International Journal on Digital Libraries (2023) 24:185–189
https://doi.org/10.1007/s00799-023-00388-9

EDITORIAL

Focused Issue on Digital Library Challenges to Support the Open Science Process

Giorgio Maria Di Nunzio[1]

Published online: 29 November 2023

Abstract
Open Science is the broad term that involves several aspects aiming to remove the barriers for sharing any kind of output, resources, methods or tools, at any stage of the research process (https://book.fosteropenscience.eu/en/). The Open Science process is a set of transparent research practices that help to improve the quality of scientific knowledge and are crucial to the most basic aspects of the scientific process by means of the FAIR (Findable, Accessible, Interoperable, and Reusable) principles. Thanks to research transparency and accessibility, we can evaluate the credibility of scientific claims and make the research process reproducible and the obtained results replicable. In this context, digital libraries play a pivotal role in supporting the Open Science process by facilitating the storage, organization, and dissemination of research outputs, including open access publications and open data. In this focused issue, we invited researchers to discuss innovative solutions, also related to technical challenges, about the identifiability of digital objects as well as the use of metadata and ontologies in order to support replicable and reusable research, the adoption of standards and semantic technologies to link information, and the evaluation of the application of the FAIR principles.

Keywords Open Science · Open data · Reproducible research · Policies · Evaluation

1 Introduction

Open Science is a transformative movement that aims to break down barriers and promote transparency, accessibility, and collaboration in the research process. The Open Science process presents a great opportunity to improve the conduct of "good practice" research in terms of openness, easy access to scientific results, reproducibility, and multidisciplinarity, resulting in better and trusted science [15]. In this context, researchers are responsible for developing and maintain interoperable frameworks that define their practices for archiving, referencing, and describing research digital objects. But deciding to practice open science can come with some hesitancy since there are positives and negatives aspects in even the smallest of Open Science practices [9]. In this sense, digital libraries (DL) play a pivotal role in supporting the Open Science process by facilitating the storage, organization, and dissemination of research outputs, including open access publications and open data [20].

In this focused issue, we invited researchers to discuss innovative solutions, also related to technical challenges, about the identifiability of digital objects as well as the use of metadata and ontologies in order to support replicable and reusable research, the adoption of standards and semantic technologies to link information, and the evaluation of the application of the FAIR principles [25]. The main topics of the focused issue dedicated to DLs that support the Open Science process are:

- Open Access and Open Data,
- Reproducible Research,
- Guidelines and Policies,
- Infrastructures, Tools, and Evaluation.

In the following sections, we focus on these topics and describe the possible current challenges.

1.1 Open access and open data

One of the main challenges that DLs encounter in their efforts to support the Open Science process is the high volume and

✉ Giorgio Maria Di Nunzio
giorgiomaria.dinunzio@unipd.it

[1] Department of Information Engineering, University of Padova, Via Gradenigo 6a, 35131 Padua, Italy

diversity of digital content being generated. With the growth of the research outputs produced today, digital libraries must develop scalable and robust search functionalities to ensure efficient discovery and retrieval of open access publications and open datasets [21]. DLs must implement rigorous data curation standards and version control systems to guarantee the integrity of open access publications and open data. They need to address issues related to data accuracy, metadata completeness, and long-term preservation to maintain the credibility and usability of the stored content [3].

Interoperability and reusability are also significant challenges, not only technical ones, in achieving seamless integration and exchange of data across different digital library platforms [7]. Open Science encourages data sharing and reuse, but variations in data formats, metadata standards, and access protocols hinder the effective aggregation and interoperability of diverse resources. In this context, DLs researchers must actively collaborate with standardization bodies and adopt common protocols to ensure compatibility and enable efficient knowledge discovery [5].

Ethical and legal considerations are also pertinent challenges since reproducible research may involve sharing sensitive data, intellectual property, or proprietary software. Digital libraries must take into account legal frameworks, copyright restrictions, and data protection regulations to enable the open sharing of research artifacts while respecting rights and ensuring privacy [17].

1.2 Reproducible research

DLs play a critical role in tackling the challenges of the dissemination and preservation of reproducible research artifacts [23]. The complex nature of research workflows and the number of tools and software used in the research process pose interesting research questions. DLs must address the issue of capturing and preserving these workflows, along with their associated dependencies, in a manner that allows for easy replication. This involves storing not only the final research output but also the code, data, parameters, and software versions used in the analysis [8].

Reproducible research often involves the use of large datasets, which may have privacy or legal constraints. Digital libraries need to develop mechanisms to handle sensitive data, such as anonymization techniques or controlled access, while still enabling reproducibility. Moreover, tracking the provenance of data and code is crucial for establishing the lineage of research objects, but it requires metadata standards and mechanisms for capturing and representing the relationships between different components of a research project [14].

Additionally, the issue of code and software sustainability poses a significant challenge to reproducible research in DLs. Software dependencies, versioning, and compatibility issues can hinder the reproducibility of research projects. Digital libraries must develop strategies to address these challenges [11].

1.3 Guidelines and policies

DLs promote research culture by providing resources to users for educational and research purposes, and in this sense, they serve as important platforms for implementing and enforcing guidelines and policies in supporting Open Science [12]. One of the main challenges is the dynamic nature of guidelines and policies that, in the Open Science domain, are continuously evolving, with new guidelines and policies emerging to address emerging issues such as data management and research reproducibility. Digital libraries must remain up-to-date with these developments and adapt their infrastructure, workflows, and services accordingly to support compliance and facilitate the implementation of these guidelines and policies [22, Chapter 12].

DLs should implement mechanisms for assessing compliance, such as automated checks for data availability or metadata completeness, and develop transparent reporting mechanisms to communicate the level of compliance achieved by individual researchers or research outputs [18].

Digital libraries should also adopt sustainable preservation strategies and technologies to ensure the longevity and integrity of Open Science content and implement metadata standards, persistent identifiers, and preservation metadata to facilitate discovery, reuse, and citation of research outputs [6].

1.4 Infrastructures, tools, and evaluation

DLs play a major role in supporting Open Science projects, infrastructures, tools, and evaluation by setting the necessary platforms and services for researchers to collaborate, share, and disseminate their work. In fact, DLs must accommodate the diverse needs of different projects, ensuring scalability and flexibility in their infrastructure and services [24].

The volume and diversity of digital content being generated and the exponential growth of research outputs poses a big challenge to DLs that must develop scalable infrastructure and robust search functionalities to ensure efficient discovery and retrieval of publications and datasets [26].

Furthermore, the sustainability of digital libraries is a constant challenge. Open access and open data initiatives require financial resources for infrastructure maintenance, system upgrades, and continuous content acquisition. Digital libraries must explore diverse funding models, engage in partnerships with research institutions, funding agencies, and stakeholders to secure long-term financial support [19].

Traditional evaluation metrics may not adequately capture the value and impact of Open Science practices such as data

sharing, software development, or community engagement. Digital libraries need to develop new evaluation frameworks that consider alternative metrics, such as data reuse, software citations, or community contributions. They must also promote the adoption of open metrics that provide a more comprehensive and nuanced understanding of the impact of open science outputs and tools [27].

2 Articles in this focused issue

After the call for papers to the focused issue, we received six contributions that were reviewed by at least two anonymous researchers who are experts in this field. Three papers were accepted after two rounds of reviews (with an acceptance rate of 50%).

In [10], the authors discuss how citation indexes are part of the research infrastructure and are in use by most scientists. They are a necessary tool in order to cope with the increasing amounts of scientific literature being published. The authors argue that the creation of a citation index for the humanities could offer substantial advantages to scholars in the humanities for several reasons: first, humanities scholars have traditionally relied on information-seeking practices that utilize citations and reference lists to uncover pertinent publications; secondly, a comprehensive citation index dedicated to the humanities would serve as a valuable data resource for researchers interested in conducting bibliometric studies within this field; lastly, by capturing the abundance of references to primary and secondary sources found in humanities literature, it becomes possible to establish connections between archives, galleries, libraries, and museums, where digitized versions of these sources are increasingly becoming available. The authors list the main aspects that are necessary to devise the creation of a Humanities Citation Index (HuCI): a decentralized and federated research infrastructure for gathering, sharing, elaborating, exposing bibliographic metadata and citation data of humanities publications that offers hooks for the development of further applications to keep track of the evolution of humanities research.

In [13], the authors present the need for sophisticated models to represent and manage documents besides the textual or conceptual elements by including the information about their layout, logical structure, their context, and broader semantic meaning. To accommodate this 'holistic' approach, the authors propose the use of Knowledge Graphs as a means of data storage. The main objectives of the paper are related to crafting an ontology for DLs that: goes beyond traditional record-based description, adopting a graph-based representation of knowledge; expands the area of description for the content and context; can describe concepts that are typical of Open Science; may improve findability and reusability of both standard and additional materials using graph-based techniques. In order to accomplish this view, this paper proposes an ontology that includes suitable concepts and relationships not considered by traditional DL schemes, but needed to support researchers in an Open Science environment. The authors also present use cases of the proposed ontology, including simple use cases on an initial set of functions that this ontology may enable, run on a prototype implementation in order to suggest how it may improve the DL and Open Science practice.

In [16], the authors discuss how institutions accomplish their knowledge dissemination objectives by providing public access to their' "knowledge products" such as Electronic Theses and Dissertations (ETD) through institutional repositories. In particular, the authors analyze and compare institutional repositories in universities from The European University of the Seas (SEA-EU) Alliance in order to determine whether universities' mandates to deposit theses influence the number of theses deposited to the institutional repositories. Furthermore, the analysis of the distribution of the number of theses across the scientific disciplines and their degree of openness shows that it largely concurs with the trends recorded in current scholarly literature. As a consequence, the university, or national policy, on open access and the university or national mandates on theses and dissertations deposits positively affect the number of deposited master's theses in the institutional repositories and the level of openness of repository deposits. At the same time, further research should be done on open access and EDT deposit policy development and its effect on the EDT repository content visibility. The benefit of policy development is best supported by clear indicators of repository impact which is most often the number of citations coming from the repository.

3 Conclusion

Open Science is a worldwide collaborative initiative that seeks to change the research process by promoting transparency, accessibility, and collaboration. In order to fully embrace Open Science, researchers need to develop and maintain interoperable frameworks for archiving, referencing, and describing research digital objects. However, the adoption of these practices is costly (both in terms of time and money) and, in this context, digital libraries can play a vital role in supporting this process.

The papers presented in this focused issue were able to address various challenges related to managing digital content, ensuring data integrity and preservation, achieving interoperability and reusability, complying with guidelines and policies, and providing robust infrastructures, tools, and evaluation frameworks.

This is just an initial step toward that is part of a broader effort carried out by a broad and active research commu-

nity such as, among others, TPDL [23] and JCDL [4] that encompasses the many meanings of the term digital libraries, including new forms of information institutions; operational information systems with all manner of digital content; new means of selecting, collecting, organizing, and distributing digital content; and theoretical models of information media, including document genres and electronic publishing.

Acknowledgements I want to express a personal thank you to the person who made this focused issue possible and made a huge contribution in this research domain: Prof. Emerita Maristella Agosti of the University of Padova. She has been an invaluable researcher and a great supervisor, and she still continues to inspire young students with her commitment and dedication. I started my career as a PhD under the guidance of Maristella within the Network of Excellence on Digital Libraries DELOS (http://delosw.isti.cnr.it/) in 2004. She taught me how to work in a collaborative and supportive research environment, and after twenty years, all the colleagues that I met during those days are now friends. And I was just a part of it, because Maristella has been a pioneer in this field for over thirty years [1, 2]. I can only be grateful and truly fortunate to have had the opportunity to learn from you.

References

1. Agosti, M., Di Nunzio, G.M., Ferro, N., et al.: Thirty years of digital libraries research at the University of Padua: the user side. In: Serra, G., Tasso, C. (eds.) Digital Libraries and Multimedia Archives. Communications in Computer and Information Science, pp. 42–54. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-319-73165-0_5
2. Agosti, M., Di Nunzio, G.M., Ferro, N., et al.: Thirty years of digital libraries research at the University of Padua: the systems side. In: Serra, G., Tasso, C. (eds.) Digital Libraries and Multimedia Archives. Communications in Computer and Information Science, pp. 30–41. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-319-73165-0_4
3. Agosti, M., Di Nunzio, G.M., Ferro, N., et al.: An innovative approach to data management and curation of experimental data generated through IR test collections. In: Ferro, N., Peters, C. (eds.) Information Retrieval Evaluation in a Changing World: Lessons Learned from 20 Years of CLEF. The Information Retrieval Series, pp. 105–122. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-22948-1_4
4. Aizawa, A., Mandl, T., Carevic, Z. (eds): JCDL '22: Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries, Association for Computing Machinery, New York (2022). https://dl.acm.org/doi/proceedings/10.1145/3529372
5. Balakireva, L., Balakirev, F.: Making FAIR practices accessible and attractive. In: Silvello, G., Corcho, O., Manghi, P., et al. (eds.) Linking theory and practice of digital libraries. Lecture Notes in Computer Science, pp. 417–424. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-16802-4_41
6. Barrueco, J.M., Termens, M.: Digital preservation in institutional repositories: a systematic literature review. Digit. Libr. Perspect. **38**(2), 161–174 (2021). https://doi.org/10.1108/DLP-02-2021-0011
7. Bartalesi, V., Pratelli, N., Lenzi, E.: Linking different scientific digital libraries in Digital Humanities: the IMAGO case study. Int. J. Digit. Libr. **23**(4), 303–317 (2022). https://doi.org/10.1007/s00799-022-00331-4
8. Bhattarai, P., Ghassemi, M., Alhanai, T.: Open-source code repository attributes predict impact of computer science research. In: Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries. Association for Computing Machinery, New York, JCDL '22, pp. 1–7 (2022). https://doi.org/10.1145/3529372.3530927
9. Castille, C.M., Kreamer, L.M., Albritton, B.H., et al.: The open science challenge: adopt one practice that enacts widely shared values. J. Bus. Psychol. **37**(3), 459–467 (2022). https://doi.org/10.1007/s10869-022-09806-2
10. Colavizza, G., Peroni, S., Romanello, M.: The case for the Humanities Citation Index (HuCI): a citation index by the humanities, for the humanities. Int. J. Digit. Libr. (2022). https://doi.org/10.1007/s00799-022-00327-0
11. Di Cosmo, R.: Should we preserve the world's software history, and can we? In: Silvello, G., Corcho, O., Manghi, P., et al. (eds.) Linking Theory and Practice of Digital Libraries. Lecture Notes in Computer Science, pp. 3–7. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-16802-4_1
12. Farid, G., Warraich, N.F., Iftikhar, S.: Digital information security management policy in academic libraries: a systematic review (2010–2022). J. Inf. Sci. (2023). https://doi.org/10.1177/01655515231160026
13. Ferilli, S., Redavid, D., Di Pierro, D.: Holistic graph-based document representation and management for open science. Int. J. Digit. Libr. (2022). https://doi.org/10.1007/s00799-022-00328-z
14. Frachtenberg, E.: Research artifacts and citations in computer systems papers. PeerJ Comput. Sci. **8**, e887 (2022). https://doi.org/10.7717/peerj-cs.887
15. Gong, K.: Open science: the science paradigm of the new era. Cultur. Sci. **5**(1), 3–9 (2022). https://doi.org/10.1177/20966083221091867
16. Gudelj, D., Matošić, M., Poljak, L., et al.: Accessibility of master's theses at SEA-EU Alliance universities in open access repositories. Int. J. Digit. Libr. (2022). https://doi.org/10.1007/s00799-022-00340-3
17. Hu, Y., LeBlanc, Z., Diesner, J., et al.: Complexities associated with user-generated book reviews in digital libraries: temporal, cultural, and political case studies. In: Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries. Association for Computing Machinery, New York, JCDL '22, pp. 1–12 (2022). https://doi.org/10.1145/3529372.3530930
18. Irrera, O., Mannocci, A., Manghi, P., et al.: A novel curated scholarly graph connecting textual and data publications. J. Data Inf. Qual. (2023). https://doi.org/10.1145/3597310
19. Khomo, M.P., Naicker, N., Chisita, C.T., et al.: Factors contributing to the successful development and use of mobile digital libraries: a systematic literature review. Digit. Libr. Perspect. (2023). https://doi.org/10.1108/DLP-08-2022-0062
20. Manghi, P., Candela, L., Lazzeri, E., et al.: Digital libraries: supporting open science. ACM SIGMOD Record **48**(4), 54–57 (2020). https://doi.org/10.1145/3385658.3385669
21. Patra, S., Sahoo, J.: A literature review on digitization in libraries and digital libraries. Preserv. Digit. Technol. Cult. **51**(1), 17–26 (2022). https://doi.org/10.1515/pdtc-2021-0023
22. Schöpfel, J., Rebouillat, V.: Research data sharing and valorization: Developments, tendencies. models. John Wiley & Sons, (2022), google-Books-ID: Fa2AEAAAQBAJ
23. Silvello, G., Corcho, O., Manghi, P., et al. (eds): Linking theory and practice of digital libraries: 26th International Conference on Theory and Practice of Digital Libraries, TPDL 2022, Padua, Italy, September 20–23, 2022, Proceedings, Lecture Notes in Computer Science, vol. 13541. Springer International Publishing, Cham (2022) https://doi.org/10.1007/978-3-031-16802-4, https://link.springer.com/10.1007/978-3-031-16802-4
24. Trisovic, A.: Cluster analysis of open research data and a case for replication metadata. In: 2022 IEEE 18th International Conference on e-Science (e-Science), pp. 423–424 (2022). https://doi.org/10.1109/eScience55777.2022.00069

25. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., et al.: The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data **3**(1), 160018 (2016). https://doi.org/10.1038/sdata.2016.18
26. Yadav, P., Pervin, N.: Towards efficient navigation in digital libraries: leveraging popularity, semantics and communities to recommend scholarly articles. J. Informetr. **16**(4), 101336 (2022). https://doi.org/10.1016/j.joi.2022.101336
27. Yılmaz, N., Kolukısa Tarhan, A.: Quality evaluation models or frameworks for open source software: a systematic literature review. J. Softw. Evol. Process **34**(6), e2458 (2022). https://doi.org/10.1002/smr.2458