Placeholder FAIR Digital Objects Conference Proceedings

Placeholder Session title

https://doi.org/10.52825/xxxx........ DOI placeholder (WILL BE FILLED IN BY TIB Open Publishing)

Published: (WILL BE FILLED IN BY TIB Open Publishing)

From Data Citations towards effective Valorisation and Tracking of Data Reuse

Claire Jean-Quartier[1,*] https://orcid.org/0000-0002-0840-2173, Stefan Reichmann[1] https://orcid.org/0000-0003-1544-5064, Miguel Rey Mazon[1] https://orcid.org/0000-0003-3116-1926, and Ilire Hasani-Mavriqi[1] https://orcid.org/0000-0003-0758-0805

[1]Graz University of Technology, Austria

*Correspondence: C. Jean-Quartier, c.jean-quartier@tugraz.at

Abstract. The principles that enabled the “web of documents” to flourish, based on the Resource Description Framework (RDF), have so far not been applied to data [1]. In the past, data has been made available as raw dumps in various formats not sufficiently linked to associated elements. Best practices for linked data based on Unified Resource Identifiers (URIs) lead to the connection of data from diverse domains [2]. The FAIR Digital Object framework has been introduced to provide Digital Objects (DO) with enough actionable information to automate the object’s discovery, access, interoperation, and reuse [[3]. In particular, reuse of scholarly data to drive discovery and innovation has been indicated as the primary motivation of the FAIR principles for scientific data management [4]. We argue that transparency should extend to data reuse including who provided and who reuses DOs, and further, that the concept of FDOs should be extended to enable automated tracking of data reuse.

At present, digital research objects are typically published and assigned Persistent Identifiers (PIDs) [5], but even minimal FAIR practices differ vastly across research contexts (as does data sharing and reuse) [6]. Citations to DOs upon reuse are not common practice [7]. Data citation is mostly informal [8], either by reference to the associated scholarly article, by indication of the underlying web resource hosting the published data, or not at all. To make research objects FAIR and enable their subsequent equitable reuse, including methods to valorise and track the reuse of FAIR DOs, requires a cultural change where researchers play an important role by advocating for open science principles and adopting best practices [9]. Data publication infrastructures, and in particular the associated metadata they store, provide an important pathway to enable FAIR data reuse by integrating scaffolding features [10]. Additionally, in order to incentivise data sharing, a “data index” has been proposed as an alternative metric [11]. However, as with other metrics (h-index, Impact Factor), the social dynamics of attention in academia results in highly skewed patterns of recognition, known as the Matthew Effect [12,13]. Although the Matthew effect can have positive effects in terms of steering researchers’ attention [14], it also results in a stratified distribution of opportunities in academia [15], and has been shown to be at play at the level of article citations [16], countries [17], institutions [18], departments [19], and journals [20]. With respect to data citation, parallel dynamics can be expected. However, unless reused data have been cited properly, no data citation link for its reuse can be created.

The community-driven OpenAIRE Graph endeavours to link and contextualize all research outcomes, integrating heterogeneous data sources and providing relationships among them [21]. Unfortunately, data reuse is not comprehensively captured due to lacking or additional citation options and limitations of the knowledge acquisition [22,23]. The initial strategic approach of EOSC also anticipated to integrate a framework that would allow unique identification, citation, avoidance of duplication, and tracking of impact through a service concept to generate, resolve and validate PIDs, including a policy framework for PIDs for research entities [24]. This concept would need to be further pursued to pave the way for valorizing data provision and create fair data reuse.
Recently, a machine-readable data reuse information tag (DRI) was proposed to enable equitable data reuse for web resources apart from repositories [25]. This tag would include information on the associated Open Researcher and Contributor ID (ORCID) to provide contact information to data consumers. In general, data reuse would need to be commonly traced and linked to data creators - such as ORCID - and mapped in a similar manner to published research output. Linking of PIDs as it is common for scholarly articles, e.g. via Crossref [26], would need to be standard practice for indexing and tracing all published digital objects making data re-findable.

In conclusion, we propose a multilevel approach to valorise and track data reuse: (1) observing Matthew effects in the transition to open science practices, with special attention to data citations. (2) Introducing workflows for capturing published digital records other than scholarly literature based on various PIDs and informally cited data. (3) An extension to existing workflows enabling automatic updating of researchers' ORCID records upon published research objects' reuse.

Keywords: Data Reuse, Data Citation, Research Output Tracking

Data availability statement

The abstract describes a conceptual work, all materials are cited in the reference list.

Author contributions

C.J-Q. (conceptualising, methodology, investigation, writing - original), S.R. (conceptualising, methodology, writing - review), M.R.M. (investigation), I.H-M. (resources)

Competing interests

The authors declare that they have no competing interests.

References

[1] Bizer, Christian & Heath, Tom & Berners-Lee, Tim. (2009). Linked Data. International Journal on Semantic Web and Information Systems. 5. 1-22. https://doi.org/10.4018/jswis.2009081901
[2] Heath, Tom, & Bizer, Christian. (2011). Linked data: Evolving the web into a global data space. Morgan & Claypool Publishers. https://doi.org/10.1007/978-3-031-79432-2
[3] Bonino da Silva Santos, L. O., Sales, T. P., Fonseca, C. M., & Guizzardi, G. (2023). Towards a conceptual model for the FAIR Digital Object Framework. In Formal Ontology in Information Systems (pp. 227-241). IOS Press. https://doi.org/10.3233/FAIA231131

[4] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[5] Kathleen Gregory, Anton Ninkov, Chantal Ripp, Emma Roblin, Isabella Peters, Stefanie Haustein; Tracing data: A survey investigating disciplinary differences in data citation. Quantitative Science Studies 2023; 4 (3): 622–649. https://doi.org/10.1162/qss_a_00264

[6] Tedersoo, L., Küngas, R., Oras, E. et al. Data sharing practices and data availability upon request differ across scientific disciplines. Sci Data 8, 192 (2021). https://doi.org/10.1038/s41597-021-00981-0

[7] Robinson-García, N., Jiménez-Contreras, E. and Torres-Salinas, D. (2016), Analyzing data citation practices using the data citation index. J Assn Inf Sci Tec, 67: 2964-2975. https://doi.org/10.1002/asi.23529

[8] Park, Hyoungjoo, Sukjin You, and Dietmar Wolfram. "Informal data citation for data sharing and reuse is more common than formal data citation in biomedical fields." Journal of the Association for Information Science and Technology 69.11 (2018): 1346-1354. https://doi.org/10.1002/asi.24049

[9] Ramachandran, R., Bugbee, K., & Murphy, K. (2021). From open data to open science. Earth and Space Science, 8, e2020EA001562. https://doi.org/10.1029/2020EA001562

[10] Paul Groth, Helena Cousijn, Tim Clark, Carole Goble; FAIR Data Reuse – the Path through Data Citation. Data Intelligence 2020; 2 (1-2): 78–86. https://doi.org/10.1162/dint_a_00030

[11] Hood, A. S. C., & Sutherland, W. J. (2021). The data-index: An author-level metric that values impactful data and incentivizes data sharing. Ecology and Evolution, 11, 14344–14350. https://doi.org/10.1002/ece3.8126

[12] Perc M. The Matthew effect in empirical data. J R Soc Interface. 2014 Sep 6;11(98):20140378. https://10.1098/rsif.2014.0378

[13] Merton Robert K. 1968 The Matthew effect in science: the reward and communication systems of science are considered. Science 159, 56-63. https://doi.org/10.1126/science.159.3810.56

[14] Merton Robert K. 1988 The Matthew effect in science, II. Cumulative advantage and the symbolism of intellectual property. Isis 79, 606-623. https://doi.org/10.1086/354848

[15] Ross-Hellauer, Tony, Stefan Reichmann, Nicki Lisa Cole, Angela Fessl, Thomas Klebel, and Nancy Pontika. 2022. "Dynamics of Cumulative Advantage and Threats to Equity in Open Science: A Scoping Review." Royal Society Open Science 9(1):211032. https://10.1098/rsos.211032

[16] Wang J. 2014 Unpacking the Matthew effect in citations. J. Informetr. 8, 329-339. https://doi.org/10.1016/j.joi.2014.01.006

[17] Bonitz M, Bruckner E, Scharnhorst A. 1999 The Matthew index—concentration patterns and Matthew core journals. Scientometrics 44, 361. https://doir.org/10.1007/BF02458485

[18] Langfeldt L, Benner M, Sivertsen G, Kristiansen EH, Aksnes DW, Borlaug SB, Hansen HF, Kallerud E, Pelkonen A. 2015 Excellence and growth dynamics: a comparative study of the Matthew effect. Sci. Public Policy 42, 661-675. https://doir.org/10.1093/scipol/scu083

[19] Weakliem DL, Gauchat G, Wright BRE. 2012 Sociological stratification: change and continuity in the distribution of departmental prestige, 1965–2007. Am. Sociol. 43, 310-327. https://doi.org/10.1007/s12108-011-9133-2

[20] Larivière V, Gingras Y. 2010 The impact factor's Matthew effect: a natural experiment in bibliometrics. J. Am. Soc. Inf. Sci. Technol. 61, 424-427. https://doi.org/10.1002/asi.21226

[21] Schirrwagen, J., Bardi, A., Czerniak, A., Loehden, A., Rettberg, N., Mertens, M., & Manghi, P. (2020). Data sources and persistent identifiers in the open science research graph of OpenAIRE. International Journal of Digital Curation, 15(1), 5-5, https://doi.org/10.2218/ijdc.v15i1.722

[22] Paolo Manghi; Challenges in building scholarly knowledge graphs for research assessment in open science. Quantitative Science Studies 2024; 5 (4): 991–1021. https://doi.org/10.1162/qss_a_00322

[23] Peng, C., Xia, F., Naseriparsa, M. et al. Knowledge Graphs: Opportunities and Challenges. Artif Intell Rev 56, 13071–13102 (2023). https://doi.org/10.1007/s10462-023-10465-9

[24] EOSC architecture working group view on the minimum viable EOSC – Report from the EOSC Executive Board Working Group (WG) Architecture, Publications Office, 2021, https://data.europa.eu/doi/10.2777/492370

[25] Hug, L.A., Hatzenpichler, R., Moraru, C. et al. A roadmap for equitable reuse of public microbiome data. Nat Microbiol 10, 2384–2395 (2025). https://doi.org/10.1038/s41564-025-02116-2

[26] Brown, J., Demeranville, T., & Meadows, A. (2016). Open Access in Context: Connecting Authors, Publications and Workflows Using ORCID Identifiers. Publications, 4(4), 30. https://doi.org/10.3390/publications4040030