews is that the solution is not a matter of just one institution. The solution should come from the collaboration between several organizations, not only public organizations but also private companies, even international cooperation. Moreover it is not enough one project of collaboration, it is necessary to have several projects with multiples partners, as the stakeholders change depending the nature of the content wishing to preserve.

However, working together is not an easy path to go. In order to make a collaborative project, it is important to get over institutional egos and jealousy. A participant has said: "We have enormous difficulties for teamwork between institutions, and it cannot happened (the digital preservation) if we do not work as a team... we need a transversal, collaborative, participative and corporate vision". It should be any way to be over personal and institutional interests and also overtake the administrative barriers that comes when several organizations from different backgrounds and natures want to work in the same project. It seems that collaboration is complicated because the lack of a leader organization that is capable to articulate and connect the work between different actors.

Collaboration should be done depending on the format and the nature of the content. Under this perspective, it is more likely to find partners and participants where the common work and results provide a direct solution of the problems faced by each institution. This is a way of generating value of a collaboration project.

CHAPTER 5. CONCLUSIONS AND IMPLICATIONS

Some Colombian scientific e-Journals are at risk of losing its digital content because they do not have plans or strategies for long-term digital preservation. This is a substantial risk not only for digital-born journals but also for journals living the transition from printed to completely digital.

Digital preservation is not a common practice of scientific e-journals, because publishers do not have institutional policies about digital preservation neither for scientific publishing and because the institutions involved are not assuming all the expected responsibilities for assuring long-term preservation and perpetual access to Colombian scientific e-journals.

The mains obstacles for providing digital preservation solutions for Colombian scientific electronic journals are lack of financing at all the levels involved; the absence of a strong publisher role that could support policies at high level; content licensing that hinder the preservation activities; low digital preservation infrastructure in Colombian depositing agencies; lack of acknowledgement about the digital preservation solutions available for scientific e-journals; a lack of understanding about what is digital preservation and its implications; poor long-term management plans for scientific e-journals and for digital preservation; lack of collaboration between institutions; low adoption of international standards and best practices; resistance to change both organizational and cultural.

Digital preservation of scientific e-journals is a multilevel practice, it means that to make the preservation feasible it requires actions and a synergy between the four stakeholders identified:

Editors, Publishers, Colciencias and National Library of Colombia. This includes national and institutional policies as well as systematic planning. Some useful tools for implementing a policy framework and planning are ERPANET's Digital Preservation Policy Tool and PLANETS Preservation Planning.

Editors in Colombia are mainly faculty staff that has to share the role with other activities of teaching and research. They usually do not get paid for their editorial time, though some universities are starting to assign hours where they can spend in the journal. Some of the wishing responsibilities for journals being able to preserve digital content are: to publish content under international standards and to have the legal requirement to preserve the submitted content; to ask an authorization from authors to preserve and transform digital content; to transfer digital copies to a third party for long term access; to adopt best practices for preserving digital content. It is the hand of the Editors where the digital preservation starts and they should come up with long-term planning including digital preservation strategies.

The role of publishers of Colombian scientific journals is diffuse. It seems like we do have Colombian high quality journals despite we do not have strong publisher organizations of scientific journals. Universities and associations are actually the publisher of the journals, but not all have the organization, policies and infrastructure required for being a journal publisher. A publisher should have clear digital preservation and copyright policies for assuring future access to the content, even if a trigger event happens. Some of their responsibilities regarding digital preservation of scientific e-journals are: to have an institutional policy about scientific publishing, including digital preservation issues; to have a long-term plan for scientific journals; to guarantee long-term funding for journals; to have an institutional policy about scientific publishing; to have an institutional policy about digital preservation; to have a long-term plan for scientific journals; to guarantee long-term funding for journals.

After doing this study I would say that, in general, we do not have publishers in Colombia; instead we have editors that have to struggle to survive and success at national or at international level, some of them with better results than others. This lack of governance or institutionalization of the publisher influences how well-prepared are the journals to move forward to a digital landscape where the long-term digital preservation is a priority as it is the quality of the published content.

In terms of digital preservation of e-journals Colciencias currently do not have an explicit role or responsibility, neither a plan nor a program for running an archiving solution. However, as they develop the national policy for scientific journals, they could play a strong role to promote digital preservation in both ways: firstly, funding a national project for developing a suit of solutions for digital preservation of scientific journals, that should be available at the National Library of Colombia and secondly to adding digital preservation as one of the quality criteria for scientific journals in Publindex. The suit of solutions should include not only the technological infrastructure at the National Library of Colombia, but also recommendations and technical guides to journals and publishers for licensing and submitting content in the digital archive. The reason for providing this support is because the National Library does not have the economic resources to offer the digital preservation solution for scientific e-journals.

People involved in the Colombian scientific publishing context expect that National Library of Colombia is doing something for preserving the content of e-journals for giving access to future generations if a trigger event happens at the publisher´s side. This generalized assumption give to the library an "ipso-facto" responsibility even if there is not an explicit law to make the digital deposits on it. National Library of Colombia is aware of its responsibility and has implemented a voluntary on-line deposit for digital works, however it does not have enough budget to invest in the infrastructure required for providing digital preservation solutions, not only for e-journals but overall. Some of the National Library responsibilities are: to offer a

national solution for preserving and granting long-term access to digital content published by scientific e-journals; to maintain and develop new features for the long-term digital preservation solutions; to offer technical and copyright guidelines to journals to guarantee long-term preservation; to promote cultural change about transition to digital formats; to promote collaboration between institutions; to have a long-term plan for digital preservation, including financing sources.

Digital preservation is influenced by the cultural and organizational change, because people involved in the process is living a transition from printed to digital-born issues. Some people involved in the scientific publishing context are still in the paper paradigm and they do not understand the implications of the digital content. This transition implies also changes with the traditional way of managing copyright issues. For instance, National Library of Colombia cannot give access to the content preserved without a license or authorization. So if the journals want the library to offer access if a trigger event happens, the publishers should have given previously a proper authorization to the library.

Lack of acknowledgement about the implications of digital preservation is part of the problem. Some editors and publishers are not aware about solutions available, so for this reason they do not use this kind of solutions. Some of them know the solutions but they do not understand its purpose. They have never been worried about long-term access, they are worried about the following issue or about current and day by day problems, and they do not have time to worry about what would happened in 20 or 30 years. Journals follow the recommendations of Colciencias but as the digital preservation it is not part of the criteria, they are not aware of it.

The transition from printed to digital has brought also some misunderstandings about economic issues. Some people though that publishing in digital is cheaper than traditional printing, but this is a fallacy. The costs of producing a digital journal with high standards could be as expensive as

publishing in paper, even more expensive. This is another obstacle for digital preservation, because people who makes the decisions could assume that producing an e-journal is cheaper and do not allocate enough budget for the new requirements of the digital environment.

Institutions at all mentioned levels have some managerial limitations. Journals do not have long-term management plans; institutions do not have digital preservation or scientific publishing policies adopting international standards and best practices. The digital preservation practice could be affected if there is not a permanent alignment with the organizational strategic objectives of all the levels involved. Some institutions are worried about the short-term actions to give access to the digital content and not about long-term issues of digital preservation.

The National Library do not need a single solution for digital preservation, instead they need several solutions at different levels to attend different sectors and formats. One of the solutions should be specific for scientific e-journals, but it needs also a solution for eBooks, other for music, other for digital video, other for multi-format, other for web archiving, for TV, for radio, etc. It is unlikely that one single solution could work for everything. The library should identify "kinds of contents" that share same characteristics and to carry out simultaneous projects for providing different solutions regarding the nature of the content and the stakeholder´s interests.

CHAPTER 6. REFLECTION AND EVALUATION

I will start my reflection about some methodological aspects of this research. At the early stage of the proposal we had decided to make a single case study about the National Library of Colombia, but as soon as the project progressed I realized that it would be fruitful if we expand the scope and perform more interviews with people from institutions with a potential role in the digital preservation of e-journals. This decision gave me a better understanding about the problem and also helps for having a holistic point of view. However, this research cannot be categorized as a single case study neither as multiple cases study. I would say that it is a case study about the National Library of Colombia within the context of the digital preservation of scientific Colombian e-journals. Theoretical propositions originated from how and why questions were extremely useful in guiding the case study.

As it is mentioned in the methodology chapter, this research has used entirely a qualitative approach. I am very glad to have discovered the grounded theory method and constant comparative analysis for analyzing the data gathered trough interviews and observations. The process of grounded theory suggested by Strauss and Corbin (1998) and the Spanish translation done by Universidad de Antioquia (2002) are excellent tools for people without experience in inductive research methods. It was amazing how after follow the iterative analysis and the combination of open, axial and selective coding you are discovering categories that actually is emerging from the data. Nevertheless, it was challenging to stay apart from my biases, as I have worked in the National Library of Colombia and also with scientific e-journals. Regarding this

issue, I would recommend to other researches using the grounded theory method to do the literature review after analyzing the data, just to avoid possible biases. I strongly agree to Pickard (2013) that although the researcher usually read articles and publisher literature before to star the project, the complete literature review is part of the data gathering process and should be included after that. This is also the reason because in the table of content I put the literature review after the methodology and not before it. I also agree Pickard (2013) in considering the grounded theory as an analysis technique more than a proper research method. I nevertheless highlight the benefits of using their inductive techniques for analyzing data.

Performing the interviews was another challenge. My first thought about interviews was that there were easy, but I was completely wrong. I chose semi structured interviews with open ended questions, and after doing the first interview I realized that even if your interview is absolutely unstructured and with open questions, you definitely need a preparation prior to the interview and you have to test the mechanisms to record the interview. I would not suggest use your smart phone for recording. It is much better to use a digital recorder because it is simply designed for that; on the contrast, the smart phone has so many distractors and can affect the fluidity of the interview. I strongly agree to Patai (1987) in the sense that the process of listening participants is an intense form of concentration and openness, and you need to experience yourself to achieve this. I performed 11 interviews and when I was transcribing I realized in which of them I was better prepared than the other ones. At the end, I could say that I definitely learnt how to perform a good semi structured interview.

Another amazing discovery during the research was the software Nvivo for Mac. It was extremely helpful for transcription; open coding, and axial coding. It is an excellent tool that gives you the possibility to annotate; to identify emerging categories, to find relationships and to merge categories as long as you are performing the analysis. You can also write memos and create links between data and memos. I also record a diary of my research as another isolated

memo. NVivo was easy to learn and use and I found plenty of online tutorials that help you to take advantage of the software.

According to Lincoln and Guba (1985, cited by Pickard, 2013) there are four criteria for evaluating trustworthiness in qualitative research: credibility, transferability, dependability and conformability. In terms of credibility, this research has used multiple data collection techniques with multiple source of evidence in order to fulfill a triangulation that give strong support for the results of the research.

For assuring transferability, this research has narrowed the scope of the study at the minimum level for the context of the digital preservation of Colombian scientific e-journals and I have described step by step the evolving of the research. I must recognized that gathering data from Editors and Publishers from just a few sources cannot led in some of the generalizations done in study, for instance when I stated that "we do not have publishers in Colombia" I want to emphasizes the message to improve the publisher role, despite I recognized that there are some Colombian organizations that are actually assuming the role of publishers of scientific journals with excellent results.

Dependability is established by an inquiry audit, where an external auditor examines the research process (Pickard, 2013). Although it was not possible to submit the whole thesis for an external auditor, my supervisor gave me feedback in each stage of the research process. In terms of confirmability, I would say that this study is completely interpretative but I have done my best to avoid biases in analyzing the data and presenting the results. It is true that I have worked in the National Library of Colombia and also with scientific e-journals for several years, but the findings are based on the data gathered and I have referred to the evidence and raw research data for supporting my statements.

Finally, this research considered just a small part of the problems related to the digital preservation and the scientific publishing in Colombia. Some of the further research that could contribute to understand better the issues involved is mixed methodological approaches for analyzing university policies about digital preservation and scientific publishing. A quantitative analysis (for example, a survey) about the policies and plans of the publishers in Colombia could help to understand the actual role of the Colombian publishers. This analysis could have done for several Latin America publishers because all of these countries could have sharing the same issues. Research about research governance in Universities and Research Agencies and their relation to the scientific publishing and digital preservation would be helpful.

The economic issues of the digital preservation are also a matter of importance not only for publishers but also for the National Library of Colombia. A study in this area could be extremely useful to take decisions about budgets and financial needs for assuring digital preservation. The cost of digital preservation should analyze the infrastructure required, including technology, human resources, policies, plans, methodologies and all the requirements for providing actual digital preservation.

Finally, analyzing the copyright issues for allow digital preservation by the National Library of Colombia or other depositing agencies would be absolutely crucial to find possible solutions for making digital preservation and perpetual access easier for all.

REFERENCES

2CUL LOCKSS. (2011). Final Report of the 2CUL LOCKSS Assessment Team Cornell University Library & Columbia University Library. Cornell University Library & Columbia University Library. Retrieved from https://www.2cul.org/sites/www.2cul.org/files/2CULLOCKSSFinalReport.pdf

Becker, C., Kulovits, H., Guttenbrunner, M., Strodl, S., Rauber, A., & Hofman, H. (2009). Systematic planning for digital preservation: evaluating potential strategies and building preservation plans. International Journal on Digital Libraries, 10(4), 133–157.

Besser, H. (2007). Collaboration for electronic preservation. Library Trends, 56(1), 216–229.

British Sociological Association. (2002). Statement of ethical practice for the British Sociological Association Durham: British Sociological Association.

Burnhill, P. (2013). Tales from the keepers registry: Serial issues about archiving & the web. Serials Review, 39(1), 3–20.

Chanod, J., Dobreva, M., Rauber, A., Ross, S., & Casarosa, V. (2010). Issues in digital preservation: towards a new research agenda. Automation in Digital Preservation, 1–14.

Charmaz, K. (2006). Constructing grounded theory: A practical guide through qualitative research. London: SAGE.

Colciencias. (2012). Indicadores generales Publindex 2002 - 2011. Retrieved from http://www.colciencias.gov.co/sites/default/files/ckeditor_files/files/INDICADORES%20GENERALES%20PUBLINDEX%202011.pdf

Colciencias. (2015). Índice Bibliográfico Nacional - IBN Publindex. I Actualización 2014. Retrieved from http://publindex.colciencias.gov.co:8084/publindex/EnIbnPublindex/resultados.do

Gaur, R. C., & Tripathi, M. (2012). Digital preservation of electronic resources. DESIDOC Journal of Library & Information Technology, 32(4).

Glaser, B., & Strauss, A. (1967). The discovery of grounded theory: Strategies for qualitative research. New York: Aldine publishing.

Jones, M., & Beagrie, N. (2001). Preservation management of digital materials: a handbook. British Library London.

Kirchhoff, A. J. (2008). Digital preservation: challenges and implementation. Learned Publishing, 21(4), 285–294.

Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook. Sage.

Morgan, C., Campbell, B., & Teleen, T. (2012). The role of the academic journal publisher and open access publishing models. International Studies Perspectives, 13(3), 228–234.

Noonan, D. (2014). Digital Preservation Policy Framework: A Case Study.

Patton, M. Q. (2005). Qualitative research. Wiley Online Library.

Pickard, A. (2013). Research methods in information. Facet publishing.

Roosendaal, H. E., & Geurts, P. A. T. M. (1997). Forces and functions in scientific communication: an analysis of their interplay.

Seadle, M. (2011). Archiving in the networked world: open access journals. Library Hi Tech, 29(2), 394–404.

Seidman, I. (2013). Interviewing as qualitative research: A guide for researchers in education and the social sciences. Teachers college press.

Sompel, H. van de, Payette, S., Erickson, J., Lagoze, C., & Warner, S. (2004). Rethinking scholarly communication: building the system that scholars deserve. D-Lib Magazine; 2004 [10] 9.

Strauss, A., & Corbin, J. (1998). Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage publications.

Strauss, A., & Corbin, J. (2002). Bases de la investigación cualitativa. Ténicas y procedimientos para desarrollar la teoría fundamentada. Universidad de Antioquia.

Strodl, S., Becker, C., Neumayer, R., & Rauber, A. (2007). How to choose a digital preservation strategy: Evaluating a preservation planning procedure (pp. 29–38). Presented at the Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries, ACM.

Thomas, G. (2010). How to do your case study: A guide for students and researchers. Sage.

Valles, M. S. (1999). Técnicas cualitativas de investigación social. Sintesis Editorial.

Yin, R. K. (2013). Case study research: Design and methods. Sage publications.

APPENDIXES