A role for medical writers in overcoming commonly held misconceptions around FAIR data

Erik Schultes
International Science Coordinator
GO FAIR International Support and
Coordination Office

Correspondence to:

Erik Schultes
Poortgebouw N-01
Rijnsburgerweg 10
2333 AA Leiden
The Netherlands
erik.schultes@go-fair.org

Abstract

More and more, computers must participate with physicians and patients as trusted partners in assessing medical options and tracking outcomes. But before the computer can become a routine medical assistant, data and computational services must become Findable, Accessible, Interoperable and Reusable (FAIR) for machines. In this article, the fundamentals of FAIR data and some technology trends are described, with clarification of commonly held and often repeated misconceptions about FAIR. As FAIR was conceived primarily in the life sciences, medical writers are uniquely positioned to help counter these misconceptions.

In March 2016, a little more than 2 years after the initial Lorentz Centre Workshop that launched the now well-known acronym,[1] a brief commentary appeared in the journal *Scientific Data* that quietly enunciated the FAIR (Findable, Accessible, Interoperable and Reusable) Guiding Principles for data stewardship.[2] The commentary gave a vision for a world where data and services could automatically interoperate and framed some of the current barriers to getting there. The FAIR Principles themselves were composed of 15 one-liners, were presented as a nondescript figure item (Box 2 in the commentary), and were accompanied by no in-depth discussion about how they should actually be implemented. Despite this innocuous beginning, the FAIR Principles would go on to take the world by storm.

The FAIR Principles make digital data and services findable, accessible, interoperable, and reusable not only to human users (e.g., doctors, patients, researchers who may be sitting behind computers or using smart phones) but also, in more and more automated ways, by computers themselves. Despite the egalitarian ring to the FAIR acronym, the FAIR Principles are actually concerned with only the technical issues regarding data handling, although data that are FAIR for machines are arguably fairer for humans.

The FAIR Principles emerged in response to the key conundrum posed by big data: although combining data from multiple sources into ever-larger collections creates immense value and opportunity, the rapid growth rates of data production mean than that such data collections cannot be built manually. Often the data are hidden in plain sight, but even when they are known, the human intervention required to process them makes their reuse impractical.[3] Without FAIR, future investments in big data, data analytics, and AI will bring only diminishing returns.[4] Under the growing weight of this conundrum, publicly declared endorsements of and commitments to the FAIR Principles came suddenly and decisively from across the stakeholder community. Among the earliest to commit were high-level organisations that wanted to harness big data analytics for high-value applications. These included powerful economic development alliances like the G20,[5] the G7,[6] and the World Economic Forum.[7] Almost as soon as the FAIR Principles were published, the European Commission announced its ambition to build the "European Open Science Cloud", a revolutionary data infrastructure based on the FAIR Principles that would safeguard member states' research data, particularly data produced with public funds (totalling €300 billion annually).[8,9] Similar large-scale public data infrastructure based on the FAIR Principles has also been launched in other countries and regions including the US, China, and Africa.[10] Now, 4 years after their publication, we begin to see serious discussions on the implementation choices and challenges around the FAIR Principles from those stakeholders in the trenches who are actually "doing" FAIR data. In addition to a growing list of publications,[11] research projects,[12] and initiatives,[13] private industry[14] is also beginning to offer solutions to help make FAIR data and services more practical.

With this explosive development of FAIR, a number of misconceptions about FAIR have taken root throughout the stakeholder community, including among policy makers and self-proclaimed practitioners of FAIR.[15] Four examples are given below, together with a few tips and tricks to keep in mind when writing about FAIR. Because FAIR originated largely in the life sciences and has been driven oftentimes by applications in the biomedical space, medical writers are uniquely positioned to help mitigate, if not actually rectify, these common misconceptions about what FAIR is, what it is not, and its relevance to various stakeholders.

First, although the ultimate purpose of FAIR is to help people (e.g., patients, doctors, epidemiologists, insurance companies) to solve complex problems and make better-informed health and medical decisions, the FAIR principles are directed at machine agents. "FAIR for machines" is therefore a redundant phrase.

Second, FAIR is not a data format, computer protocol, or standard. Rather, it is a set of principles which can be applied to data formats, computer protocols, and standards to ensure they are machine-actionable. There tends to be broad and general agreement around the FAIR Principles (like there is for world peace), but finding widespread agreement on implementation choices for FAIR can be a challenge. Among other mechanisms, the publicly funded GO FAIR International Support and Coordination Office (go-fair.org), a joint project launched by ministries in the Netherlands, Germany, and France, has as its mandate the acceleration of bottom-up convergence on standards and technologies to achieve broadly
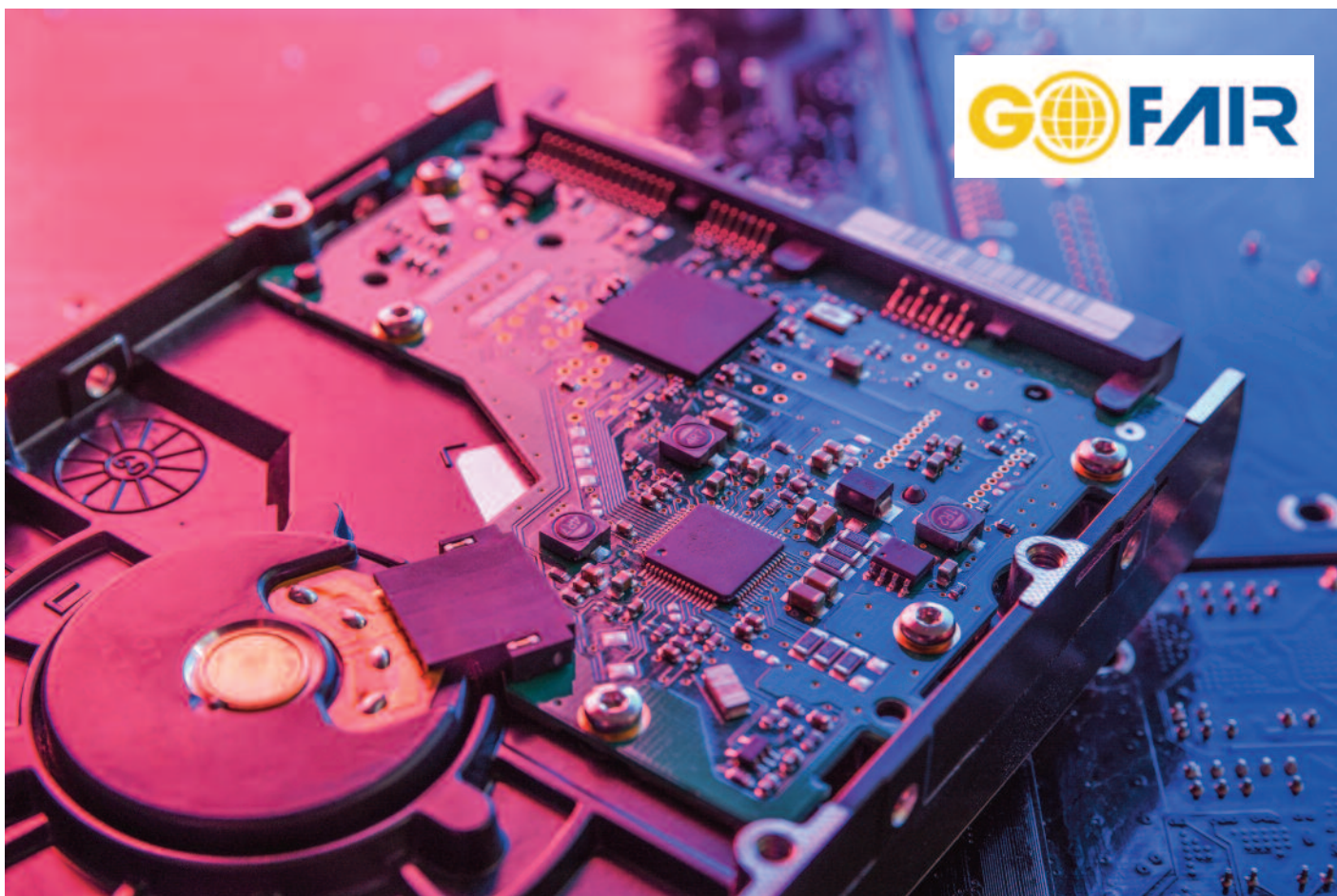
accepted FAIR implementation.

Third, and highly relevant for medical and pharmaceutical data, FAIR is often confounded with "open source", "open access", and "free". The principles under the "A" in FAIR stipulate only that the conditions to access data are so clearly spelled out that even a machine will know exactly what actions it must take before it can perform operations like data query or advanced analytics.[16] Hence, sensitive data like patient records may be rendered perfectly FAIR, but because of GDPR and other privacy restrictions, will never be made open. In other cases, like FAIR data created within private companies, part of the access procedure could be to accept a restrictive licence and pay a fee. The data in these cases are not open or free but could (and should) be made FAIR. Separating FAIR and open makes it possible for data owners to preserve control over their data while at the same time allowing these data to interoperate with other resources if and when needed. New approaches allow for fine-grained access control, allowing certified algorithms to access some data elements (e.g., patient blood pressure measurements) but not others (e.g., patient personal identification data). These types of applications open the door to so-called "distributed learning applications", also known as FAIR Data Trains.[17] In this approach, certified algorithms (Trains) are dispatched to the data (FAIR Data Stations) where access is given to whatever data are allowed in that particular case. FAIR Data Trains turn the traditional idea of data sharing upside down, giving control of data access to data owners (perhaps even patients themselves) who in turn grant permission to algorithms that visit the data. "Data visiting" rather than "data sharing" will open up the tremendous knowledge stored in patient records to researchers, while privacy and personal information remain protected.

Fourth, FAIR often raises concerns over data quality and trustworthiness: Might data (e.g., personal health data from wearable devices) be published without checks like peer review? How can we trust FAIR data? Oftentimes, researchers and policy makers propose to expand FAIR with additional principles like "Ethics" and "Responsibility" (FAIR-ER data), but by design explicit references to data quality were not included in the FAIR Principles. Data quality is context-dependent, where data of high quality in one case might be viewed as low quality in another case. The FAIR Principles prompt the creation and use of "rich" machine-readable provenance metadata (Principle R1.2), which is often missing in conventional datasets. Provenance metadata refers to any and all relevant information about the creation of the data: when and where the data were produced, and by whom; the source of funding for the data acquisition; and the methods and instruments used to collect and analyse the data. These and many other elements belong in properly constructed FAIR data. Such metadata will, by inference (either human or machine), only improve the trustworthiness of the data and ensure more accurate assessment of data quality.

After 4 years and more than 2600 citations, the FAIR Principles have already made an impact on data-intensive disciplines, especially in health, medicine, and biomedical sciences. Undoubtably, there will much more to say about FAIR in medicine in the coming years. Of all the many initiatives around FAIR, the most recent and most urgent is driven by the COVID-19 pandemic. A newly minted GO FAIR Implementation Network called the Virus Outbreak Data Network (VODAN),[18] composed of public and private regional authorities, is now attempting to install an international network of FAIR Data

Stations across the world and make COVID-19 electronic case report forms available for automated global data visiting. Driven by the urgency of the disease outbreak, FAIR is being put to the test on a scale never attempted before. Accurately communicating the aims and methods of FAIR in the coming years will continue to be a challenge, but the resources needed by journalists and medical writers to decipher the FAIR Principles are there, if you know where to look.[11,19,20] One day, even these resources themselves might be made FAIR.

Conflicts of interest
The author declares no conflicts of interest.

References
1. Dutch Techcentre for Life Sciences. Jointly designing a Data FAIRPORT. 2014 Jan 20. Available from: https://www.dtls.nl/2014/01/20/jointly-designing-data-fairport/

2. Wilkinson MD, Dumontier M, Aalbersberg I, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3(160018).

3. Press G. Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says. 2016. Available from: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#52a441026f63

4. PwC EU Services. Cost of not having FAIR research data. Brussels: Directorate-General for Research and Innovation, 2019.

5. European Commission. Article 12, G20 Leaders' Communique Hangzhou Summit. 2016. Available from: https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_16_2967

6. G7 Science and Tsukuba Communiqué. G7 Science and Technology Ministers' Meeting. 2016 May 15-17. Tsukuba, Ibaraki.

7. World Economic Forum. Value in Healthcare: Mobilizing cooperation for health system transformation. Geneva, Switzerland: World Economic Forum; 2018.

8. European Commission. Communication from the Commission to the European Parliament, the Council, The European Economic and Social Committee and the Committee of the Regions. Brussels, Belgium: European Commission; 2016.

9. Mons B. Invest 5% of research funds in ensuring data are reusable. Nature. 2020;578(7796).

10. Meerman B. OPEN SCIENCE CLOUD INITIATIVES: Region = FAR EAST. 2019. Available from: https://www.eoscsecretariat.eu/eosc-liaison-platform/post/open-science-cloud-initiatives-region-far-east

11. Mons B, Schultes E, Liu F, Jacobsen A. The FAIR Principles: first generation implementation choices and challenges. Data Intelligence 2020;2(2020):1-9. Available from: http://www.data-intelligence-journal.org

12. European Union Horizon 2020. FAIRsFAIR "Fostering FAIR Data Practices In Europe". European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. Brussels, Belgium: European Commission; 2020.

13. GO FAIR. FAIR Made Easy for Funders. Available from: https://www.go-fair.org/today/FAIR-funder

14. Phortos Consultants. The FAIR Service Provider Consortium (FSPC). Available from: http://www.phortosconsultants.com/Consortium

15. Mons B, Neylon C, Velterop J, et al. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Inf Serv Use. 2017;37(1):49-56.

16. Jacobsen A, de Miranda Azevedo R, Juty N, et al. FAIR Principles: interpretations and implementation considerations. Data Intelligence. 2020;2(1):10-29.

17. Dutch Techcentre for Life Sciences. Personal Health Train. 2016. Available from: https://vimeo.com/143245835

18. GO FAIR. Virus Outbreak Data Network (VODAN). 2020 [cited 2020 Apr 21]. Available from: https://www.go-fair.org/implementation-networks/overview/vodan/

19. Kryger Hansen K, Buss M, Sztuk Haahr L; FAIR project. A FAIRy tale. Available from: https://vidensportal.deic.dk/sites/default/files/uploads/A%20FAIRy%20tale%20book%20digital_ny.pdf

20. Go FAIR. FAIR Principles. Available from: https://www.go-fair.org/fair-principles/
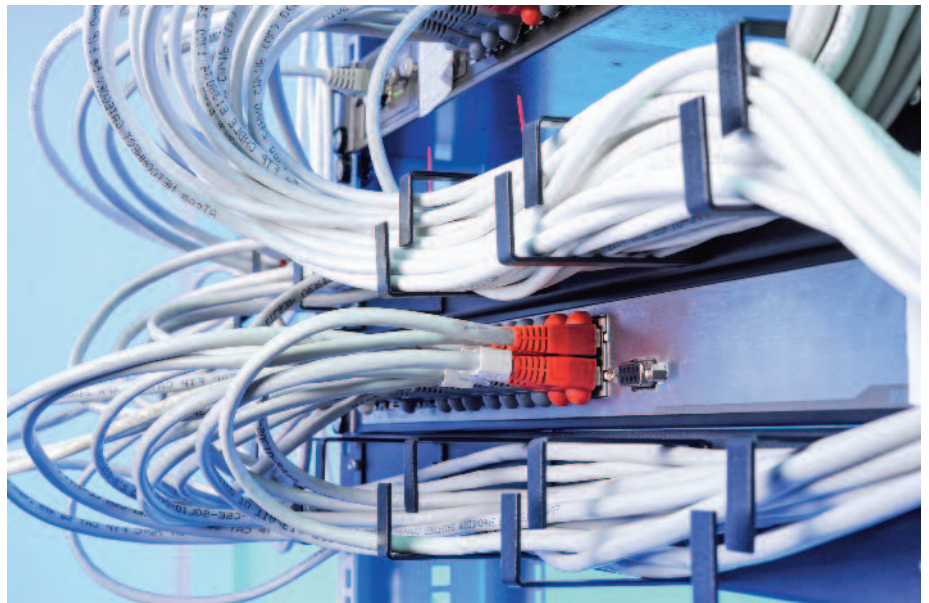
Author information
Erik Schultes is an evolutionary biologist with a long-standing interest in data-intensive research questions. As International Science Coordinator at GO FAIR, Erik has worked with a diverse community of stakeholders to develop FAIR data and services, both within and outside the life sciences. https://orcid.org/0000-0001-8888-635X.

EMWA NEEDS YOU

EMWA is a member-run organisation

When you volunteer to assist EMWA in any capacity, you are furthering the development of our association.

You can choose how you want to get involved: in a very limited way or as part of a larger project. The choice is yours, and everyone shares the benefits.

EMWA members can volunteer in the following areas:

Conference
- Planning committee
- Advertising

Finance

Journal
- Contributions
- Editorial Board

Website
- Contributions
- Web Team

Freelance Business Group

Social Media Team

Training
- Leading workshops
- Professional Development Programme Committee
- Webinar contributions
- Webinar Team

Special Interest Groups
- Pharmacovigilance
  - Regulatory Disclosure
  - Medical Devices
  - Medical Communications
  - Veterinary Medical Writing
  - Sustainability

Ambassador programme

Executive Committee
- President
- Vice President
- Journal Editor
- Public Relations Chair
- Conference Chair
- Honorary Secretary
- Professional Development Programme Committee Chair
- Treasurer

TO FIND OUT MORE
If you are a member of EMWA and eager to support ongoing initiatives, please contact info@emwa.org.

WHY VOLUNTEER?
- Help promote the role of medical writers and strengthen our association
- Help to raise standards in our field
- Increase your visibility and communication opportunities within the medical writing community
- Add some prestige to your CV
- Improve your knowledge of medical writing and related topics