BRIEFING

FAIR and Open Data

A BRIEFING FOR POLICYMAKERS AND SENIOR MANAGERS, DECEMBER 2018

SPARC Europe

1. Introduction

The past decade has seen major shifts in global policy on research data and expectations for this content to be shared. The OECD *Principles and Guidelines for Access to Research Data from Public Funding*,[1] have prompted many national research funders and governments to issue data sharing policies. Two central concepts within these policies are Open Data and more recently FAIR data.

Open data is defined as content that is freely accessible to use, modify and share by anyone for any purpose. The content must be made available under an open licence and must not pose restrictions on re-users beyond those that preserve provenance (e.g. attribution) or openness (e.g. share-alike licenses).[2]

FAIR, meanwhile, is a set of principles that describes the attributes data need to have to enable and enhance reuse by humans and machines.[3] FAIR stands for Findable, Accessible, Interoperable and Reusable, and is a set of 15 principles that emerged from a workshop in the life sciences in 2014, but is intended for broader application across all disciplines and object types.

Many people confuse and conflate the meanings of FAIR and Open Data, so this briefing demystifies some common misconceptions and advocates for both to be addressed in research data policy.

2. Common misconceptions

FAIR data has to be Open

Although some of the central concepts in FAIR and Open Data overlap, namely accessibility and interoperability, the two are not synonymous and data do not have to be made openly available to be FAIR. Data can be both Open and FAIR, one or neither. Mandates on Open Data can alienate some research communities as they face genuine ethical and commercial barriers to unrestricted data sharing. One of the strengths of FAIR is the flexibility researchers are afforded in terms of how they make their data available. Data could be shared online under an open licence, provided on request, or governed by conditions laid out in data sharing agreements. At minimum, the metadata should be available so users can discover that data exist, even if they can't get access directly.

[1] OECD (2007), Principles and Guidelines for Access to Research Data from Public Funding https://doi.org/10.1787/9789264034020-en-fr
[2] See the Open Definition: https://opendefinition.org
[3] Wilkinson, M. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, volume 3, Article number: 160018. https://doi.org/10.1038/sdata.2016.18

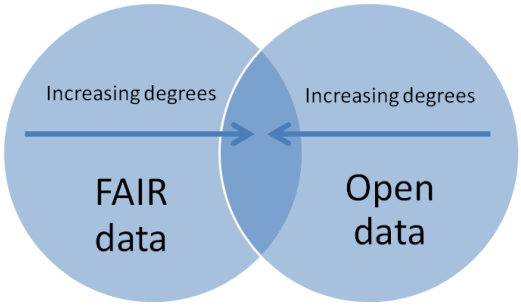

Image attribution: The relationship between FAIR and Open
CC-BY European Commission FAIR data Expert Group

FAIR is better than Open

Both concepts have important emphases and one cannot be regarded over the other. FAIR focuses on the reusability of data, but does not dictate any implementation approaches so researchers could adopt proprietary formats, share data under restrictions and provide minimal metadata while still being FAIR. It is also critical to recognise that both concepts are scales in which the reusability and usefulness of the object improves the greater the degree of openness or FAIRness. If data are made both FAIR and Open to the greatest extent possible, then the reuse potential and usefulness will increase.

FAIR is just for the life sciences

Although FAIR emerged from the life sciences and has a strong uptake within this community, the principles are articulated very broadly to apply to all types of content from all disciplines. Different research areas will have strengths and weakness in different areas. The social sciences for example have a strong history of national data centres and adoption of international standards like the Data Documentation Initiative (DDI), but machine-access and reuse of data is less prevalent. These differences in data practices and culture mean that each community should define both what it means to be FAIR in that context and appropriate metrics to measure uptake. In the FAIR Data Expert Group report, a call is made for disciplinary interoperability frameworks, giving the example of how the International Virtual Observatory Alliance (IVOA) had defined and maintained standards for the astronomy community that enables FAIR sharing.

3. A spectrum of FAIR and Open

As noted above, both FAIR and Open are scales. Tim Berners-Lee's 5 stars of Open Data articulates this very clearly. Data are technically Open if they are made available under an open licence, irrespective of the format and usefulness of the data. They may not be very reusable though, without being available as structured data, ideally in a non-proprietary format and with rich semantic links. Similarly with FAIR, it is the rich metadata, adoption of common protocols and interoperability that

enhance the potential reuse. Data may technically be FAIR without being all that reusable if only a few minimal principles are met.

★	make your stuff available on the Web (whatever format) under an open licence
★★	make it available as structured data (e.g. Excel instead of a scan of a table)
★★★	use non-proprietary formats (e.g. CSV instead of Excel)
★★★★	use URIs to denote things, so that people can point at your stuff
★★★★★	link your data to other data to provide context

Figure 2: 5 stars of Open data[4]

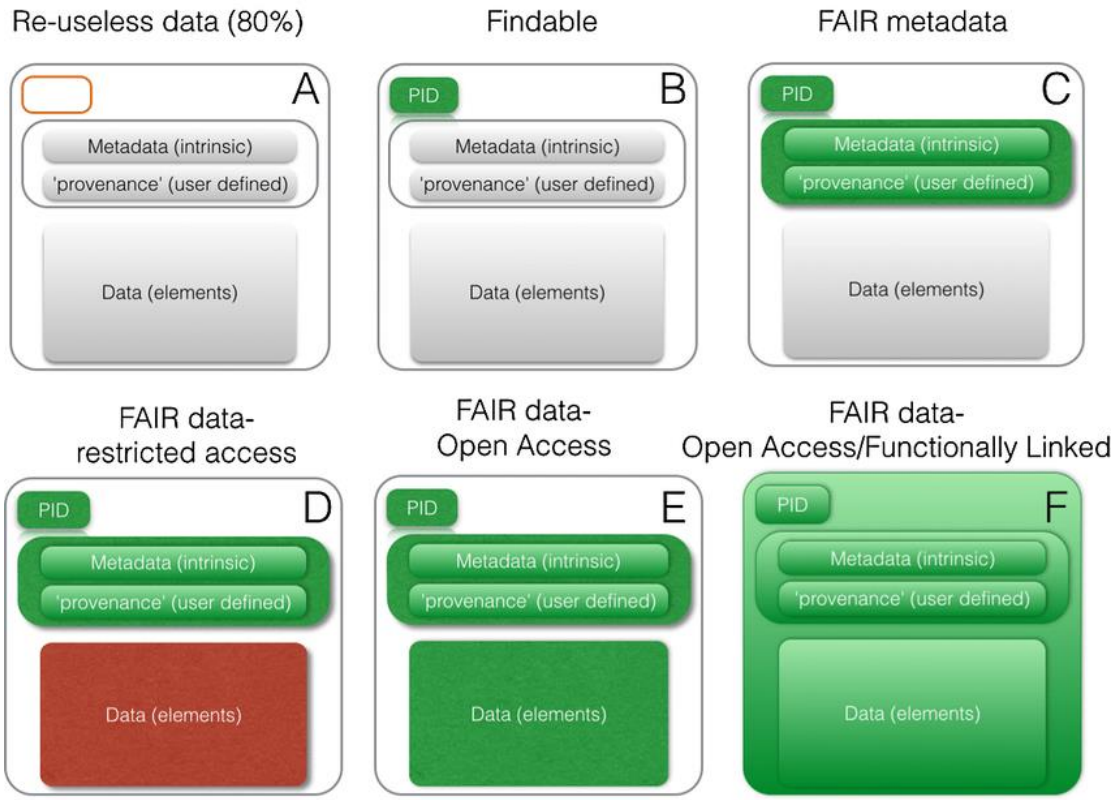

Figure 3: Data as increasingly FAIR Digital Objects[5]
From Cloudy, increasingly FAIR https://doi.org/10.3233/ISU-170824

There isn't a clear, universally-agreed minimum level of FAIR. But, if we take PIDs, metadata and access (to the data or at least metadata) as proposed in the EC Expert Group Report, data could be discovered, disambiguated and cited, but may not necessarily be in the most usable format or richly described to degrees that allow others to assess the usefulness of the data. Higher levels of FAIRness are needed to allow others to understand, recombine and reuse the data without any compatibility issues. Image F above indicates that not only the data and metadata are FAIR, but

[4] https://5stardata.info/en
[5] Notes on diagram: As elements become coloured, they have become FAIR. For example, adding a persistent identifier (PID) increases the fairness of that component. Coloured elements in green are FAIR and open, coloured elements in red are FAIR and closed. In the final panel, the mechanism for expressing the relationship between the ID, the metadata, and the data, is also FAIR (i.e. follows a widely accepted and machine-readable standard, such as DCAT or NanoPublications) and interlinked with other related FAIR data or analytical tools on the Internet of FAIR Data and Services.

the mechanism for expressing relationships between the ID, metadata and data is also FAIR. This highest degree of FAIRness parallels the linked open data approach topping the scale in the previous model.

4. How to implement FAIR and Open

In the European Commission Expert Group report on Turning FAIR Data into Reality, a triad of concepts are put forward for FAIR implementation.[6]

1. A model for FAIR Digital Objects
2. A FAIR ecosystem
3. Interoperability frameworks

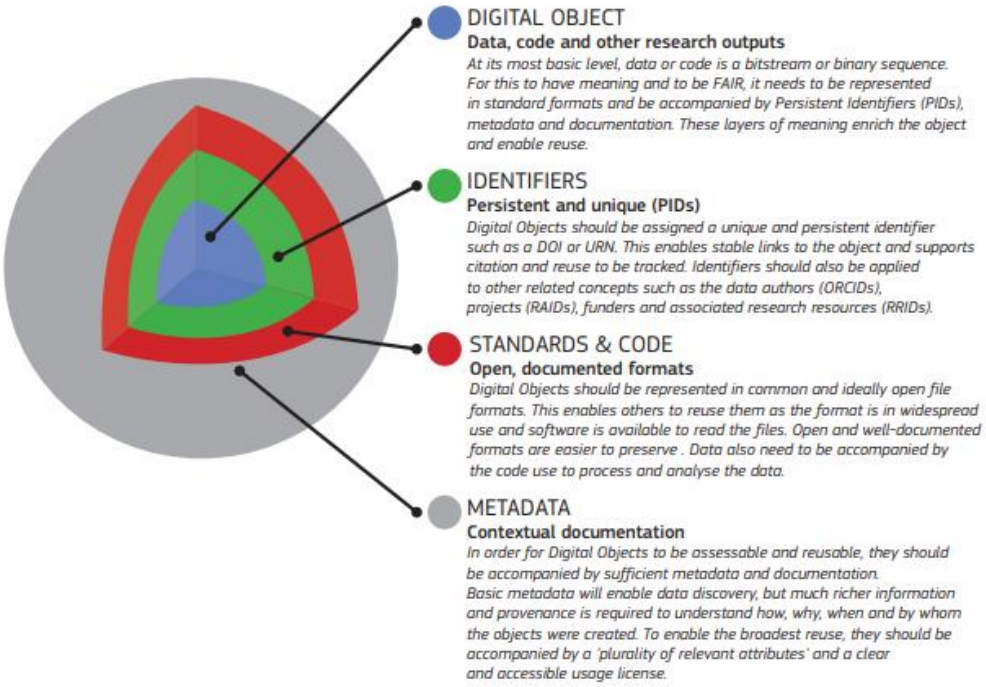

Figure 4: A model for FAIR Digital Objects
CC-BY European Commission FAIR data Expert Group

The model for FAIR Digital Objects visualises the key elements that need to be in place for data (or software and other research resources) to be FAIR. This includes Persistent Identifiers, rich metadata and documentation, and the adoption of standards. Interoperability frameworks will define these standards for a given research community.

It is important to support research communities to define their preferred data formats, metadata standards, expected timeframes for sharing, and other conventions such as deposit in a community repository. This will assist in building a common culture and practice of data sharing. In order to support the widest reuse and cross-disciplinary research, these interoperability frameworks should be articulated in common ways using global standards where possible, or, at a

[6] Hodson, S., Jones, S., et al. (2018). *Turning FAIR into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data.* http://doi.org/10.2777/1524

minimum, enable crosswalks and brokering between different disciplinary approaches.

The FAIR data Expert Group report also proposed a number of components needed in the ecosystem to enable Digital Objects to be created and shared in a FAIR way. These include:

- **Policies** to define and regulate the components of the FAIR ecosystem and their relationships.
- **Data Management Plans (DMPs)** to provide a hub of essential information on FAIR Digital Objects and the context of their creation.
- **Persistent Identifiers (PIDs)** assigned to FAIR Digital Objects and their component parts (e.g. data, metadata, code and algorithms, models, licenses), as well as other key actors such as researchers (ORCIDs), funders and institutions.
- **Standards** applied to metadata schema, to controlled vocabularies and ontologies, and to the information models of actionable DMPs and policies.
- **Trusted repositories** to store, curate and disseminate resources.

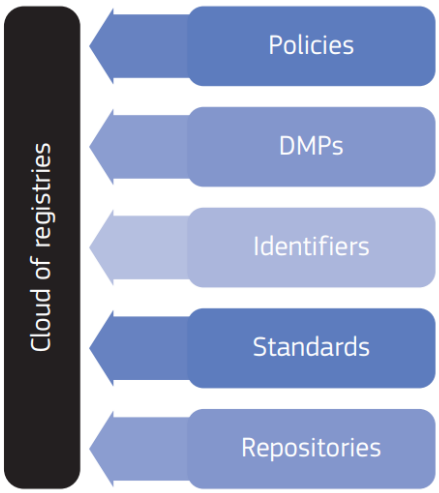

Figure 5: The components of a FAIR ecosystem
CC-BY European Commission FAIR data Expert Group

There should be registries cataloguing the component services in the FAIR ecosystem, and ideally automated workflows between them. For the ecosystem to be implemented and sustained there also needs to be a development of data science and stewardship skills, the implementation of appropriate metrics and incentives, and sufficient, well-targeted investment.

The Final Report and Action Plan released by the FAIR data Expert Group addresses a number of stakeholders including the European Commission, Member States and the international community. The primary context for implementation however is the European Open Science Cloud. As such, addressing FAIR in the broader context of Open Science and Member State policy and investments in this area is critical.

5. Recommendations

Having a **clear policy on data management and sharing** is critical. We recommend that policymakers address both concepts of FAIR and Open Data collectively. One may lend itself to your context more (e.g. if restrictions on sharing are necessary to protect personal data) but the two are complementary. In terms of Open Data, adopting the European Commission's mantra of 'as open as possible, as closed as necessary' is suggested.

Sharing expertise and lessons across institutions by **collaboration in international fora** is also key. Interoperability is inherent in both Open and FAIR and won't be achieved if institutions, countries and research disciplines take a siloed approach. The Research Data Alliance and other global coordination fora are proposed as mechanisms to align approaches and trial implementation internationally. CODATA, the Committee on Data for the International Science Council (ISC) could provide endorsement and support adoption via the Scientific Unions. It is critical that research data policy is implemented through such vehicles, with appropriate stakeholder engagement at each point.

All stakeholders should broaden their implementation approach by contributing to the definition and implementation of **interoperability frameworks**. Individual projects should aim to standardise and harmonise approaches at the research group level. Research groups and communities should work collaboratively to define disciplinary and cross-disciplinary interoperability frameworks that provide a solid platform for effective data sharing internationally. Institutional and national approaches should also compare with peers and seek common approaches. Indeed, the implementation of the European Open Science Cloud should align with global parallel initiatives such as the NIH data commons, the Australian Research Data Commons and the African Open Science Platform. By doing so we will enable data to be more FAIR, more Open and more reusable by all.

SPARC Europe would like to thank DCC for carrying out this work

www.dcc.ac.uk