DESIDOC Journal of Library & Information Technology, Vol. 44, No. 5, September 2024, pp.284-288 , DOI : 10.14429/djlit.44.05.20075

Curating the Future of Research: Navigating FAIR Challenges in Academic Repositories

Juan-José Boté-Vericad[1*], EminaAdilovich[2], Anna Caellas-Camprubí[1], and Ignasi Labastida[1]

[1]*Facultatd'Informació i MitjansAudiovisuals & Centre de Recerca en Informació, Comunicació i Cultura. Universitat de Barcelona, Barcelona - 08014, Spain*

[2]*Institut zadruštvenaistraživanja. Fakultetpolitičkihnauka.Univerzitet u Sarajevu, Sarajevu - 71000, Bosna i Hercegovina*

[*]*E-mail: juanjo.botev@ub.edu*

ABSTRACT

This study examines consortia academic research data repositories, and their curators, across five global regions. Using an ad-hoc sample, we explore the roles of curators in research data repositories and their perceptions of skills and training, Research Data Management (RDM) practices, and the limitations, strengths, and opportunities they encounter in doing their work. Furthermore, our study analyses the roles of academic research repositories and the degree to which the FAIR principles are embedded in their strategic approaches. Our findings indicate that although numerous repositories are managed through collaborations among higher education bodies and research organisations, in certain nations, the development of research data repositories is in its early stage, often led by a single entity and having limited representation. Additionally, curators face challenges when dealing with the submission of datasets by researchers, who often provide insufficiently clear information for data reuse. We argue that despite the efforts to align data repositories with FAIR principles, substantial progress is required in educating researchers on their effective utilisation.

Keywords: Research data management; Data skills; Research data repositories; FAIR data, Data steward

1. INTRODUCTION

The aim of this study is to examine the operational practices and key challenges that curators face in managing academic research repositories globally. Research data repositories are becoming more popular because of the specialisation of the repositories itself and because of mandates, especially in Europe[1] where the data management policy for projects funded by the Commission under the Horizon 2020 programme is to make data as open as possible and as closed as necessary. This study also explores the under-researched areas of operational dynamics, challenges, and development aspirations of curators in this evolving academic field. This study employs Nonaka and Takeuchi's knowledge creation theory[2] to examine how tacit knowledge in the realm of academic research is transformed and disseminated through data repositories, underscoring the pivotal role of curators in this process. Given the increasing reliance on digital repositories for the storage, preservation, reuse and dissemination of academic research, understanding the detailed dynamics of curatorial practices and hurdles is crucial for enhancing the efficacy and impact of these evolving repositories for storing, preserving, reusing, and disseminating academic research, and are central to major changes in data management policies and practices. Additionally, this study is part of a larger ongoing project examining the cycle of Research Data Management (RDM), building upon insights from a previous publication[3].

This study contributes to a deeper understanding of the complexities and future directions in the management of academic research repositories.

Received : 04 April 2024, Revised : 23 May 2024
Accepted : 07 June 2024, Online published : 24 September 2024

2. LITERATURE REVIEW

Although research demands a systematic approach to data, discrepancies exist between the final products appearing as published research results and the data utilised, partially utilised, or left entirely unused. Singh and Kumar[4] note that unlike the textual content forming the basis of research reports, researchers generate, store, and analyse their research data on a much larger scale. Consequently, research data curation involves a set of tasks which include acquiring, reformatting, adding metadata, maintaining, and delivering datasets to researchers. Behind curation stand curators who play a crucial role in these processes and also facilitate data sharing, compliance with standards, and adherence to ethical guidelines, thereby enhancing the overall quality and reliability of research outcomes. Cousijn[5], *et al*. argue sharing data to meet FAIR principles requires also sharing associated results for full context.

Additional efforts in managing research data are provided and guided by data curators. Dorst[6] *et al.* reminded us that the ultimate goal of FAIR principles, the ability for automated data discovery and processing without the need for human intervention, is crucial for accelerating research processes and increasing efficiency in scientific research. This concept not only promotes transparency and collaboration within the scientific community but also opens doors for new research opportunities and innovations arising from the wider availability and better utilisation of research data. While the ultimate aim of FAIR principles is to enable automated processes, it is important to recognise that human intervention plays a crucial role in achieving this goal, and curators provide necessary expertise and oversight to ensure the quality and reliability of data. Data curators require diverse skills to successfully perform their tasks:technical skills[7], training for researchers[8], and key data literacy skills[9].

In the context of required skills, Nahotko[10], *et al.* emphasise that one of the stages of RDM maturity development involves the application of principles, practices, and resources of traditional librarianship to research data. Data repository curators, who are not always Library and Information Science (LIS) experts by primary profession, should master skills inherent to professionals handling data, information, and their sources in various forms. Given the need for new RDM-related skills, as emphasised by Hamad[11], *et al.* one of the key challenges in the proper utilisation and organisation of RDM revolves around the adoption of necessary skills. Such challenges primarily concern larger structures upon which data repositories depend, financial resources, and the extent to which governmental structures recognize the significance of investing in them. Sheikh[12], *et al.* note that data sharing practices and the development of RDM services in libraries are more prevalent in developed countries, suggesting a causal relationship between the approach to investing in curator training and data repositories and the level of a country's development.

Masinde[13], *et al.* discuss the lack of attention given to key stakeholders such as researchers, librarians, IT departments, research offices, and legal services during the maintenance of data repositories. A coordinated approach to data management in the research environment isessential for promoting FAIR principles, asit would lead to greater transparency, interoperability, and data reuse.

Nonaka and Takeuchi's knowledge creation theory, particularly through its SECI model phases of socialisation, externalisation, combination, and internalisation, has been instrumental in exploring curatorial practices in academic research repositories within the context of LIS. This perspective not only helps understand how knowledge is curated, managed, and shared, but also offers deeper insights into the operational practices and challenges of curators, clarifying how they transform tacit knowledge into explicit to enhance knowledge management and dissemination in research settings.

3. METHODOLOGY

This study is primarily focused on the data curators of academic research repositories, examining their qualifications, experiences, operational practices, and the challenges they face in the context of digital repository management. The study's methodology was informed by the knowledge creation theory, guiding the development of the survey that probes the processes of knowledge externalisation and combination within academic repositories. The survey captures how curators externalise and combine knowledge, guiding both methodology and data analysis. This approach reveals patterns in translating tacit expertise into explicit procedures and integrating diverse knowledge sources to enhance repository management.

To enhance the study's novelty, we integrated knowledge creation Theory with practical curation insights, used multi-modal data collection, emphasised ethical data handling, applied detailed repository selection criteria, and employed advanced transcription tools for precise and comprehensive data analysis.

Ensuring informed consent, confidentiality, transparency, and ethical data handling to respect participants' autonomy and protect their privacy were integral to the study. The methodology to select repositories, explained in a prior study[3], considered institutional, national, or consortium-managed repositories hosting research data. All these repositories conducted data quality assurance measures, enhancing overall data quality[21]. From the potential candidates, a convenience sample of 28 repositories meeting the criteria was chosen[22].

3.1 Research Questions

The study is anchored around three pivotal research questions:

RQ1 - How do the qualifications and experiences of curators influence their practices in managing academic research repositories?

RQ2 - What are the principal challenges curators face in ensuring the accessibility, preservation, and reuse of data?

RQ3 - In what ways do curators perceive their developmental needs and aspirations in adapting to the evolving landscape of digital repository management?

4. RESULTS AND ANALYSIS

In this section, the participation aspect of the study is explored. Total 28 responses were received, encompassing a diverse geographical representation from continents as shown in Table 1.

The study shows that participants from diverse backgrounds in library science, data management, and various academic fields possess skills in data curation and programming, playing a pivotal role in transforming tacit knowledge into explicit forms within the knowledge creation cycle, enhancing accessibility for the academic community.

Table 1. Participants from different countries in repositories

Zone	Countries
Africa	Burkina Faso (1)
Asia	India (1)
Europe	Belgium (1), Bosnia and Herzegovina (1), Czech Republic (1), France (2), Germany (2), Hungary (1), Norway (1), Serbia (2)
North America	Canada (5), the USA (4)
South America	Argentina (2), Brazil (3)

During analysis, several specific challenges affecting research data management were identified. Challenges such as data inconsistency, privacy concerns, and technological limitations affect research data management. These issues can lead to data loss, compromised integrity, and increased handling complexity, highlighting the need for robust strategies to address these problems effectively.

Curators aim to advance their IT and coding skills, focusing on best practices in software development, programming, and data management. They seek training in data stewardship, cybersecurity, and cloud technologies, while enhancing skills in metadata, SQL, and data engineering, and exploring new training in research data management and software tools.

"Processing of primary research data" (Curator, female, Argentina)

Institutions offer training and support through workshops, tutorials, seminars, and consultations, providing extensive resources and established policies for effective data management.

"Yes, face-to-face training is done frequently for employees. Furthermore, tutorials and manuals are offered to externals/public on how to prepare and send their research data" (Curator, female, Belgium)

Participants receive training on Open Science, FAIR principles, and data management via seminars, tutorials, workshops, and consultations, both online and in-person.

"We held annual face-to-face workshops" (Curator, male, India)

In relation to obstacles, lack of time and technical skills are hindering the efficient management and development of the repository. There is also a need for data stewards, better technical support, and training on metadata standards and legal issues. Challenges include addressing ethical concerns, lack of incentives for data sharing, and limited resources. Additionally, there are issues with Intellectual Property Rights (IPR), culture of data sharing, and funding. The staff also highlight the difficulties in training users on new features and the lack of institutional recognition and support.

"No support from the institution to market/engage with the platform. Unclear expectations on which content should be in the repository." (Curator, male, Canada)

Regarding the practices in managing academic research, all data in the repositories mentioned comply with the FAIR principles to varying degrees. The repositories use persistent identifiers (such as DOIs and ORCIDs), provide rich metadata for discovery, support data exchange, control data access and licensing, and offer continuous availability. Some repositories are CoreTrustSeal certified, ensuring data quality and compliance. There are efforts to enhance and ensure FAIRness in data curation processes. Additionally, some repositories actively monitor and promote FAIR data practices and provide tools for assessing FAIR compliance.

"The repository upholds the CoreTrustSeal certification, with its Preservation Plan emphasizing adherence to FAIR principles- ensuring data are Findable, Accessible, Interoperable, and Reusable (FAIR)" (Curator, female, Serbia)

Statistics are collected in the repository to track events such as uploads and usage patterns, but not downloads. Various tools like DSpace's native usage tracking, Google Analytics, Google Search Console, and Matomo are used to gather aggregated statistics visible in OpenAIRE. Metrics are shared publicly, and efforts are ongoing to transition to automatic tracking through Dataverse for improved analysis of usage and downloads.

"Upload and download statistics are collected and analysed in the repository to enhance user services. Currently, submissions and usage requests are processed via email, with data archivists manually compiling statistical reports. The transition to DataVerse will automate the calculation of usage, download, and upload statistics." (Curator, female, Bosnia & Herzegovina)

In terms of data types, there are several types of data that can be included in a repository, ranging from computer code and statistical files to databases, spreadsheets, images, videos, multimedia, and more. The data covers a wide range of fields such as social and behavioural sciences, health, earth models, marine information, and archaeological documentation. Data formats are varied, providing a diverse collection of information to support research and publications.

"Spreadsheets, interviews, computer code, peer-reviewed and non-peer-reviewed data, whereby text publication reviewers also access corresponding data publications for review." (Curator, female, Germany)

5. DISCUSSION

This study examines how data repository curators' skills and experiences affect their ability to manage academic repositories, highlighting the challenges in data stewardship and the need for ongoing professional development.

Concerning the first research question, the discussion reveals that key challenges faced by curators include a need for more advanced IT and coding skills, along with a notable gap in training related to research data management and metadata standards[14].

In relation to the second research question, curators are often confronted with inadequate technical support, limited time resources, and complexities surrounding adherence to metadata standards and legal frameworks, crucial factors in upholding the data's integrity and

usability. Curatorial processes with data seem to influence the discovery of the datasets and allow repositories to be more effective[15].

Concerning the third research question, curators seek to enhance their expertise in areas such as IT, coding, data stewardship, and cybersecurity. These skills would ensure that the data stored in repositories are managed effectively to adhere to the FAIR principles, which encompass findability, accessibility, interoperability, and reusability over time[16]. Despite the critical nature of these competencies, technical skills in this domain are often undervalued, highlighting the need for greater recognition and investment in expertise to improve research data management practices[17]. Another study introduced a series of guidelines and methodologies aimed at data science professionals[18].

The expertise of the curator and the understanding of researcher data activities is essential for repository success[19]. The need for professional development in digital repository management is crucial for enhancing the effectiveness and adaptability of curators in this rapidly evolving field[20].

While the study provides valuable insights into research data management, it is important to acknowledge the assumptions and limitations in this research. If explicitly mentions assumptions, such as sample representativeness and measurement accuracy, and discusses limitations, like convenience sampling bias and response bias, to enhance the credibility of the research. By addressing these factors, we provide context, aid in interpreting findings, and strengthen the overall robustness of the research.

6. CONCLUSIONS

In this study, we examined the qualifications, experiences, and challenges confronting curators of academic research repositories, engaging with a diverse sample of 28 repositories worldwide. The findings highlight a broad spectrum of skills among curators, alongside a consensus about the urgent need for advanced IT and data management training. Despite the diversity of cultures, levels of development, and resources across these repositories, curators universally face similar problems and difficulties, underscoring a shared global challenge in research data management. To strengthen research data management in academic repositories, we recommend implementing advanced data consistency checks, enhancing privacy protocols, and adopting cutting-edge technologies to address identified challenges and improve overall data handling practices.

While curators are making active progress in aligning their practices with the FAIR principles, particularly through the development of strong preservation plans, they continue to face ongoing challenges in ensuring data accessibility, preservation, and reuse. The application of Nonaka and Takeuchi's Knowledge Creation Theory allowed us to gain an in-depth understanding of curators' complex roles in academic repositories and to identify ways to enhance knowledge management in research.

REFERENCES

1. England, J. & Malaguarnera, G. Horizon Europe open science requirements in practice - OpenAIRE webinar. 2023. doi: 10.5281/zenodo.7738501
2. Nonaka, I. & Takeuchi, H. The knowledge-creating company: How Japanese companies create the dynamics of innovation. Oxford University Press, 1995.
3. Boté-Vericad, J.-J; Carballo-Garcia, A.; Bautista-Villaescusa, M. & Healy, S. Research data repositories in the RDM cycle: Challenges and strengths for curators/datastewards. *AIB Studi*, 2024, **63**(3), 563–581. doi: 10.2426/aibstudi-13892
4. Singh, D.M.K. & Kumar T.K; D.G. Research data curation in academic institutions challenges & expectations. *DESIDOC J. Libr. Infor. Technol.*, 2023, **43**(01), 39-44. doi: 10.14429/djlit.43.01.18624
5. Cousijn, H.; Habermann, T.; Krznarich, E. & Meadows, A. Beyond data: Sharing related research outputs to make data reusable. *LEAP*, 2022, **35**(1), 75–80. doi: 10.1002/leap.1429
6. Dorst, T.; Gruber, M.; Vedurmudi, A.P.; Hutzschenreuter, D.; Eichstädt, S. & Schütze, A. A case study on providing FAIR and metrologically traceable data sets. *Acta IMEKO*, 2023, **12**(1), 1–6. doi: 10.21014/actaimeko.v12i1.1401
7. Costal, M.B.; Sales, L. & Zattar: Competênciaem dados: habilidades naatuação e formação do bibliotecário. BIBLIOS, 202, **34**(2). doi: 10.14295/biblios.v.34i2.11809
8. Carson, M.B.; Gonzales, S.; Shaw, P.; Schneider, D. & Holmes, K. Bridging the gap: A library based collaboration to enhance data skills for clinical researchers. *LHS*, 2023, **7**(2), e10339. doi: 10.1002/lrh2.10339
9. Burress, T.; Mann, E. & Neville, T. Exploring data literacy via a librarian-faculty learning community: A case study. *JAL*, 2020, **46**(1), 102076. doi: 10.1016/j.acalib.2019.102076
10. Nahotko, M.; Zych, M.; Januszko-Szakiel, A. & Jaskowska, M. Big data-driven investigation into the maturity of library research data services (RDS). *JAL*, 2023, **49**(1), 102646. doi: 10.1016/j.acalib.2022.102646
11. Hamad, F.; Al-Fadel, M. & Al-Soub, A. Awareness of research data management services at academic libraries in jordan: Roles, responsibilities and challenges. *NRAL*, 2021, **27**(1), 76–96. doi: 10.1080/13614533.2019.1691027
12. Sheikh, A.; Malik, A. & Adnan, R. Evolution of research data management in academic libraries: A review of the literature. *Inf. Dev.*, 2023, 026666692311574. doi: 10.1177/02666669231157405
13. Masinde, J.; Chen, J.; Wambiri, D. & Mumo, A. Research librarians' experiences of research data management activities at an academic library in a developing country. *Data and Inf. Man.*, 2021, **5**(4), 412–424. doi: 10.2478/dim-2021-0002
14. Thielen, Joanna and Amy Neeser. Making job postings more equitable: Evidence based recommendations

from an analysis of data professionals job postings between 2013-2018. *EBLIP*, 2020, **15**(3), 103–56. doi: 10.18438/eblip29674
15. Luong, H.Q.; Fallaw, C.; Schmitt, G.; Braxton, S.M. & Imker, H. Responding to reality: Evolving curation practices and infrastructure at the university of Illinois at Urbana-Champaign. *J. eSc. Lib.*, 2021, **10**(3), 1202. doi: 10.7191/jeslib.2021.1202
16. Ostler, R.; Castells, N.; Glendining, M. & Perryman, S. Linking legacies: Realising the potential of the rothamsted long-term agricultural. *In* Experiments in Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development, edited by Williamson, S. & Leonelli, H. Springer, 2023, 125-148. doi: 10.1007/978-3-031-13276-6
17. Leonelli, S. & Tempini, N. Data journeys in the sciences. Springer, UK. doi: 10.1007/978-3-030-37177-7
18. Usama Fayyad, U. & Hamutcu, H. Analytics and data science standardisation and assessment framework. 2020. *Har. Data Sc. Rev*. doi: 10.1162/99608f92.1a99e67a
19. Cooper, A.; Steeleworthy, M.; Paquette-Bigras, È.; Clary, E.; MacPherson, E.; Gillis, L. & Brodeur, J. Creating guidance for Canadian data verse curators: Portage network's data versecuration guide. *J. eSc. Lib*., 2021, **10**(3), 1201. doi: 10.7191/jeslib.2021.1201
20. Miller, R.E. Researchers may need additional data curation support. EBLIP, **14**(1), 2019, 74–76. doi: 10.18438/eblip29539
21. Kindling, M. & Strecker, D. Data quality assurance at research data repositories. *Data Sci. J*. doi: 10.5334/dsj-2022-018
22. Boté-Vericad, J.-J. & García-Carballo, A. Global research data repositories: Distribution, software platforms, and organisational structures. doi: 10.34810/data1357

CONTRIBUTORS

Dr Juan-José Boté-Vericad is Lecturer Professor at the Faculty of Information and Media Studies and member of the Centre de Recerca en Informació, Comunicació i Cultura (CRICC) at the University of Barcelona. His areas of interest includes: Open science, Cultural heritage and gender studies.
He has designed the study, contributed in writing the methodology, results, the discussion and the review of the final paper.

Ms Emina Adilović is a Senior Research Associate at the Institute for Social Science Research, Faculty of Political Sciences, University of Sarajevo. Her areas of interest includes: Information literacy, Cultural and gender studies, Knowledge organisation, Open science, and Information security.
She has contributed in writing the literature review and conclusions.

Ms Anna Caellas-Camprubí is a PhD Student at the University of Barcelona and has developed her professional career in the field of research support from a broad perspective, with a focus on Open Science. Her research interests includes: Open data, Research data repositories, Data skills and Data reusability.
She has contributed in writing the introduction and discussion.

Dr Ignasi Labastida is the head of the Research Unit at the CRAI of the University of Barcelona and a member of the Centre de Recerca en Informació, Comunicació i Cultura (CRICC), Universitat de Barcelona. He is the rector's delegate for Open Science and a member of the bioethics commission of the University of Barcelona.
He has contributed in writing the results and review the final paper.