The Journal of Electronic Theses and Dissertations

Volume 3 Article 1

Who is keeping an eye on FAIR principles?

Alcidina Cunha Costa
Brazilian Institute of Information in Science and Technology, alcidina@gmail.com

Who is keeping an eye on FAIR principles?

Cover Page Footnote

Data collection and research analysis were carried out while acting as research assistant at IBICT. I have no conflicts of interest to disclose. Correspondence concerning this article should be addressed to Alcidina Cunha Costa, SQN 202, Bloco C, Ap. 502, Asa Norte, Brasília, DF 70.832-030, Brazil. E-mail: alcidina@gmail.com

Who is Keeping an Eye on FAIR Principles?

Alcidina Cunha Costa

Brazilian Institute of Information in Science and Technology, IBICT

Author Note

Data collection and research analysis were conducted as a research assistant at IBICT. I have no conflict of interest to disclose.

Correspondence concerning this article should be addressed to Alcidina Cunha Costa, SQN 202, Bloco C, Ap. 502, Asa Norte, Brasília, DF 70.832-030, Brazil. E-mail: alcidina@gmail.com

Abstract

This systematic review is based on academic work published between 2016 and 2022 at The Networked Digital Library of Theses and Dissertations. The descriptors related to FAIR (Findable, Accessible, Interoperable, and Reusable) are to be found in the title, summary and/or keywords. The result shows that there are few academic studies on FAIR principles; the University of Porto and, consequently, Portugal are the most intensive in FAIR data research. Dissertations and PhD theses are predominant; information and computer science is the field of study most correlated with FAIR.

Keywords: FAIR principles, NDLTD, systematic literature review.

Who is Keeping an Eye on FAIR Principles?

The unequivocal contribution of academic research for constructing knowledge ignited this review on whether FAIR principles (Findable, Accessible, Interoperable, and Reusable) and their substance data, metadata, and infrastructure (GO FAIR Initiative, 2022) are studied as part of research projects. It is a narrow path but may point to the need to stimulate, promote, and establish priorities and funding research based on FAIR implementation.

The systematic review of literature is the tool favoured to address the questions: 1. where electronic theses and dissertations (ETDs) on FAIR principles have been submitted; 2. type of research work (Theses, dissertations, or others), 3. what field of study and 4. To which country they are related. The retrieval work is based entirely on searching The Networked Digital Library of Theses and Dissertations (NDLTD, 2022), the main organization dedicated to storing, preserving, and disseminating ETDs, whether open access or not. The metadata is collected from universities around the world. On March 23, 2022, the NDLTD archive contained 6,220,791 records.

Method

Steps and criteria to allow literature review to be reproducible are detailed by Okoli (2015) and summarized as follows 1. planning and protocol; 2. screening for records exclusion and inclusion; 3. search and data extract; 4. analysis and writing.

The criteria for searching NDLTD was defined by choice of Boolean operators (AND, NOT, OR). The search was limited to 2016 through 2022, with no restrictions on the type of work or language, as long as the title and abstract were translated into English.

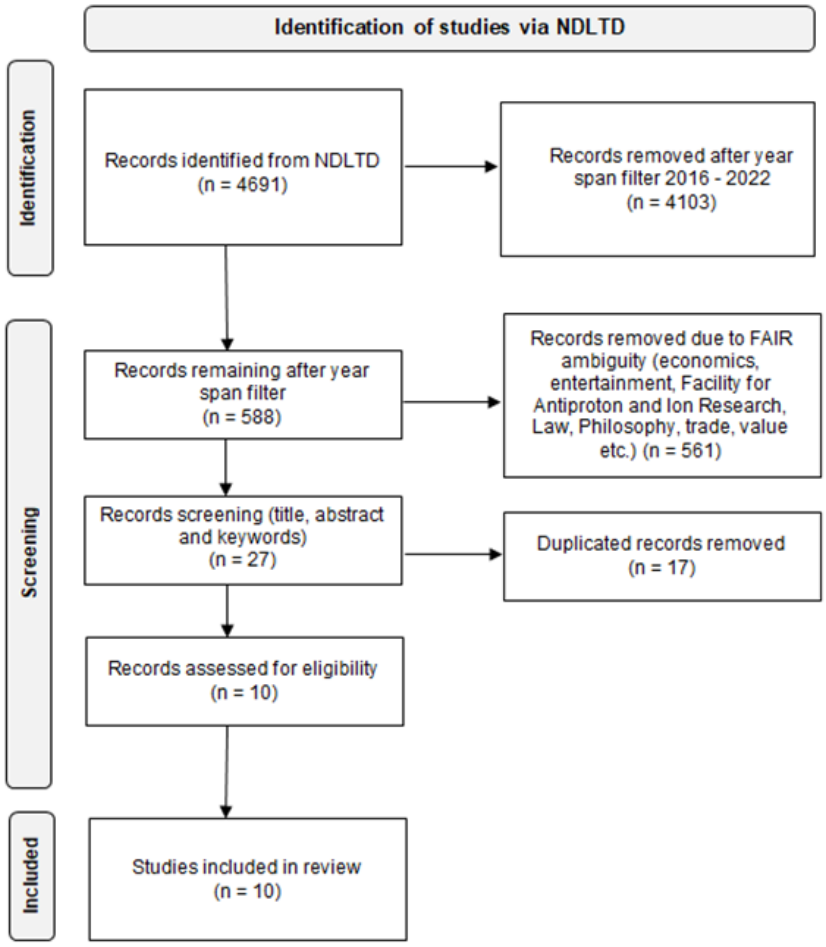
The search strategy was based on a combination of descriptors: <FAIR data OR FAIR principles>, <FAIR AND findable>, <FAIR AND accessible>, <FAIR AND interoperable>,

<FAIR AND reusable>, <FAIR AND infrastructure>, <FAIR AND metadata>, <FAIR AND dataset>.

Studies were excluded when fair was an adjective, an adverb or a noun relating to economics, education, entertainment, equal rights, FAIR (Facility for Antiproton and Ion Research), goods, health, justice, philosophy, services, trade, value, weather, and whenever principles do not apply or are not associated to FAIR digital assets. The studies found to specifically refer to FAIR principles after screening title, abstract and keywords were included in the review. Figure 1 shows the steps of the systematic literature review.

Figure 1

A systematic review based on PRISMA (2022) flow diagram



Results

The literature analysis is, therefore, based on ten studies: Haller et al. (2019); Sampaio (2019); Spicer (2019); Schubert et al. (2019); Maciel (2020); Bertò (2020); Asklöf (2021); Jones et al. (2021); Joia (2021); and Stiebe (2021).

The institutional repositories that hold the papers selected: dissertations (4), theses (2), articles (2), bachelor theses (1) and working papers (1), are Open Access (OA), offering free,

unrestricted online access to scientific and academic papers. As shown in Table 1, they are in six different countries: Portugal (3), Austria (2), Sweden (2), England, Italy, and the USA, the only one outside Europe.

Table 1

Identification and location of repositories

	Repository	University	Country
1	Repositório Aberto da Universidade do Porto	University of Porto	Portugal
2	Apollo University of Cambridge Repository	University of Cambridge	England
3	ePubWU Institutional Repository	WU Vienna University of Economics and Business	Austria
4	IRIS Institutional Research Information System	Università degli studi di Trento	Italy
5	DiVA Institutional Repository	Uppsala University	Sweden
6	Digital Commons @ ETSU Institutional Repository	East Tennessee State University	USA

There were no search results for FAIR principles or else at NDLTD from 2016 to 2018, and 2022. Among the ten selected studies, four were published in 2019, two in 2020, and another four in 2021.

The field of study was agglutinated into four areas: (1) information and computer science (60%), (2) climatological sciences and geosciences (20%), (3) media and communication (10%), (4) sustainable transport (10%).

Figure 2 shows a cloud of words with at least two occurrences amidst all titles and keywords in the texts. Initially, the total number of words was 96; after filtering and excluding articles, conjunctions, prepositions, and verbs, 72 remained. Submitting these to frequency-2, 16 words were counted as most frequent in selected papers.

The most recurrent words are data (20), fair-principles (5), repository (4), analysis (3), engineering (3), management (3), and metadata (3).

Figure 2

Word frequency map of titles and keywords

Discussion

The term FAIR was launched at a Lorentz workshop in 2014, and FAIR principles (Findable, Accessible, Interoperable, and Reusable) were published in Nature Scientific Data (FAIR 2016). Since then, it has been adopted by research institutions worldwide as they face unparalleled volume, complexity, and creation speed of data (FORCE11, 2021).

Findable data requires rich metadata and a unique and persistent identifier such as DOI or ORCID. Accessible means that data is understandable to humans and machines and is stored in repositories. Interoperable metadata implies applicable language for knowledge representation. Reusable indicates that data and collections should have a clear usage license and provide accurate information (GO FAIR Initiative, 2022).

Looking at a hands-on approach, FAIR enables computer systems to access, interoperate, and reuse data with no or minimal human intervention, provides a framework to help researchers manage their data assets and increases collaboration (Columbia University, 2022).

Considering its remarkable advantage and its extensive contribution to fostering research and better decision-making, it is surprising that there are few academic studies on FAIR data principles.

Despite the small sample, the questions were answered: 1. the University of Porto is where FAIR is most discussed; 2. post-graduate degree papers, dissertations (33.3%) and PhD theses (22.2%) are in greater number; 3. Portugal produced the most FAIR academic work (33.3%) and 4. the most prominent field of study is information and computer science (60%).

Nevertheless, other questions surface: What is blocking FAIR implementation? Are there general or specific obstacles? Are they related to a particular principle (findability, accessibility, interoperability, reusability) or substances (data, metadata, and infrastructure)? Why is FAIR

data not widely available from universities where researchers have a long-lasting and well-rooted desire to publish?

Conclusion

FAIR data is widely regarded as important, why has it not taken on a wide research significance? (Ali et al., 2022).

This literature review found few academic works available from NDLTD on FAIR principles, which have the mission and are expected to increase dissemination, visibility, transparency and evaluation of data. It is possible that the rising numbers of institutional repositories, greater investment in FAIR priorities, policies, incentives for implementation; skills development and training will favour an environment to foster FAIR grasp and awareness at the academy and elsewhere (GO FAIR Initiative, 2022).

References

Ali, B. & Dahlhaus, P. (2022). The role of FAIR data towards sustainable agricultural performance: a systematic literature review. *Agriculture*, 12 (2), 309. https://doi.org/10.3390/agriculture12020309

Asklöf, A. (2021). *Streamlining user processes for a general data repository for life science in accordance with the FAIR principles*. (Bachelor thesis, Uppsala Universitet). http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-434648

Bertò, G. (2020). *Supervised learning for white matter bundle segmentation*. (Doctoral thesis, Università degli studi di Trento). https://iris.unitn.it/handle/11572/264971

Columbia University. (2022). Augustus C. Long Health Sciences Library. *What are the FAIR Data Principles?* https://library.cumc.columbia.edu/insight/what-are-fair-data-principles

FAIR principles for data stewardship. (2016). *Nat Genet*, 48 (343). https://doi.org/10.1038/ng.3544

FORCE11. (2021). *The future of research communications and e-scholarship*. https://force11.org/info/the-fair-data-principles/

GO FAIR Initiative. (2022). https://www.go-fair.org/go-fair-initiative/

Haller, A., Fernández, J. D., Kamdar, M. R. & Polleres, A. (2019). *What are links in linked open data? A characterization and evaluation of links between knowledge graphs on the Web*. (Working Paper non-peer-reviewed, WU Vienna University of Economics and Business). https://doi.org/10.1145/3369875

Joia, R. L. (2021). *Towards reproducible and privacy-preserving analyses across federated repositories for Omics data*. (Dissertation, Universidade do Porto). https://hdl.handle.net/10216/135642

Jones, S. A., Blinman, E., Tauxe, L., Cox, J. R., Lengyel, S., Sternberg, R., Eighmy, J., Wolfman, D. & DuBois, R. (2021). *MagIC as a FAIR repository for America's directional archaeomagnetic legacy data*. (Article Faculty work, East Tennessee State University). https://doi.org/10.1029/2021JB022874

Maciel, A. F. C. (2020). *Data Management Plan in accordance with the FAIR principles for FrailSurvey project*. (Dissertation, Universidade do Porto). https://hdl.handle.net/10216/128564

Networked Digital Library of Theses and Dissertations. (2022). ETD Search: Global ETD Search. http://search.ndltd.org/

Okoli, C. (2015). A Guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, 37, 879 – 910. https://doi.org/10.17705/1CAIS.03743

PRISMA transparent reporting of systematic reviews and meta-analyse. (2022). *Flow Diagram*. https://prisma-statement.org//prismastatement/flowdiagram.aspx

Sampaio, M. C. (2019). *Metadata for the use of tools of research data management with I3S researchers*. (Dissertation, Universidade do Porto). https://hdl.handle.net/10216/122409

Schubert, C., Seyerl, G., Sack, K. (2019). Dynamic data citation service-subset tool for operational data management. *Data, 4*(3):115. http://dx.doi.org/10.3390/data4030115

Spicer, R. (2019). *Fit for purpose? A metascientific analysis of metabolomics data in public repositories*. (Doctoral thesis, University of Cambridge). https://doi.org/10.17863/CAM.34945

Stiebe, M. (2021). *A FAIR cross-platform social media analysis approach to sociotechnical sustainable transport research*. (Dissertation, Uppsala Universitet). http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-105759

WordItOut. (2022). Make a Word Cloud. https://worditout.com/word-cloud/create