Digital preservation in institutional repositories: a systematic literature review

Journal:	*Digital Library Perspectives*
Manuscript ID	DLP-02-2021-0011.R2
Manuscript Type:	Research Paper
Keywords:	digital preservation, institutional repositories, preservation policies, administrative metadata, repository certification, systematic literature review

Digital preservation in institutional repositories: a systematic literature review

Abstract

Purpose: This paper carries out a literature review on the implementation of digital preservation policies, strategies and actions by institutional repositories. The objective is to identify, out of the published experiences, at which level they are fulfilling the function of ensuring the long term availability of the deposited materials.

Design/methodology/approach: Using a systematic literature review methodology, a total of 21 articles from international refereed journals published between 2009 and 2020 are reviewed.

Findings: The research production on this subject is very limited. The scarce number of published articles proves that the interest of repository managers has been focused on issues other than to assure the long term availability of the assets they store. Our literature review has not found clear evidences about how institutional repositories are implementing digital preservation. It is particularly striking the lack of works focused on the situation in European countries. More field studies are needed. They would allow to extract conclusions and produce best practices to help managers to improve preservation strategies.

Keywords

systematic literature review ; digital preservation ; institutional repositories ; preservation policies ; administrative metadata ; repository certification

Introduction

Authors like (Koler-Povh et al., 2015), (Marsh, 2015), (Nemati-Anaraki and Tavassoli-Farahi, 2019) or (Bulock, 2016) have illustrated how institutional repositories (IRs) have become an essential element of the scholarly communication ecosystem.

The most widespread definition of IRs was provided by Clifford A. Lynch in his seminal article (Lynch, 2003): *A set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution.*

The importance of digital preservation was already pointed out in this definition. However, it has been one of the repository functions that has gathered less interest by the community.

The efforts of repository managers have been concentrated on demonstrating their usefulness and persuading academics to upload their publications. This has led to a relaxation in the basic requirements regarding digital preservation. The priority on contents recruitment has prevailed over quality concerns.

Even when IRs make clear their commitment to guarantee the availability of contents over the long term, a quick review of the literature shows that there are few evidences about how are they satisfying such purpose.

The main objective of this paper is to fill this gap, carrying out a systematic literature review on the issue of digital preservation in institutional repositories. We try to learn, out of the published experiences, at what degree IRs are fulfilling the function of assuring the long term availability of the deposited resources. Our practical purpose is to find out evidences in the literature about how repositories are actually implementing long-term preservation. The ultimate goal is to

provide useful information to researchers and practitioners to help them to develop a picture of the current situation in order to set up research priorities and action plans for the future.

Literature review

The research interest regarding IRs has evolved considerably over the last two decades. (Stevenson and Zhang, 2015) differentiate several time periods in the evolution of repository research. The initial phase was focused on introducing concepts, metadata management and potential design structures and frameworks. This is consistent with (Cho, 2014), who concluded that the main focus of interest of research in the early stages of repository development was interoperability of metadata and quality improvement. Secondly, the research moved towards the case studies, specifically related to the different user perspectives. The latest period culminated the expansion and development of IRs, with evaluation and business models as the main areas of attention.

A great deal of authors have agreed on the importance of digital preservation for IRs. Just as an example, (Hockx-Yu, 2006), (Li and Banach, 2011), (da Silva Júnior and Borges, 2017), (Francke et al., 2017), quote the already mentioned Lynch's paper and admit that preservation activities should be implemented by repositories.

Nevertheless, it seems that repository managers have paid little attention to this function. Authors like (Stevenson and Zhang, 2015) and (Cho, 2014) have shown that digital preservation is not one of the most frequent subject terms in the technical literature on IRs. It is more an independent research domain than something connected to the repositories practical activities.

Back in 2008, (Mcgovern and Mckay, 2008) already concluded that repository managers had been concentrated in enlarging the amount and quality of contents, persuading researchers about the usefulness of open access dissemination of research, or improving the interoperability and discovery of research outputs.

Recently, (Saini, 2018) published a literature review on the topic of digital preservation and IRs. From this paper we can conclude that, unfortunately, the situation has not changed over time. This work has two weaknesses in our opinion. Firstly, it is a narrative review, with not defined procedures for identification of reviewed studies. Furthermore, all the works analysed were published before 2015. In this sense, our research complements and goes a step forward by applying a scientific methodology to the literature survey.

Methodology

(Boell and Cecez-Kecmanovic, 2014) characterize the literature review as the process in which authors identify, analyse, assess and synthesize earlier research on the issues they attempt to explore. For the purpose of this paper, we have chosen the systematic literature review (SLR) approach and followed the methodology proposed by (Denyer and Tranfield, 2009). The phases they differentiate could be summarized in four: research question(s) formulation, locating studies through a bibliographic search, study selection and evaluation and analysis and synthesis.

Research Questions Formulation

Our research questions have been articulated around the three elements the ALA (American Library Association, 2007) enumerates as essential constituents of digital preservation: policies, strategies and actions. Policies prove the institution's commitment to preserve digital content for future use. Strategies and actions show how the institution is actually implementing the policies in the daily workflows of the repository. Furthermore, we consider one more element: certification

as trusted repositories. Certification is reached when the accomplishment of the previous elements is verified in the light of some international standard.

Building on these properties and the considerations pointed out in the introduction of this paper, our aim is to answer the following questions:

RQ1 Policies: To what extent do the IRs have preservation policies and plans?

The repository preservation policy must be more than a simple statement of commitment to make contents available in the future. It should be reflected in a document where at least the following elements need to be defined: the purpose of the policy, its scope, staff roles and responsibilities (Beagrie, et al., 2008).

RQ2 Strategies and actions: How are IRs implementing digital preservation?

Strategies and actions should be based on proactive digital preservation as defined by (Bountouri, 2017). This means, far from preservation being just a task for the end-of-life of a digital object, to span to the whole content life cycle. Such strategies start from the early creation stages of a digital resource, engaging the producers of the digital material and raising awareness on the importance of preservation.

Taking this in consideration, we have focused on:

- RQ2a Preservation metadata. There is a need to preserve metadata about the technological and other contexts of a digital object's creation and use (Day, 2001). Are there evidences of IRs producing and storing good quality and complete metadata to be able to resolve future issues in rendering their assets?
- RQ2b File formats. File format obsolescence has been recognized as one of the major threats to preservation. Do IRs have clear indications or policies on which formats are accepted for ingestion? Do they perform any subsequent activities after ingest of files concerning file checking, migration and so on?
- RQ2c Systems and technology. Digital preservation activities should not be limited by the technologies available. They need to be carried out in an independent mode, without being constrained by proprietary solutions. What are the technical solutions available which are being implemented by IRs?

RQ3 Certification: Are IRs being certificated as trusted repositories following international standards?

Trustworthiness is a fundamental issue for any digital repository (Corrado, 2019) (Becker, et al., 2009). Establishing a trusted and reliable archive should increase the confidence of authors and will foster their disposition to upload contents.

Checklists and criteria catalogues have been developed to be used in trustworthiness certification of digital archives. We have looked for instances in which repositories have carried out certification processes following international standards like TRAC (Trustworthy Repositories Audit & Certification: Criteria and Checklist); ISO 16363: Audit and certification of trustworthy digital repositories; Catalogue of Criteria for Trusted Digital Repositories. Version 2 (nestor, 2009), CoreTrustSeal ...

Locating studies

The source data for this literature review was collected from four of the most comprehensive bibliographic databases available: Web of Science, Scopus, LISA (Library and Information Science Abstracts) and LISTA (Library, Information Science and Tehnology Abstracts).

The search was restricted to peer reviewed articles published in English language in the period 2000-2020. The limitation by publication date was based on (Cho, 2014). In this study on the structure of the institutional repository field, the author dated the first published article on the subject in 2001.

In order to have as many source articles as possible, we took a broad approach in our database search strategy. General concepts were used in order to gain exhaustive results and to not miss relevant papers in the first stage of the review. We rather preferred to deal with non-relevant articles in the search output, than missing relevant ones due to the use of a narrower search strategy.

According to this principles, we used the broad search strategy: "institutional repositories" and preservation. Searches were executed in early December, 2020, with the results shown in Table 1.

TABLE1

The initial results set (386 records) was exported to a bibliographic database using RefWorks as reference manager. Duplicated records were deleted using the deduplicate function of RefWorks, followed by manual checking to assure the correctness of the process. The final data set to be reviewed was made up of 262 different records and it is available, as bibtex format, at: https://bit.ly/3a9I8Nn

In an initial assessment of the results, we realized that the keywords selected for the bibliographic search were used quite often without a clear and actual relation to the content of the paper they described. For such reason, it was necessary to apply well defined exclusion and inclusion criteria in order that only the most relevant documents would be chosen. These were to conform the data set object of the review.

Study selection and evaluation

The main inclusion criteria, as mentioned earlier, was: peer reviewed articles published in English language over the period 2000-2020. Additionally, we defined one more criteria:

- IC2: Selected articles must directly or indirectly address, at least, one of the research questions.

The following rejection criteria were applied:

- EC0: We excluded non relevant articles. Frequently, even when keywords were included in the abstract or subject terms, they were not descriptive of the content of the documents. For instance, digital preservation is mentioned often as one of the principal repository functions. Typically, authors include the concept in keywords and abstract, however then they go on describing the importance, usefulness and requirements of such function, but not going any further in describing practical implications. This led to a substantial reduction of 121 documents in the initial results set.
- EC1: We excluded articles where non-scientific methodology was applied or no evidence of the actual situation in IRs was provided. Under these circumstances there are two types of studies. First, we categorized the articles following the classification of research methods described by (Palvia and Sibley, 2007). Only case studies, field studies or qualitative research were taken in consideration. Secondly, we excluded surveys when the object of study was not directly related to preservation. Frequently, surveys are intended to describe usage pattern, perception of the usefulness, authors motivation to contribute, attitudes of academia, success factors... In this kind of studies, preservation is

addressed as a marginal issue. Overall, articles in this exclusion category accounted for 59 items.

- EC2: Articles focused on heritage digital repositories were also excluded. By heritage repositories we mean those that mainly handle digital surrogates of analog resources. Usually, they deal with digital rendering of special collections in libraries. Digital repositories and IRs are differentiated by type and origin of resources, document life cycles, metadata creation, and methods of arrangement (Kutay, 2014). Content in digital repositories is fairly homogeneous because it has been created following well established international standards on file formats and digitization procedures. It is made up of good quality metadata produced by curators and standardized file formats. It is ingested by the repository staff in a controlled environment and following detailed work flows. This picture differs considerably of the IRs, characterized by the heterogeneity and complexity of the digital resources, usually digital born assets, with metadata being created by non-specialized contributors. 47 articles were excluded.
- EC3: Finally, we excluded articles reporting ongoing research projects. Many of such projects are now long time finished. Probably the results were, or are going to be, incorporated into the management work flows of the IRs. Nevertheless, they clearly do not add any practical insight into the current situation. Under this category 16 articles were excluded.

We obtained the full text of all articles in the results set and read introduction, methodology and conclusions in order to apply the inclusion and exclusion criteria. From the original set, 243 articles were excluded and 19 passed the initial test.

The last step in the process of article selection was quality assessment. The purpose was to make a selection decision based on the overall quality of the candidate papers. We used two methods to carry out the process: quality criteria and citations count.

The following quality assessment criteria were set up to evaluate the selected papers.

- QA1: How relevant is the topic addressed in the research paper to our SLR?
- QA2: To what extent are IRs the main target of the study?
- QA3: Is the research methodology clearly specified in the paper?
- QA4: To what extent are the results of the research paper relevant for our SLR?

We rated each criterion with 1, 0.5 or 0 if the article satisfies fully, partially or not at all the criteria respectively. The final score of the paper was the sum of all criteria. We considered the paper's quality as low if it scored below 2. The result of applying the quality assessment criteria is given in table 2. No article was rated as low quality so no further documents were excluded from the data set.

TABLE2

The number of citations an article has received is commonly considered an indicator of its research quality. We have performed citation counts for all the initially selected papers. Table 2 shows the number of citations for each document in three citations indexes. Almost all articles in our set have been cited at least once. Only three articles had no citations. They have been published recently, in the last three years. Since a work to be cited requires a certain period of time to pass, we decided not to exclude these articles based only on the number of citations criteria.

To summarize, once the exclusion criteria and quality assessment steps were applied, the dataset of selected articles for review was reduced to those 19 which match all criteria.

In order to complete the source data, we performed reference searches. This involves examining the references and works cited in each article following the principle that one paper relevant for our research should cite and be cited by other papers that could also be of interest for our study.

With this purpose, we reviewed the references of the source papers (backward) and the citations (forward) identified in the mentioned citation indexes. For each reference and citation, we decided if the article could be of interest for our study. In such case, we downloaded the full text and applied the inclusion and exclusion criteria in the same way that we did with the bibliographic search results. This step led to the addition of 2 new articles, increasing the total set of works to be reviewed to 21. They are listed in the Appendix and constitute the object of study for this review.

Analysis and synthesis

For analysis and synthesis, the basic metadata of each source article was exported to a spreadsheet. In addition, we defined several facets in order to further characterize the scope of each paper: country and geographic region under study, publication journal, publication date and research question addressed. The resulting dataset is available at https://bit.ly/2LtEYLY. Each item in the dataset is identified with a code made up of a letter stating the phase in which the article was chosen (S = bibliographic search, C = citations search) plus the id of the article in the RefWorks database

The distribution of selected articles by publication year shows that the first studies were published in 2009. Since then, the number of new publications has been stable, with a peak in 2017. The late appearance of studies does not mean that the interest for digital preservation only arises one decade after the first IRs were born. Interest was clear since the beginning of the field development, but only after 2010 such interest crystallized in concrete policies, strategies and activities. It is worth noting that half of the studies have been published over the last three years. This fact proves that the interest in preservation is increasing as the repositories become mature tools.

Having only 21 studies published over a time spam of 20 years gives us a very small average of one single article per year. This quite limited research production confirms the scarce interest in the issue by repository managers as we already pointed out in the introduction of our work.

As for the geographic scope of research, the largest number of publications cover initiatives from North America (6 USA, 1 Canada and 5 both of them). They include case studies from large universities as well as several analyses of the general situation in the Association of Research Libraries (ARL). Europe and Africa, with three case studies each one, are in the next level due to number of publications.

It is noteworthy the small number of publications reporting on European experiences (3 papers). In our opinion, this lack of studies may be related to the role played by the European Union in the development of IRs. The interest has been driven by policies regarding open access. Such policies have fostered the promotion of mandates to make the research results resulting of projects funded by the EU available in IRs. However, they have paid little attention to digital preservation issues.

Asia and Latin America with 1 article each one and Africa with 3 complete the list.

We lack studies with broader geographic coverage, involving different countries and continents. This would allow us to draw parallels and comparison between implementations of digital preservation at different levels.

As for the distribution of articles by publication source, Digital Library Perspectives is the journal with the largest production on the issue (5 articles). The second one is D-Lib Magazine even when it was closed down in 2017 (2 articles).

RQ1: Preservation Policies and Plans

Seven articles covered the topic of digital preservation policies and plans. Two of them, C4493 and S7142, went over preservation policies within the ARL libraries. Back in 2011, S7142 confirmed that an increasing number of research libraries had started to move digital preservation programs ahead by developing preservation policies.

This study was followed in 2017 by C4493, which is the most exhaustive survey on the topic of preservation policies we have tracked down. The author found a considerable increase in the number of universities with policies compared to 2011. She goes a step forward by reviewing both the positive elements and the shortcomings of the policies. Among the top challenges for libraries listed, there are: the increasing volume of digital content, the rapid evolution of technologies, staff expertise, economic costs and lack of understanding and education on the topic.

The North America's picture is completed by S7231, a case study about policy development on the Purdue University Research Repository (PURR).

Outside North America, we found studies in Africa and Latin America. Three studies examining the situation of digital preservation in Africa have differing conclusions. S7196 carried out a survey on the whole continent and concluded that more than half of respondents had a policy statement regarding digital preservation. Policy statements are not a proper policy. Statements only express the commitment to preserve contents. In this sense, we should conclude that while IRs in Africa declare its intention to preserve contents, it is not clear if they have a true digital preservation policy.

S7360, by the same author, narrowed the scope to Nigeria. Its conclusions were constrained by the few universities that have an institutional repository in the country, only 11%. The situation is different to the rest of the continent, because almost no IRs have preservation policy statements. The same conclusions is reached by S6999 in their survey of IRs in Ghana.

Finally, S7208 went over the situation in Brazil: IRs do not have any published digital preservation policies, even though some repositories state their intention of preserving digital material in their institutional information policy.

RQ2a: Preservation metadata

There is small evidence about the usage of preservation metadata and most part of the studies focus on the situation of ARL libraries.

S7142 found that 58% of IRs in ARL libraries reported recording preservation metadata: technical data, rights information, provenance or ownership history, and change tracking of the resource.

In a survey focused on the usage of administrative metadata, S7032 concluded that there was no true consensus of administrative metadata accommodated and collected by the repositories. In her opinion, the community is not putting enough effort into administrative metadata. She identified many possible reasons for these “metadata shortfalls”: the ingest of digital materials is rapid and increasing at a rate that may well put their management beyond the means of most institutions; metadata standards are voluminous and complex; staffing may be an issue too ...

The difference between both studies could be attributed to a significant disconnection between what the community is saying in surveys and what is actually happening on the ground. In this sense, answers to surveys should be accompanied by the corresponding evidence.

Two case studies (S7231 and S7354) with geographic focus on the USA, complete the picture. They report on the usage of metadata in PURR and Colorado Alliance of Research Libraries Digital Repository respectively. In both cases, two standards are mentioned, METS and PREMIS.

If the previous research drew a daunting situation in North America, much worse is the picture outside. There is an almost complete lack of studies on the utilization of administrative metadata by IRs. The only exception is S7072 analysing the case of India. They found a desolate situation: no trace of administrative metadata used at all.

RQ2b: File formats

Once more, the situation with a better coverage is that of the ARL libraries. S7122 and S7142 examined file format policies. They found out that half of the IRs had one. Even though, many of them relied on policies associated with particular software, rather than creating policies from scratch. S7142 found that 90% of polices clearly identified supported or recommended file formats, while the rest briefly say they are committed to long-term digital preservation of all materials. Even when restrictions on file formats accepted for ingest seems to be common, S7122 points out that many IRs were willing to accept file formats used by their researchers regardless of how highly the repository trusted the formats. This is related to the promotion of contents collection instead of quality constrains.

This same situation is described by S7290 in Sweden, where about 70% of the surveyed repositories stated that they had some form of instructions or policy concerning which file formats were accepted in the repository.

Other studies focus on the preservation activities performed on the files ingested in the system. S7072 in a study on IRs in India, concluded that the performance of preservation activities on file formats was extremely low, with less than 20% of IRs performing file validation on ingestion, file audit or migration. Similarly, S7290 reported that some of the repository managers stated that they sometimes converted files from one format to another. Nevertheless, it was simply a matter of conversion from DOC to PDF. Finally, in an interesting study about file formats in two Catalan repositories, S7344 identified a discrepancy between the preservation statements declared by the repositories and the actual practice by the managers. Files were ingested in the system without any subsequent control.

Interestingly, when we look at the case studies rather than surveys, the situation seems different. C4397 reports the opposite situation, with archival file formats widely used and preservation activities performed on ingested files. In the same line, S7231 enumerates several activities related to file management included in the PURR preservation plan.

RQ2c: Systems and technology

We could differentiate systems fully oriented towards an integral solution of digital preservation and specialized software tools addressed to deal with particular activities in the preservation process.

In the first group, we have only found references to Archivematica and Rosetta. Archivematica's usage by IRs seems to be quite recent. S7122 describes its application in ARL libraries. S7101 undertook a project in 2016 to explore facets of digital accessioning and preservation using this software. In particular, they tested its capacity to transfer and ingest a selection of text, image,

audio and video formats. Furthermore, they measured its potential for migrating objects across content management systems like DSpace, CONTENTdm ...

The unique reference to Rosetta by Ex-Libris is C4397. The author limits herself to point out that the institution is considering to adopt the system without any further detail.

In the second group, specialized software tools, we include the usage of applications addressed to deal with particular preservation activities as for instance: DROID file format identification or JHOVE an application that performs format identification, validation and characterization.

C4406 concludes that few of the existing digital preservation tools and services have addressed the specific needs of IRs; in practical terms they have necessitated action that is additional rather than integral to repository workflow and describes the case of the EPrints software plugin tools. S7246 also describes the use of EPrints preservation toolkit and other additional tools in order to monitor file formats.

RQ3: Certification

S7138 reported, as early as 2011, that two repositories, at the University of the Arts London and the London School of Economics, used DRAMBORA (Digital Repository Audit Method Based on Risk Assessment). Both initiatives took a lightweight approach with the tool. While UAL failed to complete the process, at the LSE the team identified ten risks which were representative of concerns in different organizational, technical and other locally relevant areas.

S7354, S7231 and S7358 reported on the usage of TRAC and its continuation as ISO 16363 in PURR, Colorado Alliance of Research Libraries Digital Repository and Scholars Portal respectively.

Conclusions and further work

We have carried out a structured literature review on the issue of actual implementation of digital preservation in institutional repositories. After reviewing articles referenced in four of the most comprehensive bibliographic databases, the primary conclusion is that the research production on this area is very limited. The scarce number of articles published confirms that the interest of repository managers has been focused on issues other than to assure the long term availability of the assets they store.

Furthermore, the literature available in form of surveys and case studies concentrates in describing the situation in North America with insufficient reports focused elsewhere. In particular, is striking the lack of works on the situation in Europe.

The reduced number of articles dealing with each of the research questions make hard to extract accurate conclusions. This is one limitation of our work. The constrain in our review to articles published in journals ensures the quality of the source documents but it could exclude cases published in different media like conferences, research reports or working papers.

Bearing this limitation in mind, we can draw the following conclusions:

- RQ1: Our review shows that the interest for developing preservation policies and plans has been increasing in North America while it is almost missing in IRs elsewhere, where preservation is mentioned as a commitment, more a promise than a real concern for repository managers.
- RQ2a: A similar situation is found when we look to the usage of administrative metadata. IRs in North America show a high degree of application of this kind of metadata, which is needed to implement any activity related to digital preservation. There is no evidence of usage of administrative metadata in repositories in the rest of the world.

- RQ2b: With regards to file formats, we can conclude that IRs have policies to determine which formats are accepted for ingestion. Frequently, policies are driven by the requirements of the management systems. Nevertheless, digital preservation should be more than policies, and repository administrators seems to forget any further processes on file formats after ingestion. Studies show a lack of activity related to migration, checking, etc.
- RQ2c: There are few cases of software applications used to carry out integrated management of preservation activities. We come across only one software in the literature: Archivematica, but with very limited detail about how it is being implemented.
- RQ3: There is almost no proof in the literature about cases of trustworthy certification of IRs. Only two direct mentions of check lists usage for self-auditing or indirect mentions to certification processes.

Finally, our review has not found clear evidences on how IRs are implementing digital preservation at any of the research questions proposed. From this assertion, we cannot derive that IRs are not carrying out this kind of activities. On the contrary, we can only conclude that such activities are not being described in the research literature. Clearly, more work is needed on the issue. In particular, it would be necessary a study at European level and other regions to gather detailed data on digital preservation policies, strategies and activities that will allow to draw a portrait of the current situation and to extract conclusions and produce best practices to help managers and practitioners to improve or develop preservation strategies.

Appendix

Bibliography

American Library Association (2007), 'Definitions of Digital Preservation', ALA Annual Conference, Washington, D.C., June 24, 2007. Available at: http://www.ala.org/alcts/resources/preserv/defdigpres0408

Beagrie, N., Semple, N., Williams, P. and Wright, R. (2008), Digital preservation policies study: Part 1: Final report october 2008, Technical report, Charles Beagrie Limited. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.214.9056&rep=rep1&type=pdf

Becker, C., Kulovits, H., Guttenbrunner, M., Strodl, S., Rauber, A. and Hofman, H. (2009), 'Systematic planning for digital preservation: evaluating potential strategies and building preservation plans', International Journal on Digital Libraries, Vol. 10, pp. 133–157. Available at: http://doi.org/10.1007/s00799-009-0057-1

Boell, S. K. and Cecez-Kecmanovic, D. (2014), 'A hermeneutic approach for conducting literature reviews and literature searches', Communications of the Association for Information Systems 34. Available at: https://doi.org/10.17705/1CAIS.03412

Bountouri, L. (2017), Archives in the Digital Age: Standards, Policies and Tools, Elsevier.

Bulock, C. (2016), 'Feeling green', Serials Review, Vol. 42, no. 4, pp. 311–313. Available at: http://doi.org/10.1080/00987913.2016.1243077

Cho, J. (2014), 'Intellectual structure of the institutional repository field: a co-word analysis', Journal of Information Science, Vol. 40, no. 3, pp. 386–397. Available at: http://doi.org/10.1177/0165551514524686

Corrado, E. M. (2019), 'Repositories, trust, and the coretrustseal', Technical Services Quarterly, Vol. 36, no. 1, pp. 61–72. Available at: http://doi.org/10.1080/07317131.2018.1532055

da Silva Júnior, L. P. and Borges, M. M. (2017), 'Digital preservation policies of the institutional repositories at brazilian federal universities', The Electronic Library, Vol. 35, no. 2, pp. 311–321. Available at: http://doi.org/10.1108/EL-09-2015-0170

Day, M. (2001), Metadata for digital preservation: A review of recent developments, in 'ECDL 2001: Research and Advanced Technology for Digital Libraries', pp. 161–172. Available at: http://doi.org/10.1007/3-540-44796-2_15

Denyer, D. and Tranfield, D. (2009), "Producing a systematic review", Buchanan, D. and Bryman, A. (eds). The SAGE handbook of organizational research methods, SAGE.

Francke, H., Gamalielsson, J. and Lundell, B. (2017), 'Institutional repositories as infrastructures for longterm preservation', Information Research, Vol. 22, no. 2. Available at: http://www.informationr.net/ir/22-2/paper757.html

Hockx-Yu, H. (2006), 'Digital preservation in the context of institutional repositories', Program: Electronic Library and Information Systems, Vol. 40, no. 3, pp. 232–243. Available at: http://doi.org/10.1108/00330330610681312

Koler-Povh, T., Mikoš, M. and Turk, G. (2015), 'Institutional repository as an important part of scholarly communication', Library Hi Tech, Vol. 32, no. 3, pp. 423–434. Available at http://doi.org/10.1108/LHT-10-2013-0146

Kutay, S. (2014), 'Advancing digital repository services for faculty primary research assets: An exploratory study', Journal of Academic Librarianship, Vol. 40, no. 6, pp. 642–649. Available at: http://doi.org/10.1016/j.acalib.2014.08.006

Li, Y. and Banach, M. (2011), 'Institutional repositories and digital preservation: Assessing current practices at research libraries', D-Lib Magazine, Vol. 17, no. 5-6. Available at: http://www.dlib.org/dlib/may11/yuanli/05yuanli.html

Lynch, C. A. (2003), 'Institutional repositories: Essential infrastructure for scholarship in the digital age', ARL: A Bimonthly Report, Vol. 226. Available at: http://old.arl.org/resources/pubs/br/br226/br226ir print.shtml

Marsh, R. M. (2015), 'The role of institutional repositories in developing the communication of scholarly research', OCLC Systems and Services, Vol. 31, no. 4, pp. 163–195. Available at: http://doi.org/10.1108/OCLC-04-2014-0022

Mcgovern, N. Y. and Mckay, A. C. (2008), 'Leveraging short-term opportunities to address long-term obligations: A perspective on institutional repositories and digital preservation programs', Library Trends, Vol. 57, no. 2, pp. 262–279. Available at: http://doi.org/10.1353/lib.0.0038

Nemati-Anaraki, L. and Tavassoli-Farahi, M. (2019), 'Scholarly communication through institutional repositories: proposing a practical model', Collection and Curation, Vol. 37, no. 1, pp. 9–17. Available at: http://doi.org/10.1108/CC-01-2018-002

Otto, J. J. (2014), 'Administrative metadata for long-term preservation and management of resources a survey of current practices in arl libraries', Library Resources and Technical Services, Vol. 58, no. 1, pp. 4–32. Available at: http://doi.org/10.5860/lrts.58n1.4

Palvia, P. and Sibley, E. H. (2007), 'A profile of information systems research published in information and management', Information and Management (44), 1–11. Available at: https://doi.org/10.1016/j.im.2006.10.002

Saini, O. P. (2018), 'Understanding the role of institutional repository in digital preservation in academic libraries: A review of literature', Library Philosophy and Practice pp. 1–14. Available at: https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=5249&context=libphilprac

Stevenson, J. A. and Zhang, J. (2015), 'A temporal analysis of institutional repository research', Scientometrics, Vol. 105, pp. 1491–1525. Available at: http://doi.org/10.1007/s11192-015-1728-x

References accessed: 03/02/2021

Table 1: Search strategies and records found			
Database	Provider	Search strategy	Records
WoS	Clarivate	TOPIC: ("institutional repositories" and preservation) Refined by: DOCUMENT TYPES: (ARTICLE) AND LANGUAGES: (ENGLISH) Timespan: 2000-2020. Databases: WOS, CCC, DIIDW, KJD, MEDLINE, RSCI, SCIELO	69
Scopus	Elsevier	TITLE-ABS-KEY ("institutional repositories" AND preservation) AND PUBYEAR > 1999 AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (DOCTYPE , "ar"))	115
LISA	ProQuest	su("institutional repositories" AND preservation. Limit to, publication date 1999 and scholarly journals and language English and Peer review articles	94
LISTA	EBSCO	SU "institutional repositories" AND SU preservation. Publication date: 20000101-20201231. Language: English.	108
		Total records found	386
		Total records after deduplication	**262**

Table 2. Quality assessment results for selected articles

Id	Quality Assessment					Citations Count		
	QA1	QA2	QA3	QA4	QAT	GS	Scopus	WoS
S6999	1	1	1	1	4,00	2	2	2
S7032	1	1	1	1	4,00	6	4	4
S7072	1	1	1	1	4,00	0	0	0
S7085	0,5	1	1	1	3,50	0	0	nd
S7101	1	0,5	0,5	1	3,00	6	2	2
S7122	1	1	1	1	4,00	31	9	nd
S7138	1	0,5	0,5	1	3,00	11	1	nd
S7142	1	1	1	1	4,00	77	31	nd
S7187	0,5	1	1	0,5	3,00	0	0	0
S7196	1	1	1	1	4,00	2	9	9
S7203	1	1	0,5	1	3,50	4	2	2
S7208	1	1	1	1	4,00	6	5	5
S7231	1	1	1	1	4,00	15	3	nd
S7246	0,5	1	0,5	0,5	2,50	7	nd	nd
S7290	1	1	1	1	4,00	8	4	1
S7344	1	1	1	1	4,00	9	5	4
S7354	1	1	0,5	0,5	3,00	4	3	2
S7358	1	0,5	0	0,5	2,00	7	2	nd
S7360	1	1	1	1	4,00	12	5	nd
C4397	1	1	1	1	4,00	34	12	6
C4493	1	1	1	1	4,00	7	2	3

Appendix

Id	Reference	QAT	RQ
C4397	Oehlerts, Beth. "Digital Preservation Strategies at Colorado State University Libraries." Library Management 13.1/2 (2013): 83-95. *10.1108/01435121311298298*	4	RQ2
C4493	Dressler, Virginia A. "The State of Affairs with Digital Preservation at ARL Member Libraries A Survey and Analysis of Policy." Digital Library Perspectives 33.2 (2017): 137-55. *10.1108/DLP-08-2016-0030*	4	RQ1
S6999	Adjei, Emmanuel, Monica Mensah, and Eric Amponsah Amoaful. "The Story so Far-Digital Preservation in Institutional Repositories the Case of Academic Libraries in Ghana." Digital Library Perspectives 35.2 (2019): 80-96. *10.1108/DLP-12-2018-0039*	4	RQ1 RQ2
S7032	Otto, Jane Johnson. "Administrative Metadata for Long-Term Preservation and Management of Resources A Survey of Current Practices in ARL Libraries." Library Resources & Technical Services 58.1 (2014): 4-32. *10.5860/lrts.58n1.4*	4	RQ2
S7072	Shajitha, C. "Digital Curation Practices in Institutional Repositories in South India: A Study." Global Knowledge, Memory and Communication 69.8-9 (2020): 557-78. *10.1108/GKMC-10-2019-0125*	4	RQ2
S7085	Ismail, S. "A Life Well Lived: Looking Backwards and Forwards and Sideways Too: Exploring the Full Lifecycle of Institutional Scholarly Communication at Your Library." Serials Librarian 74.1-4 (2018): 116-8. *10.1080/0361526X.2018.1442968*	3,5	RQ2
S7101	Trujillo, S., et al. "Archivematica Outside the Box: Piloting a Common Approach to Digital Preservation at the Five College Libraries." Digital Library Perspectives 33.2 (2017): 117-27. *10.1108/DLP-08-2016-0037*	3	RQ2
S7122	Rimkus, K., et al. "Digital Preservation File Format Policies of ARL Member Libraries: An Analysis." D-Lib Magazine 20.3-4 (2014). *10.1045/march2014-rimkus*	4	RQ2
S7138	Pickton, M., et al. "Preserving Repository Content: Practical Tools for Repository Managers." Journal of Digital Information 12.2 (2011): 1-14.	3	RQ2 RQ3
S7142	Li, Y., and M. Banach. "Institutional Repositories and Digital Preservation: Assessing Current Practices at Research Libraries." D-Lib Magazine 17.5-6 (2011). *10.1045/may2011-yuanli*	4	RQ2
S7187	Ayla, Stein Kenfield. "Metadata Documentation Practices at ARL Institutional Repositories." Portal : Libraries and the Academy 19.4 (2019): 667-99. *10.1353/pla.2019.0041*	3	RQ2
S7196	Anyaoku, Ebele N., U. Nwabueze Echedom Anthonia, and Ebikabowei Emmanuel Baro. "Digital Preservation Practices in University Libraries." Digital Library Perspectives 35.1 (2019): 41-64. *10.1108/DLP-10-2017-0041*	4	RQ2

S7203	Awre, Chris, and Richard Green. "From Hydra to Samvera: An Open Source Community Journey." Insights 30.3 (2017): 82-8. *10.1629/uksg.383*	3,5	RQ2
S7208	da Silva Júnior, Laerte Pereira, and Maria Manuel Borges. "Digital Preservation Policies of the Institutional Repositories at Brazilian Federal Universities." The Electronic Library 35.2 (2017): 311-21. *10.1108/EL-09-2015-0170*	4	RQ1
S7231	Dearborn, Carly C., Amy J. Barton, and Neal A. Harmeyer. "The Purdue University Research Repository: HUBzero Customization for Dataset Publication and Digital Preservation." OCLC Systems & Services: International Digital Library Perspectives 30.1 (2014): 15-27. *10.1108/OCLC-07-2013-0022*	4	RQ2
S7246	Hitchcock, Steve, and David Tarrant. "Characterising and Preserving Digital Repositories: File Format Profiles." Ariadne.66 (2011)	2,5	RQ2
S7290	Francke, Helena, Jonas Gamalielsson, and Bjorn Lundell. "Institutional Repositories as Infrastructures for Long-Term Preservation." Information Research 22.2 (2017): 1-16.	4	RQ2
S7344	Termens, Miquel, Mireia Ribera, and Anita Locher. "An Analysis of File Format Control in Institutional Repositories." Library Hi Tech 33.2 (2015): 162-74. *10.1108/LHT-10-2014-0098*	4	RQ2
S7354	Colati, Jessica Branco, Robin Dean, and Keith Maull. "Describing Digital Objects: A Tale of Compromise." Cataloging & Classification Quarterly 47.3 (2009): 326-69. *10.1080/01639370902737331*	3	RQ2 RQ3
S7358	Johnston, Wayne. "Digital Preservation Initiatives in Ontario: Trusted Digital Repositories and Research Data Repositories." Partnership: The Canadian Journal of Library & Information Practice & Research 7.2 (2012): 1-8.	2	RQ2 RQ3
S7360	Kari, Kingdom H., and Ebikabowei E. Baro. "Digital Preservation Practices in University Libraries: A Survey of Institutional Repositories in Nigeria." Preservation, Digital Technology & Culture 45.3 (2016): 134-44. *10.1515/pdtc-2016-0006*	4	RQ2