

Introduction – FAIR data, systems and analysis

Paul Groth ^{a,*} and Michel Dumontier ^b

^a *University of Amsterdam, The Netherlands*

E-mail: p.groth@uva.nl; ORCID: <https://orcid.org/0000-0003-0183-6910>

^b *Maastricht University, The Netherlands*

E-mail: michel.dumontier@maastrichtuniversity.nl; ORCID: <https://orcid.org/0000-0003-4727-9435>

Abstract. The FAIR principles outline key attributes to make digital resources more Findable, Accessible, Interoperable, and Reusable. Globally endorsed and widely adopted, there is now a pressing need to enable the establishment of an Internet of FAIR Data and Services, to demonstrate how these can be used to generate new insights, and to assess the overall value for FAIR across different sectors. Realizing the value of the FAIR principles will require a combination of scientific, technical, social, legal, and ethical advances for the production, sharing, discovery, assessment, and reuse of data. This special issue highlights research and software that is making FAIR data and services a reality.

The FAIR principles have established themselves as key focal point for encouraging the findability, accessibility, interoperability and reusability of data [5]. Given this foundation, there is a need to develop concrete systems and tools that make the production and consumption of FAIR data possible. Likewise, there is a need to analyze, evaluate, measure and understand the ramifications of FAIR data. This special issue highlights research and systems tackling these challenges.

Unique and persistent identifiers are critical for FAIR data in particular for being able to refer to the vocabularies and ontologies that define a data set semantics. The Open Biological and Biomedical Ontologies (OBO) Foundry community maintains many high quality ontologies that use PURLs for identifiers [1]. In their paper [3], because of the change in the underlying service provide for PURLs, they describe (somewhat dramatically!) how they engineered a new PURL identifier service designed to be easily and affordably maintained. This study shows the importance of redirection for persistent identifiers and the importance of maintenance of infrastructures for FAIR data.

To many readers of Data Science, benchmark datasets and results will be familiar. Knowing how our algorithms and models perform on common datasets is critical to measuring the progress of the data science field. However, given the central importance of benchmarking, the results and experiments behind them are not always available in a machine readable fashion. This is where the second paper in this special issue comes in. Röder et al. present their benchmarking platform, HOBBIT (Holistic Benchmarking of Big Linked Data), which provides a benchmarking environment specifically for linked data methods [4]. The results of benchmarking are available as FAIR data. The benchmarking platform has hosted 13,000 experiments and more than 40 benchmarks. The platform provides identifiers and uses a well defined ontology to describe benchmark experiments and their results. Importantly, this paper demonstrates how designing software infrastructure is key to generating FAIR data.

*Corresponding author. E-mail: p.groth@uva.nl.

The final paper in this special issue looks at applying and extending the FAIR principles to consider research software [2]. The authors argue that it is imperative to fold software into the FAIR ecosystem and look at it through the same lens. The paper shows the challenges of expressing interoperability with respect to software. Indeed, reading this paper, it is hard not to wonder if the software ecosystem with its strong global infrastructure and repeatable practices might not become FAIR faster than data. Furthermore, it is worth considering that software is a critical link in helping to define the semantics of data and thus allowing for reusability.

In summary, the papers in this special issue illustrate the gamut of work being done to implement the vision of the FAIR data principles from maintaining fundamental infrastructure and supplying platforms that generate FAIR to reflecting deeply on the applicability of these guidelines.

References

- [1] J. Klump and R. Huber, 20 years of persistent identifiers – Which systems are here to stay?, *Data Science Journal* **16** (2017), 09. doi:[10.5334/dsj-2017-009](https://doi.org/10.5334/dsj-2017-009).
- [2] A.-L. Lamprecht, L. Garcia, M. Kuzak, C. Martinez, R. Arcila, E. Martin Del Pico, V. Dominguez Del Angel, S. van de Sandt, J. Ison, P.A. Martinez, P. McQuilton, A. Valencia, J. Harrow, F. Psomopoulos, J.Ll. Gelpi, N. Chue Hong, C. Goble and S. Capella-Gutierrez, Towards FAIR principles for research software, *Data Science* **3**(1) (2020), 37–59. doi:[10.3233/DS-190026](https://doi.org/10.3233/DS-190026).
- [3] J.A. Overton, M. Cuffaro, The OBO Foundry Operations Committee Technical Working Group and C.J. Mungall, String of PURLs – Frugal migration and maintenance of persistent identifiers, *Data Science* **3**(1) (2020), 3–13. doi:[10.3233/DS-190022](https://doi.org/10.3233/DS-190022).
- [4] M. Röder, D. Kuchelev and A.-C. Ngonga Ngomo, HOBbit: A platform for benchmarking Big Linked Data, *Data Science* **3**(1) (2020), 15–35. doi:[10.3233/DS-190021](https://doi.org/10.3233/DS-190021).
- [5] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao and B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* **3**(1) (2016), 160018. doi:[10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).