**Digital curation practices in institutional repositories in South India: a study**

Purpose: The purpose of this study was to identify the digital curation practices in Institutional Repositories (IRs) in South India.

Design/methodology/approach: A voluntary survey was conducted among the IR managers of 23 South Indian IRs, and the response rate was 87%.

Findings: This study found that the active participation of South Indian IRs was only seen in a few digital curation activities. However, Of the 33 digital curation activities analyzed, the active participation of IRs was only seen in ten digital curation activities. The performance of preservation activities was extremely low, and disagreements were recorded by the survey participants towards several digital curation activities. The most disagreed digital curation activities were emulation and cease data curation. All the participants had assigned metadata and allowed file download in their IRs. Raman Research Institute had provided a good number of digital curation services in their IR.

Originality/Value: This is an in-depth study investigating the digital curation practice currently underway in South Indian IRs, and the researcher could not find similar studies in this niche.

**1. Introduction**

Although institutional repositories (IRs) are initiated to break the monopoly of toll-access publishers, they have a greater responsibility, which is research data curation. Even though there is a cost, anyone can access the research results published from toll-access journals on payment of subscription fee. Research data access, however, is not an easy task. With regard to a user, it may not be possible to access the research data which are stored in other departments in a university. Preserving and sharing, in other words, curating the data will be helpful for interrelated disciplines and may eliminate unnecessary intellectual efforts of scholars, saving them time.

Data curation is a collaborative process of librarians, information professionals, faculty and researchers. It requires the application of a range of data expertise, beginning with research planning and extending through phases of long-term stewardship and the re-use of data for new purposes (Karasti, Baker, &Halkola, 2006; Weber, Palmer, & Chao, 2012; Punzalan & Kriesberg, 2017). The e-Science Curation report (Lord and Macdonald, 2003) defined curation as “the activity of managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and re-use’’ (p. 12).

The DCC curation life cycle model of Digital Curation Centre (DCC) has elucidated preservation and curation of data in three categories: full life cycle actions, sequential actions, and occasional actions. Full lifecycle actions contain a set of actions that have to be performed throughout the life cycle of digital objects (Constantopoulos et al., 2009. Sequential actions must be completed in a specific order to simplify the data curation. Occasional actions comprise activities undertaken occasionally or infrequently (Higgins, 2008).

 In India, Open Government Data (OGD) platform was formed in 2010 to increase the transparency of government and all the data collected by ministries/ department for public use are available in this portal. In 2012, the Government of India introduced the National Data Sharing and Accessibility Policy (NDSAP) (Ministry of Science and Technology, 2012). The essence of this policy is that all the non-sensitive data generated using public fund should be shared and enable access for the development of the nation. The Indian Council of Social Science Research (ICSSR) has mandated the pre-arrangement for data preservation to release the project fund, and it is also reserving the right to demand raw data. The policies of UGC which directed the research scholars to submit the e-version of theses, ShodhGanga, the ETD (Electonic Theses and Dissertations) repository of INFLIBNET, and Shdodhgangotri the repository of research in progress also facilitated the access of research findings in India (Tripathi et al., 2017b). However, due to a lack of knowledge, there is an inconsistency about research data management among the authors and librarians in India (Anilkumar, 2018; Shrivastava, 2019). The lack of policies and lack of technical assistance are also the reasons that prevent research data curation becoming a common practice in India. As a new practice, data curation needed practitioners, who have skills for working in an IT-intensive environment and technical skill along with tools and applications for digital curation. In digital curation, policies and procedures are also important (Kim, Warga & Moen, 2013).

A culture of sharing research data through institutional repositories is not yet prevalent in India. However, with the introduction of the NDSAP, it is evident that the Government of India has started such an initiative. Even the research data is scarce; South Indian IRs contained a variety of item types such as audio/video files, interviews, case studies, working papers, presentations, teaching resources etc. (Shajitha & Abdul Majeed, 2018). If not performing digital curation activities, these digital objects may become obsolete. Thus, an IR workflow must be changed to give importance to digital curation. This study tries to understand how the existing digital objects are managed by IRs in South India and whether such digital objects are curated or not. The study is highly relevant as the researcher could not find any evidence of similar studies in this niche. The current research implies that technical expertise is required to perform digital curation. It can be acquired by providing training in digital curation for existing librarians and by including digital curation in LIS education. Creation of discussion forums or groups exclusively for those who are using particular repository software such as Dspace or Eprints and inclusion of computer centre of the institutions in IR management along with the library staff are the other strategies to overcome this challenge. The current study investigates the digital curation practices in IRs in South India. South India is the southern part of India encompassing the five states of Kerala, Tamil Nadu, Karnataka, Telangana, Andhra Pradesh and the Union Territories of Lakshadweep, Andaman and Nicobar, and Puducherri.

**2. Review of Literature**

Nowadays, the position of the print materials is being replaced by digital objects, and it necessitates digital curation. However, the digital curation practice among the authors was found to be low. This was clear from the survey conducted among the Science and Mathematics faculty at California Polytechnic State University, San Luis Obispo. The study revealed that many small science researchers regularly had back up their data on an office computer or external hard drive. Any failure of the hardware would lead to data loss (Scaramozzino, Ramirez & McGaughey, 2012). Likewise, lack of finance and lack of skill prevented the authors from curating their research data. Lack of tools, Intellectual property rights and dark data were the other barriers of data curation (Heidron, 2008).

 As an information manager, librarians are the best choice to perform digital curation. Libraries can curate research data, and the research data is crucial for the development of the country because libraries have the skill sets, longevity, and most of the infrastructure required to execute data curation activity (Heidorn, 2011). The author observed that if the libraries are not ready to do this task that may lead to the new type of institution to curate the data.

However, several issues are associated with the data curation. Lee and Stvilia (2017) found that lack of best practices, norms, skilled persons and funding were adversely affected the data curation. The researchers conducted a semi-structured interview on IRs of 13 research Universities in the USA and analyzed the activity structure of IR. It also explored the metadata schema, identifier schema, controlled vocabularies, tools, policies, rules, and norms adopted by IRs. The study found that data curation services restricted as per the supporting features of IR software and researchers requirements imposed by publishers. Both users and IRs were frightened into adopting a new system rather than the familiar one. As per the findings of Shen and Varvel (2013), human technological and financial aspects influence the success of data curation practices. The investigators described that social and technical readiness was required for the implementation of data management services. Technical aspects included hardware and software needed for data conservation and social aspects included administration, consultation, and customer services. Such an observation was made by the researchers in light of the experiences of establishing a data management service at John Hopkins University through the JHU data archive.

Due to the issues quoted above, most institutions may not be able to engage in data curation activities. Data Curation Network (DCN) came up with a solution to these problems. DCN would help the partner institution in data curation activities by collaboratively sharing expert data curation staff. The implementation plan of DCN was based on their planning face research conducted from 2016-17 (Johnston et al., 2018a).

Lage, Losoff and Maness (2011) investigated the data curation practices and existing support for science faculty and students at the University of Colourado Boulder. It is found that the researcher’s receptivity to a library role in data curation did not correlate with the level of need for assistance or the existence of obstacles.

Depending on the nature of the data, there is a difference in digital curation activities. Palmer et al. (2007) noted that large volume and homogenous data (astronomy and seismology) are more curated than small and numerous, specialized heterogeneous data. So a mix of data management solutions is required to address this problem. In the observation of Witt (2009), flexible and highly scalable infrastructures are inevitable to accommodate massive and heterogeneous datasets. The author continued that meaningful data collection, proper metadata, point of access between human and machine, unique identifier, and policies were the main challenges of data curation. The study put forth a distributed Institutional repository of the Purdue Libraries as an example to describe how to conquered the mentioned challenges. It was distributed in three repositories, and these three repositories were connected with web services and other middleware.

Just as the nature of data, influences the digital curation activity, the disciplinary differences also influence the curation activity. Akers and Doty (2013) studied the research data management practices of faculty at Emory University and found that basic science researchers were likely to keep a high volume of digital data in the university-based servers than other researchers. However, social scientists were least likely to share their data via email with others due to the highly confidential nature of the data, and they required a more secure method of transmission. These issues were also relevant to medical science researchers.

Sharing of experience of institutions about the ways they have performed data curation activities in their repositories can provide a sense of direction and encourage those who do not engage in such activities. Choudhury (2008) demonstrated the current project of John Hopkins University to perform the data curation activities in its Institutional Repository. In ordered to develop data curation prototype system which connects digital archiving and electronic publishing, JHU Sheridan libraries and National virtual observatory worked with American Astronomical Society. To achieve a compound object publication model, this project used open archive initiative object reuse and exchange (OAI- ORE) Protocol.

 It is clear from previous studies that Indian institutions are lagging behind developed countries in providing data curation services. Tripathi et al. (2017a) conducted a study on 47 central Universities and found that librarians were in the initial stage of implementing such services. The study also suggested a model for university libraries in India to implement Research Data Management (RDM) services. The same observation also made by Kaushik (2017) and Anilkumar (2018). Anilkumar (2018) analysed the perception of South Indian library professionals in RDM and found that their perception was very low. Kaushik (2017) also tried to find out the perception of library professionals in data curation activities. As per this study, most of the Indian LIS professionals had been aware of this concept for a year. Lack of adequate skill was highlighted as the major reason for the non-association in data curation activities by the respondents. The findings also showed that most of the institutions had a lack of data curation policy. However, the study of Tripathi et al (2017b) revealed that faculty and researchers in Baba Saheb Bimrao Ambedkar University and Jawaharlal Nehru University had a culture of data sharing, and the data storage problem was faced by most of them. This study points out the need for libraries in India to indulge in data curation activities.

 Shivarama et al. (2016) described the digital curation strategies for information management and the study suggested that the information generated system has to be designed with the aim of long term preservation, and this standardisation should arise from ground level to the top management of organizations. Shrivastava and Gupta (2018) analysed the Indian research data repositories (RDR) available in re.dataorg and found that of the 43 registered Indian RDR, only nine repositories were functional. All the RDR were open access except for one, and the repositories had a lack of repository standards or certification.

 As per the e-Science Curation report, curation is a broader term that encompasses the activities like archiving and preservation (Lord and Macdonald, 2003). Hence, the studies which discussed the preservation methods adopted by institutions in countries like India, where curation is less familiar, are also important. Katre (2011) compared the digital preservation programs launched by India and USA and found that the difference between these two programs mainly on its origin and driving force. Ganesan et-al (2015) explored the issues related to digital preservation in India. The study pointed out that scanning of the documents is performed according to the Indian copyright act 1957, and the issues like removal of documents which is requested by the authors have to be addressed. The other issues were the selection of software, file format, metadata, and accuracy of Optical Character Recognizer (OCR). Kumar (2014) studied about the digital preservation practices in engineering institutional libraries in Andhra Pradesh. The study revealed that refreshing and emulation were the most used digital preservation techniques by the librarians, and most used format was HTML. The constraints faced by the librarians were intellectual properties rights issue followed by a lack of interest of users in digital services.

Most of the studies in India addressed the needs of digital curation, problems related to curation activities and perception of library professionals towards digital curation. However, the digital curation activities currently taking place in Indian repositories are yet to be discussed. The current research provides a clear picture of the digital curation practices in IRs in South India.

**3. Objectives and Methods**

The purpose of the study was to explore the digital curation activities offered by the IRs in South India. The main objectives of the present study are as follows:

1. To find out the tools used by the IRs in South India to curate the digital content.
2. To identify the digital curation practices in IRs in South India.
3. To find out whether the IRs have any disagreement towards the digital curation activities.

The identification of IRs in South India was necessary to satisfy the objectives of the study. For this purpose, the investigator relied on the Registry of Open Access Repositories (ROAR), and Directory of Open Access Repositories (OpenDOAR) and identified 23 operational IRs at 22 institutions (Table 1).

**Table 1** **Institutional Repositories in South India**

|  |  |  |
| --- | --- | --- |
| **Sl No** | **Name of Institutional Repository** | **Name of Institute** |

|  |  |  |
| --- | --- | --- |
| 1 | Indian Academy of Sciences: Publications of Fellows | Indian Academy of Sciences, Bangalore |
| 2 | ePrints@IISc | Indian Institute of Science, Bangalore |
| 3 | Eprints@CMFRI | Central Marine Fisheries Research Institute (CMFRI), Cochin |
| 4 | ePrints@UoM | University of Mysore |
| 5 | Open Access Repository of ICRISAT | International Crops Research Institute for the Semi Arid Tropics ([ICRISAT ) Hyderabad](http://www.icrisat.org/) |
| 6 | Dspace@IIA, | Indian Institute of Astrophysics, Bangalore |
| 7 | NAL-IR | [Information Centre for Aerospace Science and Technology](http://www.icast.org.in/)(ICAST), Bangalore |
| 8 | RRI  Digital Repository | [Raman Research Institute, Bangalore](http://www.rri.res.in/) |
| 9 | Dyuthi | Cochin University of Science and Technology (CUSAT), Kerala |
| 10 | Open Access Digital Repository of Ministry of Earth Sciences | Indian National Centre for Ocean Information Services (INCOIS), Hyderabad |
| 11 | IR@CECRI | CSIR-Central Electrochemical Research Institute, Thamizhnad,  |
| 12 | RAIITH | IIT Hyderabad |
| 13 | etd@IISc | Indian Institute of Science Bangalore |
| 14 |  Mahatma Gandhi University Online Theses Library | Mahatma Gandhi University, Kerala |
| 2315 | University of Agricultural Sciences, Online theses | University of Agricultural Sciences, Dharwad, Karnataka |
| 16 | [Eprints @MDRF](http://mdrf-eprints.in/) | [Madras Diabetes Research Foundation, Chennai](http://mdrf.in/) |
| 17 | ePrints@NIRT | [National Institute for Research in Tuberculosis, Chennai](http://www.nirt.res.in/) |
| 18 | Dspace@IIMK,  | Indian Institute of Management, Kozhikode |
| 19 | Librarians'Digital Library,  | Documentation Research and Training Centre (DRTC), Bangalore |
| 20 | Dspace@IMSc Library | Institute of Mathematical Sciences- Chennai |
| 21 | NAARM Digital Repository | National Academy of Agricultural Research Management, Hyderabad |
| 22 | ePrints@ATREE , | [Ashoka Trust for Research in Ecology and the Environment- Bangalore](http://atree.org/) |
| 23 | eprints@BU | Bangalore University |

An online survey was conducted among the IR managers of 23 institutional repositories, and voluntary participation was encouraged. The questionnaire was prepared based on previous studies. To prepare the questions about tools used for digital curation, the studies of Park & Tosaka (2010), Andrade & Baptista (2015), and Lee and Stvilia (2017) were used. The study of Hudson-Vitale, et al. (2017) was useful for the preparation of the questions related to support for curation activities, and all these questions were Likert scale type’s questions. An online questionnaire was prepared by using a Google form and circulated through email. The questionnaire was not anonymous. Data were collected from the IR managers subject to the condition that the data would be used for academic purpose only and kept confidential. Maintaining anonymity is an important ethical matter followed in quantitative studies (Allen, 2017). However, the study used the non-anonymous questionnaire since it does not address any sensitive data.

In total, 60 remainders were sent through email and institutions were contacted many times by telephone. Though the process was lengthy, participants responded to the questionnaire and recorded their remarks towards open-ended question also. To clarify the doubts about the response, the participants have been again contacted by the researcher, and all of them have responded. Of the 23 institutional repositories identified, 20 IRs of 19 institutions responded, and the response rate was 87% (Table 2). The data collected through the survey were coded and analysed with the help of Microsoft Excel, and it was also used for the generation of figures. The technique applied to analyse the data was percentage analysis.

**Table 2** **List of Institutional Repositories participated in the survey**

|  |  |  |
| --- | --- | --- |
| **Sl No** | **Name of Institutional Repository** | **Name of Institute** |

|  |  |  |
| --- | --- | --- |
| 1 | Indian Academy of Sciences: Publications of Fellows | Indian Academy of Sciences, Bangalore |
| 2 | ePrints@IISc | Indian Institute of Science, Bangalore |
| 3 | etd@IISc | Indian Institute of Science Bangalore |
| 4 | Eprints@CMFRI | Central Marine Fisheries Research Institute (CMFRI), Cochin |
| 5 | Open Access Repository of ICRISAT | International Crops Research Institute for the Semi Arid Tropics ([ICRISAT ) Hyderabad](http://www.icrisat.org/) |
| 6 | Dspace@IIA, | Indian Institute of Astrophysics, Bangalore |
| 7 | NAL-IR | [Information Centre for Aerospace Science and Technology](http://www.icast.org.in/)(ICAST), Bangalore |
| 8 | RRI  Digital Repository | [Raman Research Institute, Bangalore](http://www.rri.res.in/) |
| 9 | Dyuthi | Cochin University of Science and Technology (CUSAT), Kerala |
| 10 | Open Access Digital Repository of Ministry of Earth Sciences | Indian National Centre for Ocean Information Services (INCOIS), Hyderabad |
| 11 | IR@CECRI | CSIR-Central Electrochemical Research Institute, Thamizhnad,  |
| 12 | RAIITH | IIT Hyderabad |
| 13 |  Mahatma Gandhi University Online Theses Library | Mahatma Gandhi University, Kerala |
| 2314 | University of Agricultural Sciences, Online theses | University of Agricultural Sciences, Dharwad, Karnataka |
| 15 | [Eprints @MDRF](http://mdrf-eprints.in/) | [Madras Diabetes Research Foundation, Chennai](http://mdrf.in/) |
| 16 | Dspace@IIMK,  | Indian Institute of Management, Kozhikode |
| 17 | Dspace@IMSc Library | Institute of Mathematical Sciences- Chennai |
| 18 | NAARM Digital Repository | National Academy of Agricultural Research Management, Hyderabad |
| 19 | ePrints@ATREE , | [Ashoka Trust for Research in Ecology and the Environment- Bangalore](http://atree.org/) |
| 20 | eprints@BU | Bangalore University |

**4. Results and Discussion**

**4.1 Tools used for Digital Curation**

Lee and Stvilia (2017) described IR software, metadata schemas, identifier schemas, controlled vocabularies and applications as tools used for digital curation.

***IR software***

As observed in the study about content growth of IRs in South India (Shajitha & Abdul Majeed, 2018), this research also found that the Eprints software is used by the majority of the institutional repositories (60%) in South India. DSpace is in the second position with 30% usage and other software used by two institutional repositories. Indian Institute of Sciences (IISc) Bangalore, Karnataka is the first South Indian institution to establish an Institutional repository in India,. IISc used Eprints as their software. Other South Indian Institutions may have adopted IISc as their role model. Most of the South Indian IRs are found to be from Karnataka State. This may be the reason for the high use of EPrints software among South Indian IRs.

***Metadata***

Qualified Dublin Core metadata was used by the majority of the respondents (75%), and a large-scale difference was observed between the usage of Qualified Dublin Core and the other metadata. Standard Eprints metadata stands behind the qualified Dublin core with 15% usage, and 10% of respondents followed Unqualified Dublin core. However, 10% of respondents followed Adapted Dublin core in their Institutional Repositories. No other metadata standards exceeded 10% (Table 3). Park and Tosaka (2010) also noticed that the use of unqualified Dublin core was less than that of qualified Dublin core. The metadata standards like MODS (Metadata Object Description Schema), METS (Metadata Encoding and Transmission Standard), TEI (Text encoding initiative), and COBISS (Co-operative Online Bibliographic System and Service) were not used by any IRs. Except for Bangalore University and University of Agricultural Sciences, all the participants were using a single metadata schema. Bangalore University followed qualified Dublin core and unqualified Dublin Core, while University of Agricultural Sciences followed Qualified Dublin core, Archival metadata, print on demand metadata, standard Eprints metadata and Adapted Dublin core. This result proved that the use of diversified metadata schema was less in South Indian IRs and this result is in contrast with the previous studies outside India (Park & Tosaka, 2010; Andrade & Baptista, 2015; Moulaison, 2015; Lee and Stvilia, 2017).

***Controlled Vocabulary***

This study found that 55 % of the respondents were familiar with the use of controlled vocabulary in IRs, while 45% of the survey participants did not use it. Hence the question about the name of controlled vocabulary used by the IR was restricted to 11 participants, who claimed that they had used controlled vocabulary schema. This finding supports the conclusion of Steele & Sump- crethar (2016) that the controlled vocabulary was not considered important by the libraries. Lee and Stvilia (2017) also found that usage of controlled vocabularies is very less among Institutional Repositories.

It was observed that the Library of Congress Subject Heading (LCSH) and Dewey Decimal Classification (DDC) were the most used controlled vocabulary (36.4% each) by the respondents. The in- house controlled vocabulary was used by 27.3% of respondents. However, 18.2% of the respondents used LCC (Library of Congress Classification), and one institution used other controlled vocabulary, NASA thesaurus for their IR (Table 3). At the same time the controlled vocabularies like AAT (Art and Architecture Thesaurus), TGM (Library of Congress Thesaurus for Graphic Materials), TGN (Getty Thesaurus of Geographic Names), and ULAN (Getty Union List of Artists Names) were not used by single IR.

***Persistent Identifier***

Digital preservation handbook (Digital preservation Coalition, n. d.) defined persistent identifier (PID) as a “long-lasting reference to a digital resource" (para. 2). This study found that all the respondents were assigned PID (90%) for the items in their IRs except for two. Hence, the question with regards to the type of identifier system used by IRs was limited to 18 respondents, who had assigned PID to its digital resources. As shown in Table 3, 61.1% survey respondents used URI (Uniform Resource identifier) followed by Handle (44.4%). Only one respondent used DOI (Digital Object Identifier), and the IR manager commented that they are planning to use a Handle identifier. The PURL (Persistent Uniform Resource Locator) identifier was used by a single respondent, while ARKS (Archival Resource Key Identifiers) was not used by any IRs. By default, the URI identifier comes with the software Eprints, and most of the South Indian IRs were using Eprints software. Likewise, the Dspace software provides Handle identifier. These things may be the reason behind the widespread use of URI and Handle among the South Indian IRs. Lee and Stvilia (2017) found that the institutions are using the identifier schema, which is embedded in the IR software, and they are afraid to adopt new system according to their needs and goals.

Table 3- Results regarding usage of metadata, controlled vocabularies, and identifier schema by the IRs in South India

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Metadata**  | **Frequency****(N=20)** | **Percent** | **Controlled Vocabulary**  | **Frequency****(N=11)** | **Percent** | **Persistent Identifier (N=18)** | **Frequency****(N=18)** | **Percent** |
| Qualified Dublin Core | 15 |  75% | LCSH | 4 | 36.4% | URI | 11 | 61.1% |
| Standard Eprints Metadata | 3 |  15% | DDC | 4 | 36.4% | Handle | 8 | 44.4% |
| Unqualified Dublin Core | 2 |  10% | In house built | 3 | 27.3% | DOI | 1 | 5.6% |
| Adapted Dublin Core | 2 |  10% | LCC | 2 | 18.2% | PURLS | 1 | 5.6% |
| Archival Metadata | 1 |  5% | Other (NASA thesaurus) | 1 | 9.1% | ARKS | 0 | 0% |
| Print-on-demand Metadata | 1 |  5% |  |  |  |  |  |  |
| Others Standards | 1 |  5% |  |  |  |  |  |  |

***Applications used to manage the digital content of the Institutional Repositories***

This study investigated the applications used by the institutional repositories to manage the digital content. It was found that half of the surveyed IRs (50%) did not use any application to manage the digital content. Some participants (20%) responded that they used apache solr in their repository. Apache solr is an open-source search platform that automatically comes with the Dspace, and four IRs that claimed to use the apache solr were also using Dspace. At the same time, 15% of the respondents used altmetrics (views, downloads , discussion in social media and Mendeley) in their IRs. Altmetrics measures the impact of digital resources based on mentions on social media networks, citations on Wikipedia and in public policy documents. It also measures discussions on research blogs, mainstream media coverage, bookmarks on reference managers like Mendeley and peer reviews on faculty of 1000 (https://www.altmetric. com/about-altmetrics/what-are-altmetrics/). Very few respondents used the file information toolset and the simple text editor (10% each) for handling their digital content. However, one respondent (5%) used Openrefine (5%) for data clean up. Other applications (10%) were also used by survey participants (Table 4). CMFRI commented that they had used the tools such as VSDC Video converter to convert the video format, mega and Google Drive for storage, and we transfer for sending files. Likewise, CUSAT commented that in order to avoid the spelling mistakes, they used to copy the author name, title, and keywords from Pdf file for metadata creation and for this, they were using Pdf password remover.

Table 4. Applications used to manage the content in the repository,

|  |  |  |
| --- | --- | --- |
|  Tools | Frequency ( N=20) | Percent |
| No tool | 10 | 50% |
| Apache solr | 4 | 20% |
| Altmetrics | 3 | 15% |
| File information tool set | 2 | 10% |
| Simple text editor | 2 | 10% |
| Openrefine | 1 | 5% |
| Other tools | 2 | 10% |

**4.2 Support for Curation activities**

Data curation services included a variety of activities, and that is grouped into the following five different aspects of data curation (Hudson-Vitale, Cynthia et al., 2017)

* Ingest
* Appraisal
* Processing and review
* Access
* Preservation

Of the 20 IRs who participated in the survey, two of them had not responded to all questions in the survey questionnaire, which are related to support for curation activities. Hence, the response of 18 participants had only been considered for the analysis of support for curation activities. All the questions related to support for digital curation activities were ‘Likert’ scale model, and it asked the respondents to rate the status of activities in their IR on a scale of 1 to 5 where 1= provide the service, 2= intended to provide the service shortly, 3= would like to provide the service but unable currently, 4= would not like to provide the service, 5= not sure.

***Support for Ingest activity***

The ingest activities included digital curation activities such as authentication, the chain of custody, deposit agreement, documentation, file validation, and metadata (Hudson-Vitale, Cynthia et- al., 2017).

Table 5. Support for ingest activity

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Ingest Activities (N=18) | Provide | Intended to provide shortly | Would like to provide but unable currently | Would not like to provide | Not Sure |
| Authentication | 9(50%) | 1(5.6%) | 2 (11.1%) | 2 (11.1%) | 4 (22.2%) |
| Chain of Custody | 6 (33.3%) | 1 (5.6%) | 4 (22.2%) | 3 (16.7%) | 4 (22.2%) |
| Deposit agreement of depositor | 5 (27.8%) | 1 (5.6%) | 2 (11.1%) | 2 (11.1%) | 8 (44.4%) |
| Documentation | 5 (27.8%) | 3 (16.7%) | 4(22.2%) | 2 (11.1%) | 4 (22.2%) |
| File validation  | 4 (22.2%) | 1 (5.6%) | 5 (27.7%) | 3 (16.7%) | 5 (27.7%) |
| Metadata | 18(100%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |

Table 5 indicates that all (100%) participants provided metadata to the repository contents for the retrieval purposes. Authentication was the next supported ingest activity by the survey participants (50%). Nine years ago, a study of Association of Research Libraries (ARL) member institutions found that 83.3 % of the repositories were using authentication mechanism for depositors to ensure the quality of submitted content (Li & Banach, 2011). It is disappointing that the South Indian IRs have not yet reached that level.

Chain of custody is the process of intentional recording of metadata about who created the file, date of the last editing etc., which was supported by 33.3% of respondents. Fritz (2019) observed that the chain of custody is one of the things IRs need to perform to facilitating access and safeguarding information. The activities like the deposit agreement of the depositor and documentation activities were supported by 27.8% respondents each. Johnston et al. (2018b) found that 80.2% of the researchers in the USA are engaged in documentation activities. Digital objects become more understandable and identifiable by doing documentation, and it gives necessary information about the object. Therefore the documentation activities of South Indian IRs need improvement.

At the same time, 22.2% of respondents were actively engaged in file validation. A study among 13 large universities in the USA, it is found that the IRs had been using software tools such as DROID and PRONOM for file identification and validation (Lee & Stvilia, 2017)

Unfortunately, a reasonable percentage of survey participants had recorded their disagreement, such as would not like to provide or not sure to provide ingest activities in their IRs (Fig. 1). Deposit agreement of depositor was one of the most disagreed ingest activity by the participants (10). Perhaps this disagreement will be due to their concern that the implementation of such things would repel the contributors from IR. At the same time, in 2010, 72.4% of the IR of ARL member institutions had an agreement with the depositor (Li & Banach, 2011) and the percentage was increased to 77.5 among ARL institutions in 2017 (Hudson-Vitale, Cynthia et- al., 2017). Likewise, 67% of the IRs of Nigerian Universities had agreements with the content contributors (Kari and Baro, 2016).

Fig.1 Disagreement of IRs towards ingest activities

***Support for Appraisal activities***
The appraisal activity is the activity in which data managers are required to  “evaluate data and select for long-term curation and preservation”(DCC, n.d.). The appraisal process is a crucial part of digital curation. For the appraisal, some repositories seek help from subject experts, and some rely on the resource producers themselves (Niu, 2014). This study analysed three digital curation appraisal activities; these are rights management, risk management and selection.

Table 6 indicates that more than half of the survey participants (55.6%) performed the selection activity to ensure whether the digital content is suitable for their IRs. The risk management was done by 27.7% of the respondents. At the same time, only 22.2% of the respondents performed rights management. The lack of manpower is a significant challenge faced by the repositories in India (Roy, Mukhopadhyay, and Biswas, 2012).The reason why most south Indian IRs do not engage in risk management and rights management might be due to the inadequacy of the workforce.

The study of DeRidder and Helms (2016) revealed that institutions collect rights information from the content providers. These were information on access and reuse restrictions, copyright information, intellectual property right information, rights to make copies and derivatives, rights to preserve and migrate content, digitization permissions, and rights to make representations and metadata.

Table 6 Status of appraisal activities in South Indian IRs

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Appraisal activity (N=18) | Provide | Intended to provide shortly | Would like to provide but unable currently | Would not like to provide | Not Sure |
| Rights Management | 4 (22.2%) | 3 (16.7%) | 3 (16.7%) | 2 (11.1%) | 6 (33.3%) |
| Risk Management | 5 (27.7%) | 2 (11.1%) | 4 (22.2%) | 0 (0.0%) | 7 (38.9%) |
| Selection | 10 (55.6%) | 2 (11.1%) | 3 (16.7%) | 0 (0.0%) | 3 (16.7%) |

It was found that most have agreed to engage in appraisal activities, even if most have not been involved in activities like rights management and risk management (Figure 2).

Figure 2 Disagreement of IRs towards appraisal activities.

 ***Support for Processing and Review activities***
The current study investigated the status of nine digital curation processing and review activities in the Institutional Repositories in south India. These are, code review, contextualize, data cleaning, file format transformation, interoperability, peer review, software registry, transcoding and persistent identifier.

Table 7 summarizes the responses of survey participants about processing and review activities in the South Indian IRs. Assigning persistent identifier (88.9%) was the most supported processing and review activity by the participants, and 55.6% of the respondents had supported to contextualize the digital content. The methods adopted by the UK Higher Educational Institutions to contextualize the deposited data include linking datasets to related publications, connecting datasets to institutions or organizations, linking datasets to the author institutional profile, linking datasets to the author, and linking datasets to a related subject collection (Pham, 2018).

The data cleaning was provided by 44.4% of the respondents. Very interestingly, 33.3% of respondents managed software registry in order to maintain copies of the modern and obsolete version of the software, and it was higher than that of the ARL member libraries institutions (8.3%) (Hudson-Vitale, Cynthia et- al., 2017). At the same time code review and Interoperability were provided by 27.7% and 22.2% respondents respectively. The activities such as peer review, transcoding, transform file in to open nonproprietary file format were the least provided activity by the respondents (16.7% each). However, in the UK, 55 % of the higher education institutions used non-proprietary or open file format as a preservation solution (Pham, 2018).

 As illustrated in Figure 3, all survey respondents agreed to assign persistent identifier while most respondents (12) showed their disagreement towards the peer review process. Callicott, B. B., Scherer, D., and Wesolek, A. (Eds.) (2016) argued that IR managers have a responsibility to the scholars at their institution and an IR can publish the scholars rejected work also. In 2010, 20% of the Association of Research Libraries (ARL) member institutions had used peer review system with their IR (Li & Banach, 2011). However, a survey of Hudson-Vitale, Cynthia et- al., (2017) found that 52 % of the ARL member institutions disagreed with the peer review system.

Table 7 Status of processing and review activities in South Indian IRs

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Processing and Review activities (N=18) | Provide | Intended to provide shortly | Would like to provide but unable currently | Would not like to provide | Not Sure |
| Code Review  | 5 (27.7%) | 0 (0.0%) | 4 (22.2%) | 2 (11.1%) | 7 (38.9%) |
| Contextualize | 10(55.6%) | 4 (22.2%) | 2 (11.1%) | 1 (5.6%) | 1 (5.6%) |
| Data cleaning  | 8 (44.4%) | 3 (16.7%) | 2 (11.1%) | 1 (5.6%) | 4 (22.2%) |
| Transform file in to open nonproprietary file format | 3 (16.7%) | 5 (27.8%) | 1 (5.6%) | 1 (5.6%) | 8 (44.4%) |
| Interoperability  | 4 (22.2%) | 1 (5.6%) | 3 (16.7%) | 2 (11.1%) | 8 (44.4%) |
| Peer review  | 3 (16.7%) | 2 (11.1%) | 1 (5.6%) | 3 (16.7%) | 9 (50%) |
| Software registry  | 6(33.3%) | 0 (0.0%) | 2 (11.1%) | 1 (5.6%) | 9 (50%) |
| Transcoding  | 3 (16.7%) | 4 (22.2%) | 1 (5.6%) | 0 (0.0%) | 10 (55.6%) |
| Persistent Identifier | 16 (88.9%) | 2 (11.1%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |

Figure 3. The disagreement of IRs towardsprocessing and review activities

***Support for Information/services to accelerate the information access in IR***

This study evaluated the attitude of South Indian IRs towards eight digital curation access activities. These are the contact information of depositor/data creator, data citation, discovery services, embargo, file download, metadata brokerage, full-text indexing, and terms of use.

The results disclosed that all the participants had allowed their users to download files. Likewise, a high percentage had supported to provide embargo facility (72.2%) in their IRs. The embargo facility was also high in Spanish repositories (Serrano-Vicente, Melero and Abadal, 2018). More than half of the respondents (55.6 %) were provided contact information of depositor to the users, and this percentage was higher than that of Social Science data repositories in the US (28.6 %) (Yoon and Tibbo, 2011). This facility would help the users to ask doubt regarding the study and allow requesting the other works of an author if required. Full-text indexing and terms of use were provided by half of the IRs (50% each). In total, 44.4% of the participated IRs had discovery services to enhance the retrieval functionality, and 38.9% of the participants displayed recommended bibliographic citation of data to its users. Yoon and Shultz (2017) found that the US academic libraries provided information about data reuse (49.2%) and how to properly cite the data (37.8%) in their library websites. Metadata brokerage is the active dissemination of metadata, and only 22.2% of the respondents offered metadata brokerage in their IRs (Table 8). This result contrasts with the findings of the study of Hudson-Vitale, Cynthia, et al., (2017), which reported that metadata brokerage is provided by 62.5% of the ARL libraries.

Table 8. Support for access activities

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Access activities (N=18) | Provide | Intended to provide shortly | Would like to provide but unable currently | Would not like to provide | Not Sure |
| Contact information of depositor/data creator | 10 (55.6%) | 1 (5.6%) | 0 (0.0%) | 2 (11.1%) | 5(27.8%) |
| Data citation | 7 (38.9%) | 1 (5.6%) | 4 (22.2%) | 2 (11.1%) | 4 (22.2%) |
| Discovery services | 8 (44.4%) | 4 (22.2%) | 2 (11.1%) | 2 (11.1%) | 2 (11.1%) |
| Support embargoes and /or restricted access condition | 13(72.2%) | 2 (11.1%) | 0 (0.0%) | 2 (11.1%)) | 1 (5.6%) |
| Allow file download | 18(100%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |
| Metadata brokerage  | 4 (22.2%) | 5 (27.8%) | 2 (11.1%) | 1 (5.6%) | 6 (33.3%) |
| Full text indexing | 9 (50%) | 5 (27.8%) | 1 (5.6%) | 1 (5.6%) | 2 (11.1%) |
| Terms of use  | 9 (50%) | 4 (22.2%) | 0 (0.0%) | 1 (5.6%) | 4 (22.2%) |

Figure 4. The disagreement of IRs towards access activities

As figure 4 illustrated above, everyone has agreed with the file download. Further, most respondents (15 each) recorded their agreement towards full-text indexing and embargo. However, metadata brokerage and contact information of a depositor (7 each) were the other access activities that were disagreed by a good number of respondents.

***Support for preservation activities***

This study assessed the digital curation preservation activities, which takes place in South Indian IRs. Seven digital curation preservation activities were considered for the study, which includes cease data curation, emulation, file audit, migration, repository certification, secure storage, and versioning.

Table 9. Support for preservation activities

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  Preservation activities (N=18) | Provide | Intended to provide shortly | Would like to provide but unable currently | Would not like to provide | Not Sure |
| Cease data curation | 2 (11.1%) | 0 (0.0%) | 1 (5.6%) | 2 (11.1%) | 13 (72.2%) |
| Emulation | 2 (11.1%) | 1 (5.6%) | 0 (0.0%) | 1 (5.6%) | 14 (77.8%) |
| File audit | 4 (22.2%) | 1 (5.6%) | 2 (11.1%) | 1 (5.6%) | 10 (55.6%) |
| Migration | 5 (27.8%) | 3 (16.7%) | 0 (0.0%) | 1 (5.6%) | 9 (50%) |
| Repository Certification | 2 (11.1%) | 2 (11.1%) | 2 (11.1%) | 2 (11.1%) | 10(55.6%) |
| Secure storage | 8 (44.4%) | 1 (5.6%) | 1 (5.6%) | 1 (5.6%) | 7 (38.9%) |
| Versioning  | 3 (16.7%) | 3 (16.7%) | 2 (11.1%) | 1 (5.6%) | 9 (50%) |

It has been observed that relatively least percentage of the respondents supported the preservation activities. The secure storage was the most supported digital curation preservation activity (44.4%). Secure storage is an important data curation activity and 75% of the researchers in six Universities in the USA recorded that secure storage happened for their data (Johnston et al, 2018b)., . In all, 27.8% of the respondents performed migration, which means they have transformed the obsolete file format into a new format. In Nigeria also, 50 % of the IRs have used migration as their preservation strategy (Kari and Baro, 2016). About 22.2 % of the respondents performed file auditing. No other preservation activity exceeded 20 % (Table 9).

Figure 5 Disagreement of IRs towardspreservation activities

The result revealed that except secure storage, all digital curation preservation activities were not supported by half of the respondents; whereas more than seven respondents (40%) out of 18 had recorded their disagreement, such as would not like to provide or not sure towards all digital curation preservation activities. Emulation is one of the important digital preservation strategies and cease data curation means the provision of terminating data access by providing only metadata. Surprisingly, emulation and cease data curation (15 each) were the most disagreed digital curation preservation activities. This result must be read in conjunction with the finding that 30 per cent of the repositories in Africa were engaged in emulation (Anyaoku, Echedom, and Baro, 2019). The activities such as repository certification (12) and file audit (11) were the other most disagreed activities by the respondents (Figure 5).

 The survey investigated the status of 33 digital curation activities which are grouped into ingest, appraisal, processing and review, access and preservation activities in South Indian IRs. It has been observed that the most provided digital curation services were access activities. The access activities included eight activities such as the contact information of depositor/data creator, data citation, discovery services, embargo, file download, metadata brokerage, full-text indexing and terms of use. All the access activities were provided by more than one-third of the IRs except metadata brokerage. All the participants have provided two or more of these services and 66.7% of the libraries have provided four or more of these services. This result is in contrast with the finding of the survey conducted in ARL libraries which revealed that most provided curation services are ingest activities (Hudson-Vitale, Cynthia, et al., 2017).

The next supported digital curation services were ingest activities, which include authentication, chain of custody, deposit agreement of depositor, documentation, file validation and metadata. All the libraries have provided one or more of these services, and 55.6% of the participants have presently provided two or more of these services. However, performances of IRs with respect to preservation activities were extremely low. Preservation activities include cease data curation, emulation, file audit, migration, repository certification, secure storage, and versioning. Only 38.9 % of the libraries have provided one or more of these preservation services.

As figure 6 illustrates, all the respondents had assigned metadata and allowed file download in their IRs. Likewise, the high percentages had assigned a persistent identifier (89%) to its digital document. Nearly three- fourth of the respondents had supported embargoes and/or access restricted condition (72.2%) in their IRs. At the same time, more than half of the respondents have performed the selection process, used metadata to link the dataset to related publications, and provided the contact information of a depositor to its users (55.6% each). The digital curation activities such as authentication, full-text indexing, and terms of use were provided by half of the IRs (50% each). However, 23 digital curation activities out of 33 were not performed by more than half of the IRs and the active participation of IRs were only seen in ten digital curation activities.

It was observed that the Raman Research Institute (RRI) had provided 24 digital curation services in their IR. The digital curation services provided by the IRs of Mahatma Gandhi University (22) and Indian Institute of Astrophysics (20), Bangalore University (18), and CMFRI Cochin (16) were also not less (fig.7). Compared to universities, the number of staff at other research institutions in South India is low. Likewise, the participation of library professional in IR management was very less (0-20%) among half of the IRs in South India. However, the participation was 100% in the Raman Research Institute (Shajitha, 2020). The cooperation and wholehearted involvement from the entire library team might have given this output to RRI. Universities such as Mahatma Gandhi University, Bangalore University and Cochin University of Science and Technology could do better if they included more staff in IR management. The findings of Shajitha (2020) also explored that more than one-third of the south Indians IRs had no professionals with job titles like IR manager, metadata specialist, subject specialist, and data curator. The lack of peoples with technical expertise may also have affected the performance of IRs.

Figure 6. The digital curation activities provided by IRs in South India

Fig. 7 Performance of IRs in digital curation activities

The participants have recorded their disagreement, such as would not like to provide or not sure towards some activities. Out of 18 respondents, 15 recorded their disagreement towards emulation and cease data curation. Likewise, repository certification and peer review (12 each) were the next most disagreed activities. At the same time, 11 respondents recorded their disagreement towards file audit. More than half of the IRs did not like to provide or not sure about versioning, migration, transcoding, software registry, interoperability, and deposit agreement of a depositor (Figure 8)

Figure 8 Disagreement of IRs towards digital curation activities

**5 Conclusions**

The current study tried to explore the digital curation practices in South Indian IRs, and it proves that although digital curation activities are new to a developing country like India, they have not completely abstained from engaging in such activities. The most provided digital curation services were access activities. All the participants had assigned metadata and allowed file download in their Institutional repositories. The high percentages had assigned a persistent identifier to its digital document. It was observed that the Raman Research Institute had provided a good number of digital curation services in their IR. However, 23 digital curation activities out of 33 were not performed by more than half of the IRs and the active participation of IRs were only seen in ten digital curation activities. Likewise, the performance of preservation activities was extremely low. Similarly, disagreements were recorded by the survey participants towards several digital curation activities. The most disagreed digital curation activities were emulation and cease data curation. Such hurdles can be overcome by forming a digital curation policy at the national level.

 Eprints and qualified Dublin core metadata were the most used software and metadata scheme by the IRs in South India, respectively. Likewise, URI Identifier was used by most IRs. The most used controlled vocabularies were LCSH and DDC, while approximately half of the respondents did not have use any controlled vocabularies in their IR. Even though most IRs did not use any applications to manage the digital content, 20 per cent used Apache solar. Lee and Stvilia (2017) observed that the IRs provide the data curation services as per the supporting feature of IR softwares. Parallel to this observation, the study also found that the tools used for digital curation have a strong association with the software used by the IR.

 However, some South Indian IRs were hesitant to offer certain features that Eprints and DSpace provide. The features such as full-text indexing and authentication are supported by EPrints and DSpace software’s. But it was not offered by 50% of the IRs in South India. One of the reasons that deter the South Indian IRs from implementing advanced features in addition to the fundamental features may be the lack of technical knowledge. Creation of the discussion forums or groups exclusively for the IRs which are using particular repository software such as DSpace, EPrints or Digital Common may enable all the IRs to utilize the entire features currently provided by repository softwares. Such type of platforms will be useful for the IRs in developing countries to share their experiences, to clear doubts and understand best practices regarding digital curation. As technical skills are essential to managing the digital objects of an IR, it is advisable to develop a working methodology that includes the computer centre of the institutions besides the library staff, and the respective heads of the departments should take the initiative to implement the same. Providing training in digital curation to existing librarians and incorporating digital curation in LIS education will enable library professionals to deal with this situation.

 Unlike in developed countries, the South Indian IRs face challenges such as lack of technical skill and low workforce (Roy, Mukhopadhyay, and Biswas, 2012). The current research sheds light towards digital curation practices in IRs and investigates how much attention has been paid by South Indian IRs to digital curation activities despite these limitations. The findings of this study may persuade the institutions in developing countries to evaluate whether digital objects available in IRs are currently curated and to check whether the IRs have utilized all the possibilities that repository software provides for digital curation.

**References**

Akers, K. G. and Doty, J. (2013), “Disciplinary differences in faculty research data management practices and perspectives”, *International Journal of Digital Curation*, Vol. 8 No. 2, pp. 5–26.

Allen, Mike (2017), *The SAGE encyclopedia of communication research methods,* SAGE Publications, Thousand Oaks, US.

Anilkumar, N. (2018), “Research Data Management in India: A Pilot Study”, *European Physical Journal Web of Conferences,* Vol. 186, available at: <https://www.epj-conferences.org/articles/epjconf/pdf/2018/21/epjconf_lisaviii2018_03002.pdf> (accessed 12 March 2019)

Anyaoku, E., Echedom, A. and Baro, E. (2019), "Digital preservation practices in university libraries: An investigation of institutional repositories in Africa", *Digital Library Perspectives*, Vol. 35 No. 1, pp. 41-64. <https://doi.org/10.1108/DLP-10-2017-0041>

Callicott, B. B., Scherer, D., & Wesolek, A. (2016), *Making institutional repositories work*, Purdue University Press, West Lafayette, IN.

Choudhury, G. S. (2008), “Case study in data curation at Johns Hopkins University”, *Library Trends*, Vol. 57 No. 2, pp. 211–220.

Constantopoulos, P., Dallas, C., Androutsopoulos, I., Angelis, S., Deligiannakis, A., Gavrilis, D., …Papatheodorou, C. (2009), “DCC&U: An extended digital curation lifecycle model”, *International Journal of Digital Curation*, Vol. 4 No. 1, pp. 34-45.

# DCC, (n.d.), “What is digital curation?”, available at: <http://www.dcc.ac.uk/digital-curation/what-digital-curation> (accessed 12 March 2019)

DeRidder, J. L. and Helms , Alissa Matheny (2016), “Intake of Digital Content: Survey Results From the Field” D-Lib Magazine, Volume 22, Number 11/12, available at: <http://www.dlib.org/dlib/november16/deridder/11deridder.html> (Accessed 14 December 2019)

Digital preservation Coalition (n. d.). “Digital preservation handbook”, available at: [https://www.dpconline.org/handbook/technical-solutions-and-tools/ persistent-identifiers](https://www.dpconline.org/handbook/technical-solutions-and-tools/persistent-identifiers) (accessed 12 March 2019)

Fritz, A. (2019), ‘So many ptions, so little time: how to evaluate a digital preservation system that is right for your institution’, In Myntti, J. and Zoom, J. (eds.). *Digital preservation in libraries: preparing for a sustainable future*, Chicage, ALA, pp. 77-92.

Ganesan, P., Sridhar, K., Kirana Kumar, D. and Tutu, P. (2015), “Preservation of Digital Content in Indian Libraries: Issues and Strategies”, paper presented at 16th General Conference on Southeast Asian Librarians (CONSAL XVI), 10-13 June, Bangkok, Thailand, available at: <http://myrepositori.pnm.gov.my/bitstream/123456789/4201/1/CONSALXVI_P.Ganesan.pdf> (accessed 12 March 2019).

Heidorn, P. B. (2008), ‘’Shedding light on the dark data in the long tail of science’’, *Library Trends*, Vol. 57 No. 2, pp. 280–299.

Heidorn, P. B. (2011), “The emerging role of libraries in data curation and e-science’’, *Journal of Library Administration*, Vol. 51 No. 7–8, pp. 662–672.

Higgins, S. (2008), “The DCC curation lifecycle model”, *International Journal of Digital Curation*, Vol. 3 No. 1, pp. 134-140.

Hudson-Vitale, C., Imker, H., Johnston, L. R., Carlson, J., Kozlowski, W., Olendorf, R., & Stewart, C. (2017). *SPEC Kit# 354: Data Curation*. Association of Research Libraries, Washington, DC.

Johnston, L. R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., and Stewart, C. (2018a), “Data Curation Network: A Cross-Institutional Staffing Model for Curating Research Data”, *International Journal of digital curation,* Vol. 13 No. 1, available at: [http://www.ijdc.net/article/ view/616](http://www.ijdc.net/article/view/616)

Johnston, L.R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R. and Stewart, C. (2018b), “How important is data curation? gaps and opportunities for academic libraries”, *Journal of Librarianship and Scholarly Communication*, Vol. 6 No. 1, pp. 1-24, available at: http://doi.org/10.7710/2162-3309.2198

Karasti, H., Baker, K. S., and Halkola, E. (2006). “Enriching the notion of data curation in e-science: Data managing and information infrastructuring in the long term ecological research (LTER) network”, *Computer Supported Cooperative Work (CSCW)*, Vol. 15 No. 4, pp. 321–358.

Kari, Kingdom H. and Baro, Ebikabowei E. (2016), “ Digital Preservation Practices in University Libraries: A Survey of Institutional Repositories in Nigeria”, *Preservation, Digital Technology & Culture*, Vol. 45 No. 3, Pages 134–144

Katre, D. (2011), “Digital preservation: Converging and diverging factors of libraries, archives and museums: an Indian perspective”, IFLA journal, Vol. 37 No. 3, pp. 195–203,

Kaushik, Anna (2017), “Perceptions of LIS Professionals about Data Curation”, *World Digital Libraries* Vol. 10 No. 2, pp. 89–98.

Kim, J., Moen, W. and Warger, E. (2013), “Competencies required for digital curation: An analysis of job advertisements”, *The International Journal of Digital Curation,* Vol. 8 No. 1 [doi.org/10.2218/ijdc.v8i1.242](https://doi.org/10.2218/ijdc.v8i1.242)

Kumar, K. (2014), "Analytical Survey on Digital Preservation and Techniques among Engineering Education Institutional Libraries in Rayalaseema Region of Andhra Pradesh", *Library Philosophy and Practice* , available at: <http://digitalcommons.unl.edu/libphilprac/1045> (accessed 12 March 2019)

Lage, K., Losoff, B., & Maness, J. (2011). “Receptivity to library involvement in scientific data curation: A case study at the University of Colorado Boulder’’, *Portal: Libraries and the Academy*, Vol. 11 No. 4, pp. 915–937.

Lee, D. J., and Stvilia, B. (2017), “Practices of research data curation in institutional repositories: A qualitative view from repository staff’’, *PloS One*, Vol. 12 No. 3, e0173987. doi.org/10.1371/journal.pone.0173987

Li, Y., & 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> (accessed 12 October 2019)

Lord, P., and Macdonald, A. (2003), “e-Science Curation Report, prepared for The JISC Committee for the Support of Research” available at : <https://www.cs.york.ac.uk/ftpdir/pub/leo/york-msc-2007/information/vsr-curation/science-dc-report.pdf> (accessed 12 March 2019)

Ministry of Science and Technology, (2012), “National Data Sharing and Accessibility Policy”, available at: <https://data.gov.in/sites/default/files/NDSAP.pdf> (accessed 12 March 2019)

Morgana, A and Baptista, A. A. (2015), “ The use of Application Profiles and metadata schemas by digital repositories: findings from a survey”, available at: <http://dcpapers.dublincore.org/pubs/article/view/3770> (accessed 12 March 2019)

Moulaison, H. L., Dykas, F., and Gallant, K. (2015), “OpenDOAR repositories and Metadata practices’’, *D-Lib Magazine*, Vol. *21* No. ¾, available at: <http://www.dlib.org/dlib/march15/moulaison/03moulaison.html> (Accessed 12 March 2019)

Niu, Jinfang (2014), “Appraisal and Selection for Digital Curation”, - *International Journal of Digital Curation*, Vol. 9 No. 2, pp. 65–82.

Palmer, C.L., Cragin, M.H., Heidorn, P.B. and Smith, L.C. (2007), “*Data curation for the long tail of science: the case of environmental sciences*”, paper presented at the 3rd International Digital Curation Conference, Washington, DC, January 1-5.

Park, J., and Tosaka, Y. (2010), “Metadata creation practices in digital repositories and collections: Schemata, selection criteria, and interoperability”, *Information Technology and Libraries*, Vol. 29 No. 3, pp. 104–116.

Pham, Amy (2018), ‘Surveying the state of data curation: a review of policy and practice in UK HEIs’, MSc.thesis, University of Strathclyde, Glasgow.

Punzalan, R. L., and Kriesberg, A. (2017), “Library-mediated collaborations: Data curation at the National Agricultural Library’’, *Library Trends*, Vol. 65 No.3, pp. 429–447.

Roy, B. K., Mukhopadhyay, P., & Biswas, S. C. (2012). “An analytical study of institutional digital repositories in India”. *Library Philosophy and Practice*, *1329 .* available at: <http://digitalcommons.unl.edu/libphilprac/1329> (accessed 12 March 2019)

Scaramozzino, J. M., Ramírez, M. L., & McGaughey, K. J. (2012), “A study of faculty data curation behaviors and attitudes at a teaching-centered university”, *College & Research Libraries*, Vol. 73 No. 4, 349–365.

Serrano-Vicente, R., Melero, R and Abadal, E. (2018) "*Evaluation of Spanish institutional repositories based on criteria related to technology, procedures, content, marketing and personnel*", Data Technologies and Applications, Vol. 52 Issue: 3, pp.384-404, https://doi.org/10.1108/ DTA-10-2017-0074

Shajitha, C. and Abdul Majeed, K.C., (2018), "Content growth of institutional repositories in South India: a status report", [*Global Knowledge, Memory and Communication*](https://www.emerald.com/insight/publication/issn/2514-9342), Vol. 67 No. 8/9, pp. 547-565. <https://doi.org/10.1108/GKMC-02-2018-0018>

Shajitha, C. (2020), “Institutional Repositories in South India: an exploratory study”. [Manuscript submitted for publication].

Shen, Y., & Varvel Jr, V. E. (2013), “Developing data management services at the Johns Hopkins University’’, *The Journal of Academic Librarianship*, Vol. 39 No. 6, pp. 552–557.

Shivarama, J., Sheela V., Deepak, A. B. and Agadi, K. B. (2016), “Digital Curation Strategies for Information Management in Higher Education Institutions” In *10th Convention PLANNERNEHU, North-Eastern Hill University Shillong, Meghalaya, November 09-11, 2016*, INFLIBNET Centre, Gandhinagar, pp. 117-126.

Shrivastava, P. and Gupta, D. K. (2018), “Research Data Preservation in India: An Analysis based on Research Data Registry”, *World Digital Libraries*, Vol. 11 No. 2, pp. 107-121.

Shrivastava, P. and Gupta, D. K. (2019), “Emergence of Research Data Literacy with Special Reference to India”, ***SRELS Journal of Information Management***, Vol. 56 No. 2, pp.112-118.

Steele, T., and Sump-crethar, N. (2016), “ Metadata for Electronic Theses and Dissertations: A Survey of Institutional Repositories’’, *Journal of Library Metadata*, Vol. 16 No. 1, pp. 53–68. <https://doi.org/10.1080/19386389.2016.1161462>

Tripathi, M., Shukla, A., & Sonkar, S. (2017a), “Research data management practices in university libraries: a study”, DESIDOC *Journal of Library and Information Technology*, Vol. 37 No. 6, pp. 417–424.

Tripathi, M., Chand, M., Sonkar, S.K. and Jeevan, V.K.J. (2017b), “*A brief assessment of researchers’ perceptions towards research data in India*”, IFLA Journal, Vol. 43 No. 1, pp. 22-39.

Weber, N. M., Palmer, C. L., and Chao, T. C. (2012), “Current trends and future directions in data curation research and education”, *Journal of Web Librarianship*, Vol. 6 No. 4, pp. 305–320.

Witt, M. (2009), “Institutional repositories and research data curation in a distributed environment”, *Library Trends*, Vol. 57 No. 2, pp.191–201.

Yoon, Ayoung, and Tibbo, H. (2011), “ Examination of Data Deposit Practices in Repositories with the OAIS Model”, *IASSIST Quarterly,* Vol 35 No. 4, pp. 6-13.

Yoon, A. and Schultz, T. (2017), “ Research Data Management Services in Academic Libraries in the US: A Content Analysis of Libraries’ Websites”, *College & Research Libraries*, Vol 78 No. 7, 920-933.