Who is keeping an eye on FAIR principles?

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 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 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.

1 Introduction

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 [1] are studied as part of research projects. It is a narrow path, but may point to the need to stimulating, promoting, establishing priorities and funding research 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. which country they are related to. The retrieval work is entirely based on searching The Networked Digital Library of Theses and Dissertations, NDLTD [2], 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[1].

2 Methodology

Steps and criteria to allow literature review to be reproducible are detailed by Okoli [3] 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 in time to 2016 through 2022, with no restrictions on type of work or language, as long as title and abstract were translated into English.

[1] http://search.ndltd.org/

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.

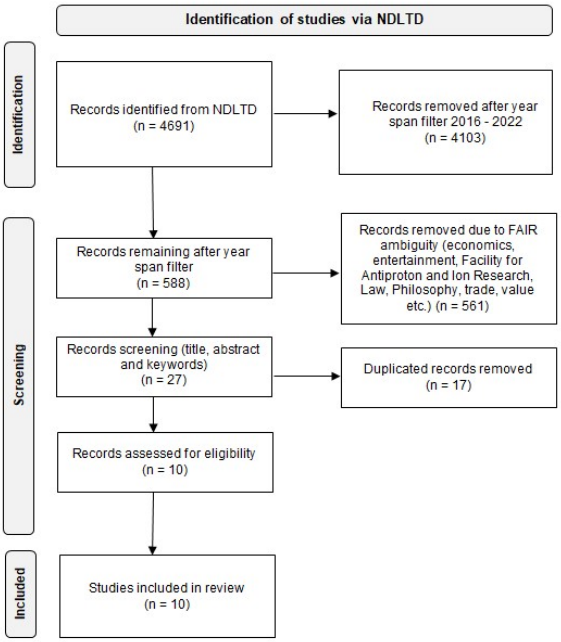

Fig. 1. Systematic review based on PRISMA 2020 flow diagram [4].

3 Results

The literature analysis is, therefore, based on ten studies [5,6,7,8,9,10,11,12,13,14].

The institutional repositories that hold the papers selected dissertations (4), theses (2), articles (2), bachelor thesis (1) and working paper (1), are Open Access (OA), offering free, unrestricted online access to scientific and academic papers. As shown on Table 1, they are located in six different countries: Portugal (3), Austria (2), Sweden (2), England, Italy, and USA, the only one outside Europe.

There were no search results for FAIR principles or else at NDLTD from 2016 through 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%).

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	Upsalla University	Sweden
6	Digital Commons @ ETSU Institutional Repository	East Tennessee State University	USA

Figure 2 shows a cloud[2] for words with at least two occurrences amidst all title and keywords in the texts. Initially, the total number of words was 96; after filtering and excluding articles, conjunctions, prepositions, and verbs, there were 72 left. Submitting these to frequency-2, there were 16 words to be 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).

Fig. 2. Word frequency map of titles and keywords

4 Discussion

The term FAIR was launched at a Lorentz workshop in 2014 and FAIR principles (Findable, Accessible, Interoperable, and Reusable) were published in 2016 on Nature Scientific Data [15]. Since then it has been adopted by research institutions worldwide as they face unparalleled volume, complexity, and creation speed of data [16].

[2] https://worditout.com/word-cloud/create

Findable data asks for rich metadata and a unique and persistent identifier, such as DOI, ORCID. Accessible means that data is understandable to humans and machines, and are 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 [1].

Looking at a hands-on approach, FAIR enables computer systems access, interoperate, and reuse data with no or minimal human intervention; provides a framework to help researchers manage their data assets and increase collaboration [17].

Considering its remarkable advantage and its extensive contribution to foster 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. University of Porto is where FAIR is most discussed at; 2. post-graduate degree papers, dissertation (33.3%) and PhD theses (22.2%), are in greater number; 3. Portugal is the country that produced more FAIR academic work (33.3%) and 4. the field of study most prominent is information and computer science (60%).

Nevertheless, other questions come to surface: what is blocking FAIR implementation? Are general or specific obstacles? Are they related to a specific principle (findability, accessibility, interoperability, reusability) or substances (data, metadata and infrastructure)? Why FAIR data is not widely available from Universities where researchers have a last longing and well rooted desire to publish?

5 Conclusion

FAIR data is widely regarded as important, why is it not widely adopted? [18]

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 on FAIR priorities, policies, incentives for implementing; skills development and training will favour an environment to foster FAIR grasp and awareness at the academy and elsewhere [1].

References

1. GO FAIR Initiative. (2022). Retrieved March 29, 2022 from https://www.go-fair.org/go-fair-initiative/.

2. Networked Digital Library of Theses and Dissertations. (2022). ETD Search: global ETD search. Retrieved March 16, 2022 from https://ndltd.org/.

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

4. PRISMA transparent reporting of systematic reviews and meta-analyse. (2022). Flow Diagram. Retrieved March 3, 2022 from http://prisma-statement.org/prismastatement/flowdiagram.aspx.

5. 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). ID oai:union.ndltd.org:VIENNA/oai:epub.wu-wien.ac.at:7193. Retrieved March 11, 2022 from http://epub.wu.ac.at/view/p_series/S1/.

6. Sampaio, M. C. (2019). Metadata for the use of tools of research data management with I3S researchers. (Dissertation, Universidade do Porto). ID oai:union.ndltd.org:up.pt/oai:repositorio-aberto.up.pt:10216/122409. Retrieved March 11, 2022 from https://hdl.handle.net/10216/122409.

7. Spicer, R. (2019). Fit for purpose?: a metascientific analysis of metabolomics data in public repositories. (Doctoral thesis, University of Cambridge). ID oai:union.ndltd.org:bl.uk/oai:ethos.bl.uk:763861. Retrieved March 11, 2022 from https://www.repository.cam.ac.uk/handle/1810/287634.

8. Schubert, C., Seyerl, G. & Sack, K. (2019). Dynamic data citation service-subset tool for operational data management. (Article peer reviewed, WU Vienna University of Economics and Business). ID oai:union.ndltd.org:VIENNA/oai:epub.wu-wien.ac.at:7180. Retrieved March 11, 2022 from http://dx.doi.org/10.3390/data4030115.

9. Maciel, A. F. C. (2020). Data Management Plan in accordance with the FAIR principles for FrailSurvey project. (Dissertation, Universidade do Porto). ID oai:union.ndltd.org:up.pt/oai:repositorio-aberto.up.pt:10216/128564. Retrieved March 11, 2022 from https://hdl.handle.net/10216/128564.

10. Bertò, G. (2020). Supervised learning for white matter bundle segmentation. (Doctoral thesis, Università degli studi di Trento). ID oai:union.ndltd.org:unitn.it/oai:iris.unitn.it:11572/264971. Retrieved March 11, 2022 from https://iris.unitn.it/handle/11572/264971.

11. Asklöf, A. (2021). Streamlining user processes for a general data repository for life science in accordance with the FAIR principles. (Bachelor thesis, Uppsala Universitet). ID oai:union.ndltd.org:UPSALLA1/oai:DiVA.org:uu-434648. Retrieved March 11, 2022 from http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-434648.

12. 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). ID oai:union.ndltd.org:ETSU/oai:dc.etsu.edu:etsu-works-10769. Retrieved March 11, 2022 from https://doi.org/10.1029/2021JB022874.

13. Joia, R. L. (2021). Towards reproducible and privacy-preserving analyses across federated repositories for Omics data. (Dissertation, Universidade do Porto). ID oai:union.ndltd.org:up.pt/oai:repositorio-aberto.up.pt:10216/135642. Retrieved March 11, 2022 from https://hdl.handle.net/10216/135642

14. Stiebe, M. (2021). A FAIR cross-platform social media analysis approach to sociotechnical sustainable transport research. (Dissertation, Uppsala Universitet). ID oai:union.ndltd.org:UPSALLA1/oai:DiVA.org:lnu-105759. Retrieved March 11, 2022 from http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-105759

15. FAIR principles for data stewardship. (2016). Nat Genet 48 (343). Retrieved March 21, 2022 from https://doi.org/10.1038/ng.3544.

16. FORCE11. (2021). The future of research communications and e-scholarship. Retrieved March 22, 2022 from https://force11.org/info/the-fair-data-principles/.

17. Columbia University. Augustus C. Long Health Sciences Library. Retrieved March 22, 2022 from https://library.cumc.columbia.edu/insight/what-are-fair-data-principles.

18. Ali, B., & Dahlhaus, P. (2022). The role of FAIR data towards sustainable agricultural performance: a systematic literature review. *Agriculture*, 12 (309), 1-17. Retrieved March, 27, 2022 from https://www.mdpi.com/2077-0472/12/2/309.