How FAIR are pharma publication data?

Eniola Awodiya,[1] Joana Osório[2]

1   Medical Sciences Division,
     University of Oxford, UK
2   Oxford PharmaGenesis, Oxford, UK

Correspondence to:

Joana Osório
joana.osorio@pharmagenesis.com

Abstract

Sharing research data increases reusability, reduces waste, supports reproducibility and promotes innovation. In medical research, sharing data also promotes transparency and access to information relevant to patient care.

While important advancements have been made in data sharing by regulators, the pharmaceutical industry and academic publishers, several barriers remain. Some of these barriers stem from concerns about data privacy and patient safety, but others are related to the need for confidence in sharing, which can be improved through agreed standards and systems for reuse of research data, including the application of FAIR (findable, accessible, interoperable, and reusable) principles and the overarching principle of "as open as possible, as closed as necessary".

Medical writers, who are key links between the pharmaceutical and publishing industries, can contribute to making pharma publication data FAIRer. They also have an important role in educating others about the path to more findable, accessible, interoperable, and reusable data.

Introduction

Good data management is essential in a healthy research ecosystem. Greater access to appropriately shared data increases reusability, reduces waste of resources, supports reproducibility, and promotes innovation.

The landscape of healthcare data sharing has changed considerably over the past decade. Registration of clinical trials and disclosure of the results of many types of trials are mandatory in the US, the EU, UK and other countries.[1–2] Publicly available lay language summaries of clinical trials are also mandatory in the EU.[2] The data sharing policies of most pharma companies adhere to the principles and positions developed by industry bodies such as the Pharmaceutical Research and Manufacturers of America (PhRMA) and the European Federation of Pharmaceutical Industries and Associations (EFPIA).[3–5] These commitments include publishing all human trial results, including "negative" results, in appropriate peer-reviewed journals.

While the principles of registering and disclosing clinical trial results are now widely accepted, even if incompletely adopted,[6] sharing individual patient data is more complex. The principles of open science and open data championed by many funders, regulators, and national and international policy organisations have to be balanced with responsibilities towards patient privacy, legal consent, data ownership, and intellectual property.[7–21]

The use of repositories that meet data sharing requirements while protecting individual privacy can help ensure that research data are "as open as possible, as closed as necessary". This statement is often associated with the FAIR principles, which state that research data should be findable, accessible, interoperable, and reusable by both humans and machines (Table 1).[22–23] The FAIR principles

> Repositories that meet data sharing requirements while protecting individual privacy [ensure] data are "as open as possible, as closed as necessary".

seek to address the rising need to strengthen the infrastructure supporting the reuse of scholarly data, so that data use can be automated and standardised. Importantly, FAIR principles apply to both the raw data and to their associated metadata – the data that enable discovery, linkage, and reuse across multiple systems.

The FAIR principles are domain-agnostic, and can be applied to many types of data including clinical trial and healthcare data.[24] As medical writers have key roles in communicating research findings, and preparing data to be shared, they are well placed to influence data sharing best practice in publications and promote FAIR data sharing efforts where possible.

FAIR principles in medical publishing
Most medical publishers have endorsed the International Committee of Medical Journal Editors (ICMJE) recommendations on data sharing.[25–26] Authors of articles that report clinical trials must submit a data sharing statement with their manuscript and, for all trials that began enrolment after January 1, 2019, they must also include a data sharing plan when registering the trial.

Some publishers have also endorsed the FAIR principles, at least for a subset of their journals (Table 2), and more journals may join them as the support for FAIRer data grows.

How findable are pharma publication data?
Findability refers to how easily identifiable published data are. Table 1 highlights the four aspects of this principle.

Aggregated and summary data produced by pharma companies are fairly easy to discover through platforms such as ClinicalTrials.gov[31] and publications. The link to the raw data used to produce these summary outputs, however, is not always obvious.

Pharma companies often deposit patient-level data from clinical trials on repositories that are used by multiple companies or institutions. These repositories include Vivli,[32] ClinicalStudyDataRequest.com,[33] and the Yale University Open Data Access (YODA) Project.[34]

On Vivli, for example, all data sets are assigned a unique digital object identifier (DOI), in line with the F1 criterion. Furthermore, Vivli uses patient population, intervention, comparison and outcomes (PICO) searches designed to yield more precise search results from broader clinical questions to optimise findability. Vivli has a process to extract and curate metadata from source documents and other indexing platforms, although the current catalogue does not yet enable highly precise search and browse functionality.

Some pharma companies also use company-specific databases to hold and share some types of data, such as non-interventional trial data. In these cases, it is often unclear whether F1–F4 criteria are being met.

Journals with more stringent data sharing

Table 1. The FAIR principles[22]

Findable

F1.	(Meta)data are assigned a globally unique and persistent identifier.
F2.	Data are described with rich metadata (defined by R1 below).
F3.	Metadata clearly and explicitly include the identifier of the data they describe.
F4.	(Meta)data are registered or indexed in a searchable resource.

Accessible

A1.	(Meta)data are retrievable by their identifier using a standardised communications protocol.
A1.1	The protocol is open, free, and universally implementable.
A1.2	The protocol allows for an authentication and authorisation procedure, where necessary.
A2.	Metadata are accessible, even when the data are no longer available.

Interoperable

I1.	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2.	(Meta)data use vocabularies that follow FAIR principles.
I3.	(Meta)data include qualified references to other (meta)data.

Reusable

R1.	(Meta)data are richly described with a plurality of accurate and relevant attributes.
R1.1.	(Meta)data are released with a clear and accessible data usage license.
R1.2.	(Meta)data are associated with detailed provenance.
R1.3.	(Meta)data meet domain-relevant community standards.

policies (Table 2) require data to have a persistent identifier, meeting FAIR criterion F1. They also recommend that data are properly labelled and described, and that they are associated with the appropriate metadata for the kind of data being shared, supporting criteria F2–F3. Many of these journals also recommend data deposition in a searchable repository, supporting criterion F4.

A key role of medical writers is to prepare the metadata and documentation needed by secondary users of research data. They may also be asked to advise on the choice of repository for specific, less common types of data. By understanding that FAIR requirements also extend to metadata and by understanding the characteristics of a FAIR-compliant repository, writers are instrumental in increasing the FAIR findability of data associated with publications.

How accessible are pharma publication data?

Accessibility supports data reuse and integration. Importantly, accessible data are not the same as open data. Data that are not in the public domain but that are accessible to qualified researchers, after evaluation by a review panel, are not open but can be FAIR. Two FAIR criteria address the principle of accessibility (Table 1).

Criterion A1 refers to the ability of retrieving data or metadata using an open, free, and standardised protocol that also includes an authorisation and authentication process when necessary. Following regulatory and industry guidelines, pharma companies have committed to deposit summary results for eligible trials conducted in the US on the ClinicalTrials.gov platform[31] and for all trials conducted in the EU on the EudraCT platform,[35] within specified time frames. The next step towards accessibility is to continue to increase the rates of clinical trial data sharing on these platforms by both pharma and academic researchers.

> Unusual or complex data sets might be more difficult to standardise owing to the time and costs involved in data curation.

Increasing the accessibility of patient-level data is a harder issue to tackle. Pharma companies have justifiable concerns about data privacy and patient safety. They must also overcome hurdles associated with the costs of data management and curation, and the potential delays in publication timelines that may result from preparing data to be shared in appropriate formats and deposited in specific platforms.

However, even heavily protected and private data can be FAIR, if the metadata clearly states the data privacy requirements restricting access to data.[22] The YODA repository provides detailed information on frequent reasons data access requests may be denied, such as restrictions arising from informed consent agreements with patients.[36]

Vivli stores data for up to 10 years and maintains the persistent DOI associated with the description of a data set even after the data set is

---

Table 2. Publishers' data sharing policies

Policy type	Data sharing	Data citation	Data availability statement	Peer review of data	Licence applied to data set	FAIR standards for data
Encourages data sharing	Encouraged	Encouraged	Optional	Optional	Author's choice	Optional
Encourages data sharing with evidence	Encouraged	Encouraged	Encouraged	Optional	Author's choice	Optional
Expects data sharing	Encouraged	Encouraged	Required	Optional	Author's choice	Optional
Mandates data sharing	Encouraged	Required	Required	Optional	Author's choice	Optional
Mandates data sharing and peer review of data	Required	Required	Required	Required	Author's choice	Optional
Mandates data sharing and peer review of data, which must be open	Required	Required	Required	Required	CC0, CC BY or equivalent (open data)	Optional
Mandates data sharing and peer review of data, which must be open and fully FAIR	Required	Required	Required	Required	CC0, CC BY or equivalent (open data)	Required

The table shows an overview of data sharing policies, combining information from four publishers: Taylor & Francis,[27] Springer Nature,[28] Wiley,[29] and the Public Library of Science (PLOS).[30] Springer Nature and Wiley have four tiers of data sharing for different journals, whereas Taylor & Francis has five tiers. PLOS has a single data sharing policy for all its journals.

CC0, Creative Commons Zero; CC BY, Creative Commons Attribution.

no longer accessible,[37] supporting the A2 criterion. Whether metadata are accessible after the data are no longer available in other repositories is less clear. Unification and standardisation of repository accessibility criteria could help to increase FAIRness.

Journals that mandate data sharing often recommend authors use an external repository, which can help meet A1.1 and A1.2 criteria. Journals do not typically require publication-associated data to meet criterion A2, although publications in some Taylor & Francis journals must meet fully FAIR criteria.[27]

At the bridge between pharma research and publishers, medical writers are uniquely positioned to help pharma companies prepare even protected data in the most accessible way possible and support authors with information they may need to overcome data accessibility barriers.

How interoperable are pharma publication data?

Interoperable data can be integrated with other data, applications, or workflows for analysis, storage, and processing. Table 1 shows the three criteria within this principle.

Pharma companies are making increasing efforts to structure and annotate their data in a way that enables and facilitates interoperability and reuse.[38] Enabling interoperability of data from multiple sources is also one of the main stated goals of the Vivli platform.[32] Despite these advances, it is often difficult to reuse and reanalyse data across different systems, even when the data are accessible. The COVID-19 pandemic, which highlighted the healthcare benefits of data sharing at scale, spurred several pharma voices to call for greater efforts to increase interoperability of data across the industry.[39]

Medical writers can advance this FAIR principle by helping to produce data and metadata in standardised formats and with controlled vocabularies that make them easy to integrate in multiple workflows, users and systems. Using standard file types, machine-readable text rather than PDFs and open-source software rather than proprietary software for storing data (for example, .csv rather than Excel) can help, as well as clearly annotating table headings, scales and other elements that make the data easy to reuse across systems. In addition to preparing data for publications, medical writers can also help to promote repositories that favour interoperable data.

How reusable are pharma publication data?

Publishing (meta)data in a manner that increases its use(ability) for the community is the primary objective of FAIRness.[22] This FAIR principle has one criterion with three components.

Criterion R1.1 states that (meta)data are released with a clear and accessible data usage licence. The extent to which this criterion is being met across all pharma publication data is unclear, but the Pistoia Alliance, an industry colla-boration, does recommend that there are always human-readable and machine-readable pointers in the metadata to the data owner or license.[38]

Several journals, including the most prestigious medical journals, request that protocols and statistical analysis plans are available for clinical trial publications, and there is evidence that researchers, including pharma companies, are complying with this require-ment.[40] In addition, many publishers require data citations (Table 2), in alignment with the Joint Declaration of Data Citation Principles.[41] These efforts support criterion R1.2, which recommends that data and metadata are associated with their provenance.

The data sets published as part of a pharma publication typically adhere to domain-relevant community standards, in line with criterion R1.3.

Those data sets that are submitted to a known repository do so as well, as the requirements for data submission and entry into these platforms encourage standardisation. However, unusual or complex data sets might be more difficult to standardise owing to the time and costs involved in data curation. Recognition of the value of data curation and an appropriate set of incentives for this type of work could promote further adherence to R1.3.

Conclusion

Open Pharma exists to improve the transparency, accountability, accessibility, and discoverability of pharma publications;[42] this ultimately has an impact on patient care. Responsible data sharing supports all of these goals and can increase trust in the pharma industry and its research outputs.

Responsible data sharing recognises the public health benefits of access to reusable data, but also the rights of patients and other people involved in clinical research. To make data available while protecting the personal informa-tion of research participants, researchers may need to anonymise or randomise data, and even omit data that, when shared, would be likely to reveal a patient's identity. Any data sharing increases transparency, but not all data sharing enables reproducibility and reuse.

While pharma publication data cannot always be open, for legal, privacy and safety reasons, there's an opportunity to make them FAIRer. Improvements are in progress but there is still much to be done, including: standardisation of data structure, metadata and systems to enable interoperability; unification of policies across regulators, publishers, pharma companies, and other research data producers; clearer guidance about metadata that advances all four FAIR principles while decreasing the burden of data management; and education of authors on the benefits of FAIRer data for their research.

Medical writers, linking academic authors, pharma companies, and publishers, are in an influential position to drive these changes.

Acknowledgements

Disclosures and conflicts of interests

References
1. ClinicalTrials.gov. FDAAA 801 and the final rule [cited 2022 Sept]. Available from: https://clinicaltrials.gov/ct2/manage-recs/fdaaa

2. EMA. Clinical trials regulation. Implemented January 31, 2022 [cited 2022 Sept]. Available from: https://www.ema.europa.eu/en/human-regulatory/research-development/clinical-trials/clinical-trials-regulation

3. PhRMA Principles for responsible clinical trial data sharing [cited 2022 Sept] . Available from: https://www.phrma.org/-/media/Project/PhRMA/PhRMA-Org/PhRMA-Org/PDF/P-R/PhRMAPrinciplesForResponsibleClinicalTrialDataSharing.pdf

4. IFPMA, EFPIA, JPMA, PhRMA. Joint position on the disclosure of clinical trial information via clinical trial registries and databases (November 10, 2009, revised January 15, 2018) [cited 2022 Sept]. Available from: https://www.ifpma.org/wp-content/uploads/2010/11/Joint-Position-on-Disclosure-of-CT-Info-via-CT-Registries-Revised-Jan2018-vFINAL.pdf

5. IFPMA, EFPIA, JPMA, PhRMA. Joint position on the publication of clinical trial results in the scientific literature (June 10, 2010) [cited 2022 Sept]. Available from: https://www.ifpma.org/wp-content/uploads/2016/03/June2010_Joint_Position_CT_Data_Publication-scientific_literature.pdf

6. Joint Letter by the European Commission, EMA and HMA (June 2019). Letter to stakeholders regarding the requirements to provide results for authorised clinical trials in EudraCT [cited 2022 Sept]. Available from: https://eudract.ema.europa.eu/docs/guidance/1R%20-%20Joint%20letter%20signed%20by%20EC,%20EMA,%20HMA_en.pdf

7. Ursin G, Malila N, Chang-Claude J, et al. Sharing data safely while preserving privacy. Lancet. 2019;394(10212):P1902. doi:10.1016/S0140-6736(19)32603-0

8. World Medical Association. WMA Declaration of Taipei on ethical considerations regarding health databases and biobanks [cited 2022 Sept]. Available from: https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/

9. Centers for Disease Control and Prevention. Health Insurance Portability and Accountability Act of 1996 (HIPAA) [cited 2022 Sept]. Available from: https://www.cdc.gov/phlp/publications/topic/hipaa.html

10. General Data Protection Regulation (GDPR) [cited 2022 Sept]. Available from: https://gdpr-info.eu/

11. EMA. Publication and access to clinical-trial data, draft for public consultation (June 24, 2013) [cited 2022 Sept]. Available from: https://www.ema.europa.eu en/documents/other/draft-policy-70-publication-access-clinical-trial-data_en.pdf

12. CARE principles for indigenous data governance (November 08, 2018) [cited 2022 Sept]. Available from: https://www.gida-global.org/care

13. NIH. Data management & sharing policy overview [cited 2022 Sept]. Available from: https://sharing.nih.gov/data-management-and-sharing-policy/about-data-management-and-sharing-policy/data-management-and-sharing-policy-overview#after

14. Medical Research Council. MRC Policy and guidance on sharing of research data from population and patient studies [cited 2022 Sept]. Available from: https://www.ukri.org/wp-content/uploads/2021/08/MRC-0208212-MRC-policy-and-guidance-on-sharing-of-research-data-from-population-and-patient-studies-Word-version-v01.02.pdf

15. Wellcome. Data, software and materials management and sharing policy [cited 2022 Sept]. Available from: https://wellcome.org/grant-funding/guidance/data-software-materials-management-and-sharing-policy

16. Gates Foundation. Data sharing requirements [cited 2022 Sept]. Available from: https://openaccess.gatesfoundation.org/how-to-comply/data-sharing-requirements/)

17. Cancer Research UK. Data sharing guidelines (April 01, 2009) [cited 2022 Sept]. Available from: https://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/submission-of-a-data-sharing-and-preservation-strategy/data-sharing-guidelines

18. Concordat on open research data (July 28, 2016) [cited 2022 Sept]. Available from: https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf

19. EMA policy on publication of clinical data for medicinal products for human use (March 21, 2019) [cited 2022 Sept]. Available from: https://www.ema.europa.eu/en/documents/other/european-medicines-agency-policy-publication-clinical-data-medicinal-products-human-use_en.pdf

20. NHS #Make it Public, Transparency and openness in health and social care research (July 2020) [cited 2022 Sept]. Available from: https://s3.eu-west-2.amazonaws.com/www.hra.nhs.uk/media/documents/8828_transparency_strategy_2020_V4.pdf

21. UNESCO. Draft recommendation on open science (2021). Available from: https://unesdoc.unesco.org/ark:/48223/pf0000378841

22. GO FAIR. FAIR Principles (2016). Available from: https://www.go-fair.org/fair-principles/

23. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 2016;3:160018.

24. Pistoia Alliance. FAIR implementation supported by a FAIR toolkit for the life science industry [cited 2022 Sept]. Available from: https://www.pistoiaalliance.org/projects/current-projects/fair-implementation/

25. Taichman DB, Sahni P, Pinborg A, et al. Data sharing statements for clinical trials — a requirement of the International Committee of Medical Journal Editors. N Engl J Med 2017;376(23):2277–9. doi:10.1371/journal.pmed.1002315

26. ICMJE. Defining the role of authors and contributors [cited 2022 Sept]. Available from: https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html

27. Taylor & Francis. Data sharing policies (2018) [cited 2022 Sept]. Available from: https://authorservices.taylorandfrancis.com/wp-content/uploads/2019/04/Author-Services-Data-sharing-policies.pdf

28. Springer Nature. Research data policy types [cited 2022 Sept]. Available from: https://www.springernature.com/gp/authors/research-data-policy/data-policy-types/12327096

29. Wiley Author Services. Wiley's Data Sharing Policies [cited 2022 Sept]. Available from: https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html

30. PLOS Biology. Data Availability (March 03, 2014). Available from: https://journals.plos.org/plosbiology/s/data-availability#loc-recommended-repositories

31. ClinicalTrials.gov [cited 2022 Sept] Available from: https://clinicaltrials.gov/

32. Vivli. The platform [cited 2022 Sept]. Available from: https://vivli.org/platform/

33. ClinicalStudyDataRequest.com (CSDR) [cited 2022 Sept] Available from: https://www.clinicalstudydatarequest.com/

34. The YODA project [cited 2022 Sept]. Available from: https://yoda.yale.edu/

35. EudraCT. European Union Drug Regulating Authorities Clinical Trials Database [cited 2022 Sept]. Available from: https://eudract.ema.europa.eu/

36. The YODA project. Policies & procedures to guide external investigator access to clinical trial data [cited 2022 Sept]. Available from: https://yoda.yale.edu/policies-procedures-guide-external-investigator-access-clinical-trial-data

37. Vivli. FAQs [cited 2022 Sept]. Available from: https://vivli.org/resources/faqs/

38. Pistoia Alliance. FAIR Toolkit. Interoperability maturity indicators [cited 2022 Sept]. Available from: https://fairtoolkit.pistoiaalliance.org/methods/interoperability-maturity-indicators/

39. Pharmaphorum. The largest pharma company in Europe calls for inter-operability. April 19, 2022 [cited 2022 Sept]. Available from: https://pharmaphorum.com/digital/the-largest-pharma-company-in-europe-calls-for-interoperability/

40. Spence O, Hong K, Onwuchekwa Uba R, Doshi P. Availability of study protocols for randomized trials published in high-impact medical journals: a cross-sectional analysis. Clin Trials 2020;17(1):99–105. doi:10.1177/1740774519868310

41. Data Citation Synthesis Group: Joint declaration of data citation principles. Martone M. (ed.). San Diego, CA: FORCE11: 2014

42. Open Pharma. Driving positive change in the communication of pharma-sponsored research [cited 2022 Sept]. Available from: https://www.openpharma.blog/

Author information

Eniola Awodiya is a fourth-year medical student at Green Templeton College, Oxford. She has a keen interest in medical writing and spent the summer of 2022 as an intern with Oxford PharmaGenesis, working with Open Pharma to drive transparency, accountability, accessibility, and discoverability in pharma-sponsored research communications.
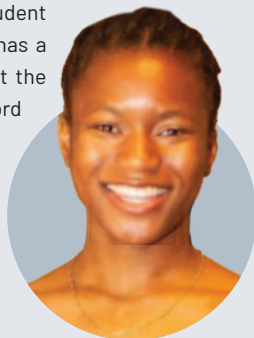
Joana Osório is a communications consultant at Oxford PharmaGenesis. She has worked in the communication of science and medicine for nearly 20 years, within academic research, publishing, the charity sector, and medical communications. She is project lead for Open Pharma, a multi-stakeholder collaboration that aims to drive positive change in the communication of pharma-sponsored research.