

Guidelines for best practices in the publication of scientific data

March 17 2016

March 15 2016, the FAIR Guiding Principles for scientific data management and stewardship were formally published in the Nature Publishing Group journal *Scientific Data*. The problem the FAIR Principles address is the lack of widely shared, clearly articulated, and broadly applicable best practices around the publication of scientific data. While the history of scholarly publication in journals is long and well established, the same cannot be said of formal data publication. Yet, data could be considered the primary output of scientific research, and its publication and reuse is necessary to ensure validity, reproducibility, and to drive further discoveries. The FAIR Principles address these needs by providing a precise and measurable set of qualities a good data publication should exhibit - qualities that ensure that the data is Findable, Accessible, Interoperable, and Re-usable (FAIR).

At the 98th Dies Natalis of Wageningen University, 9th March 2016, Prof. Arthur Mol, Rector of Wageningen UR stated: "Within Wageningen, we will develop a policy and implement practices so that our future data systems are operated according to the FAIR principles of data sciences: findable, accessible, interoperable and re-usable". This will lead to a more transparent research data ecosystem and enhance visibility of the data we are working on, both for the public and for computers. Dr. Richard Finkers, senior scientist Plant Breeding at Wageningen UR is one of the authors of the paper in *Scientific data*. Finkers: "Having datasets computer readable will enhance our ability to re-use data and incorporate published information more effectively in our analysis pipelines and establish links between this information more

effectively (Big data for Discovery). This will allow plant breeders to precisely develop new cultivars, tackling the societal challenge of feeding the world with an ever increasing world population"

The principles were formulated after a Lorentz Center workshop in January 2014 where a diverse group of stakeholders, sharing an interest in scientific data publication and reuse, met to discuss the features required of contemporary [scientific data](#) publishing environments. The first-draft FAIR Principles were published on the Force11 website for evaluation and comment by the wider community - a process that lasted almost two years. This resulted in the clear, concise, broadly supported principles that were published today. The principles support a wide range of new international initiatives, such as the European Open Science Cloud and the NIH Big Data to Knowledge (BD2K), by providing clear guidelines that help ensure all data and associated services in the emergent 'Internet of Data' will be Findable, Accessible, Interoperable and Reusable, not only by people, but notably also by machines.

The recognition that computers must be capable of accessing a data publication autonomously, unaided by their human operators, is core to the FAIR Principles. Computers are now an inseparable companion in every research endeavor. Contemporary scientific datasets are large, complex, and globally distributed, making it almost impossible for humans to manually discover, integrate, inspect and interpret them. This (re)usability barrier has, until now, prevented us from maximizing the return-on-investment from the massive global financial support of [big data](#) research and development projects, especially in the life and health sciences. This wasteful barrier has not gone unnoticed by key agencies and regulatory bodies. As a result, rigorous data management stewardship - applicable to both human and computational 'users' - will soon become a funded, core activity within modern research projects. In fact, FAIR-oriented data management activities will increasingly be made mandatory by public funding bodies.

The high level of abstraction of the FAIR Principles, sidestepping controversial issues such as the technology or approach used in the implementation, has already made them acceptable to a variety of research funding bodies and policymakers. Examples include FAIR Data workshops from EU-ELIXIR, inclusion of FAIR in the future plans of Horizon 2020, and advocacy from the American National Institutes for Health. As such, it seems assured that these principles will rapidly become a key basis for innovation in the global move towards Open Science environments. Therefore, the timing of the Principles publication is aligned with the Open Science Conference under the Netherlands EU Presidency in April 2016.

With respect to Open Science, the FAIR Principles advocate being 'intelligently open', rather than 'religiously open'. The Principles do not propose that all data should be freely available - in particular with respect to privacy-sensitive data. Rather, they propose that all data should be made available for reuse under clearly defined conditions and licenses, available through a well-defined process, and with proper and complete acknowledgement and citation. This will allow much wider participation of players from, for instance, the biomedical domain and industry where rigorous and transparent data usage conditions are a core requirement for data reuse.

"I am very proud that the FAIR Principles were published, just over two years after the Lorentz meeting where we came up with the early principles. They play such an important role in many forward looking policy documents around the world and the authors on this paper are also in positions that allow them to follow these Principles. I sincerely hope that FAIR [data](#) will become a 'given' in the future of Open Science, in the Netherlands and globally", says Barend Mons, Professor in Biosemantics at the Leiden University Medical Center, Head of Node ELIXIR-NL at the Dutch Techcentre for Life Sciences, Integrator Life Sciences at the Netherlands eScience Center, and board member of the

Leiden Centre of Data Science.

More information: Mark D. Wilkinson et al. The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* (2016). [DOI: 10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

Definition of FAIR data: www.dtls.nl/fair-data/

Provided by Wageningen University

Citation: Guidelines for best practices in the publication of scientific data (2016, March 17) retrieved 19 April 2024 from <https://phys.org/news/2016-03-guidelines-scientific.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.