Guidelines for Raising FAIR Adoption in Health Data and Health-related Research Performing Organisations (HRPOs)

RDA Recommendation

DOI: 10.15497/RDA00078

Authors: Celia Alvarez-RomeroORCID, Anupama E GururajORCID, Kristan KangORCID, Shanmugasundaram VenkataramanORCID, Raising FAIRness in Health Data and Research Performing Organisations (HRPOs) WG members

Published: 5 May 2022

Version: 1.2

Abstract: Adoption and application of the FAIR principles to research data has grown significantly in the few years that they have existed. They have been embraced in many quarters and have paved the way to creating a level playing field for data reuse. Nevertheless, they also pose challenges in some areas of research where reuse may not be an automatic right due to issues of confidentiality, privacy, commercial interests and sensitivities in general.

One sector of research where this is particularly evident is health research, and this proposed working group aims to address global disparities in uptake of the FAIR principles in health research and within health research performing organisations (HRPOs). The work will build upon existing and ongoing work being conducted at the European level (please see below the attached document "D2.3 Guidelines for implementing FAIR open data policy in health research"), and which has in itself identified differences in policies within that region. The aim will be to expand on this work to create a global analysis of policies and to subsequently draw on commonalities to propose a set of guidelines (RDA outputs or recommendations) that can be utilised by HRPOs in their local contexts to address FAIRification of their research data.

Impact:

<u>Societal</u>

Siloed datasets cause barriers hindering full exploitation of health research data. Although many data have been made open, that does not necessarily translate to data that can be easily reused and shared, or even discovered, due to mismanagement or socioeconomic, political and/or legal barriers. Legal and ethical barriers for the implementation of FAIR policies in health research performing organisations (HRPOs) limit the reproducibility of research, while data variability and lack of access to large volumes of data make this more pronounced.

The FAIR principles will provide an answer to the challenges of proper research data management (RDM) of big data that will have significant impacts on society. The rapid response to the Covid crisis[1,2,3], show the benefits of sharing and making data FAIR, while several other diseases that could be targeted for concerted action using this model could also benefit. To provide indicators, the increase in secondary use of data could be measured once FAIR policies have been implemented, related to the publication and sharing of FAIR data, and compared to implementation of FAIR policies.

In terms of the UN's Sustainable Development Goals[4], and especially Good Health and Well-being, the FAIR principles provide more equity, especially for low- and middle-income (LMIC) countries through better data access and sharing. The imbalance between high income countries (HICs) and LMICs in levels of healthcare leads to disparities in health research and lack of funding and infrastructure. In today's digital world, LMICs will benefit from shared resources and this can be achieved through application of the FAIR principles.

Good scientific practice aided by the FAIR principles will also provide better understanding of research and allow better transparency and trust: for health research, this is particularly true to combat misinformation and distrust in health research solutions for many of the world's common diseases.

Economic

Several billion dollars of taxpayer funded research is conducted every year globally and this necessitates proper accountability. This data deluge will benefit from the FAIR principles and provides a framework in which accountability can be fulfilled to a large extent.

The European Commission's analysis of costs of not having FAIR data[5] concluded that: (i) this amounts to approximately €10.2bn lost per year for the EU; (ii) the open data economy suggests that the impact on innovation of FAIR could add another €16bn lost to the minimum cost estimated; and (iii) that would make a total of at least €26.2bn per year.

Unnecessary duplication of research could also be avoided through implementing the FAIR principles and the reproducibility of the research studies. Allowing reuse and sharing of data and increasing their discoverability will reduce the need to repeat previous research and thus reduce their economic outlay while also freeing up effort. These will perhaps have most significant impacts in LMICs where reuse of data could have solutions to local health problems.

Electronic health records (EHRs) provide a practical example of the effort and use of resources in a HRPO when using or not using the FAIR principles. Data extraction and collection from these and other healthcare sources is not trivial and requires extensive efforts, due mainly to: (i) raw data complexity (EHRs are typically very complex including information in several tables), (ii) free text, and (iii) differences between the nature of the source raw data. To address this complexity, different types of researchers are involved, including experts in Natural Language Processing (NLP) to address the information in free text fields of the EHRs and data scientists to analyse in-depth each source raw dataset. Therefore, to provide indicators, the increase in data reuse could be measured once FAIR policies have been implemented.

Technical and Scientific

Since open research is not the same as research adhering to the FAIR principles, sensitive data should not automatically be made available for reuse. However, metadata in health research data can be

---

[1] Digital Health COVID-19 Impact Assessment: Lessons Learned and Compelling Needs - National Academy of Medicine

[2] Preparedness needs research: How fundamental science and international collaboration accelerated the response to COVID-19 | PLOS Pathogens

[3] Open science saves lives: lessons from the COVID-19 pandemic | BMC Medical Research Methodology | Full Text

[4] UNESCO Recommendation on Open Science

[5] Cost-benefit analysis for FAIR research data - Publications Office of the EU

made open and reusers can seek permission to access the underlying data through existing mechanisms for access control and/or anonymisation/ pseudonymisation. Meanwhile, data from FAIR repositories will facilitate multi-centre and cross-domain studies which can be measured, increasing the scope of research in general, improving health research in particular, and facilitating further scientific advances. This could be done, for example, through surveys to evaluate the time involved for researchers, aiming to analyse the improvements of the FAIR policies in health research.

Rich metadata enables FAIR compliance and is also true for non-sensitive data: using appropriate licences, controlled and common vocabularies and domain specific repositories all add value to the data and uphold reproducibility and research integrity that are fundamental to good science. Implementing FAIR policies in HRPOs will allow the use of larger and more heterogeneous datasets, thus increasing the variability of the data and one indicator that could be measured is the size of the datasets used in specific research studies not applying FAIR compared to the size of a more heterogeneous and FAIR compliant dataset. FAIR dataset reuse from international repositories could also lead to more demographic, environmental and social information being taken into account, and variability of data and inclusion of more variables can be measured.

Language: English

RDA webpage: https://www.rd-alliance.org/groups/raising-fairness-health-data-and-health-research-performing-organisations-hrpos-wg

Citation and Download: Alvarez-Romero, C., Gururaj, A., Kang, K., & Venkataraman, S. (2022). *Health Research Performing Organisations (HRPOs) FAIR Guidelines* (Version 1.0). Research Data Alliance. https://doi.org/10.15497/RDA00078

Guidelines for raising FAIR adoption in health data and health-related research performing organisations (HRPOs)

PRINCIPLES

1. To implement a data policy, including FAIR, implies managing a complex change. The policy needs **strategic vision and leadership**.
2. There are no policies without **resources** and **incentives** supported by necessary infrastructure.
3. It is necessary to rely on the right **knowledge and skills** about FAIR data and research data management in the current Open Science landscape.
4. Before implementing any FAIR data policy, the institution needs a clear action plan identifying the main **actors** and a credible timeline.
5. The policy must be **written** and approved by the institution.

STEPS

1. Define the **vision and objectives** of the policy. This should include:
   a. A description of the current research performed in the institution: research funders and their policies regarding research data; types of research outcomes that the institution produced in the last 5 years; types of data used/reused by the researchers; current infrastructures used at institutional level (virtual research environments (VREs), data repositories, etc.).
   b. A definition of the objectives of the policy, such as: comply with funders' requirements; improve transparency and reproducibility; include datasets in health research infrastructures; increase return on investment into data acquisition by maximising use of data; minimise waste by improving visibility and accessibility of research already completed or underway; address health consumer/research participants' expectations that maximum value will be achieved by their contribution to research (e.g. improved health outcomes); enable data citation; facilitate research translation by making data more readily available for systematic review, etc.
   c. The adoption of a comprehensive approach to enabling mechanisms that can enhance data sharing and productivity while reducing risk.
2. Determine the agreed level of **data governance, openness, transparency and re-usability** for research data produced in the HRPO, including licensing and provenance as well as the intended mechanisms for personal data protection (e.g. GDPR compliance).
3. Identify/name a **responsible person/unit for the FAIR data policy** as well as the envisaged team to put it in practice. Design an action plan, state resources to be committed for Research Data Management and a feasible timeline.
   a. Define clearly the roles and responsibilities of the data creators, data owners and data processors in a FAIR context.

    b. Use existing frameworks to promote, establish and define roles and competencies wherever possible.
4. **Raise awareness** among researchers, as well as providing adequate training and assistance (data stewards, data scientists, or similar in the institution). Definition of needed skills and a training programme.
5. Identify and describe current and **required data infrastructures**. This includes data storage and architecture definition but also the provision of tools to make data description and formatting easy and affordable for researchers.
    a. Use of existing infrastructures is encouraged, but attention must be paid when using commercial solutions when dealing with sensitive data with respect to data ownership considerations.
    b. Gap analysis should be performed where possible to determine current and desired readiness levels.
6. Establish **responsible Research Data Management practices** within the institution and define FAIR for implementation. This should include:
    a. Analysis of the scope of the FAIR principles, that includes concepts like: data selection/curation, long-term stewardship, legal interoperability and the timeliness of sharing.
    b. Creation of a standard institutional template for Data Management Plans (DMPs).
7. FAIR data require several technical decisions and standards adoption: the policy should include decisions, at institutional level, around **technical standards**. For example:
    a. PID policy and control for identification of data, publications and other outputs, as well as researchers themselves.
    b. Metadata policy for data accessibility and interoperability: selection of metadata vocabularies, best practices on metadata completeness, etc. These decisions might include a technical analysis in domain relevant (health) standards including metadata schemas, data modelling and vocabularies.
    c. All research data should be deposited in a trustworthy repository.
    d. Anonymisation/pseudonymisation and access control mechanism decisions need to be clearly defined in line with the HRPO's own policies.
8. Devise **credit and reward mechanisms** in order to incentivise researchers and promote FAIR data management/sharing.
9. **Write the policy** on FAIR research data management:
    a. Submit it for approval by governance bodies.
    b. Disseminate the policy inside the HRPO.
    c. Monitor uptake and compliance through e.g. DMPs.
10. Ensure **FAIR data assessment** within the institution, using existing tools, or devising in-house metrics where necessary. Re-align and consolidate the policy with funder mandates to guarantee that publicly-funded research data are made FAIR (and open), except for legitimate restrictions.

Selected Resources

Competencies and skills	1. https://eosc-fair4s.github.io/ 2. https://www.digcomptest.eu/index.php 3. https://www.unesco.org/en/communication-information/digital-competencies-skills
Gap analysis and infrastructure readiness	4. ARDC Institutional Underpinnings Framework 5. https://sparceurope.org/evaluate-your-rdm-offering/ 6. https://ardc.edu.au/services/advisory-services/rise-workshop/
Data Management Plans (DMPs)	7. http://dmponline.dcc.ac.uk 8. http://argos.openaire.eu 9. https://dmptool.org/ 10. https://ds-wizard.org/
Metadata catalogues/ registries	11. https://bartoc.org/ 12. https://rdamsc.bath.ac.uk/
Example metadata	13. http://ecrin-mdr.online/index.php/Summary_Tables
Licences	14. https://chooser-beta.creativecommons.org/
Trustworthy repositories	15. https://www.re3data.org
Anonymisation/ pseudonymisation	16. https://amnesia.openaire.eu/
Access control	17. https://ukdataservice.ac.uk/learning-hub/research-data-management/data-protection/access-control/
FAIR assessment/ metrics	18. https://www.f-uji.net/ 19. https://ardc.edu.au/resources/aboutdata/fair-data/fair-self-assessment-tool/ 20. https://op.europa.eu/en/publication-detail/-/publication/ced147c9-53c0-11eb-b59f-01aa75ed71a1
Training and guidance	21. http://www.dcc.ac.uk 22. http://openplato.eu