



FAIRSFair
Fostering Fair Data Practices in Europe

Defining Data Interoperability Frameworks

Key issue #5 in Assessing Capability Maturity and Engagement with FAIR-enabling Practices (ACME-FAIR)

Josefine Nordling (CSC), Angus Whyte (DCC), Ricarda Braukmann, René van Horik (DANS)

Public Consultation Version 1 December 2021

DOI: 10.5281/zenodo.5775552

| | |
|-----------------------|--|
| Project Title | Fostering FAIR Data Practices in Europe |
| Project Acronym | FAIRSFair |
| Grant Agreement No | 831558 |
| Instrument | H2020-INFRAEOSC-2018-4 |
| Topic | INFRAEOSC-05-2018-2019 Support to the EOSC Governance |
| Start Date of Project | 1st March 2019 |
| Duration of Project | 36 months |
| Project Website | www.fairsfair.eu |

Defining Data Interoperability Frameworks

Introduction

Data interoperability is key to the FAIR principles, yet can be challenging to put into practice. This document provides guidance on practices involved in achieving data interoperability, more specifically, practices around data citation, persistent identifiers (PIDs), semantic resources, and metadata. All of these create a data interoperability framework and are important building blocks of a FAIR ecosystem. The purpose of such a framework is to set some specific requirements for the digital objects that it will be applied to. Generally these include that the digital object/data need to be accompanied by standardised metadata for it to be cited and be unambiguously identified, using a persistent identifier. The metadata should also describe the object according to a community-endorsed vocabulary, richly enough for it to be understandable and reusable by anyone in that community. In addition, the data files that comprise the object need to be represented in common and open formats.

The report 'Turning FAIR into Reality' (TFIR) issued by the European Commission in 2018 provides guidance for research practitioners and research performing organisations (RPOs) on FAIR practices resulting in data interoperability. TFIR summarises the relevant recommendations on data interoperability in: *Recommendation 4 - Develop interoperability frameworks*. TFIR recommends implementing mechanisms to facilitate sharing good practices and lessons learned on implementing FAIR practices within and across disciplines. TFIR also states that these frameworks are supported through common standards, which need to be developed through international collaborations. Advancing semantic technologies and common standards collaboratively will ultimately support interdisciplinary research and break down silos between communities. To enable research communities to effectively advance disciplinary frameworks, there needs to be adequate funding in place for doing so.

We can also find related recommendations made in the FAIRsFAIR project¹ described how data stewards can play an important role in making sure the semantic resources are kept up-to-date and in providing the researchers with practical support in finding the relevant standards to be used. Data stewards should collaborate with research communities and software engineers in developing and maintaining machine-readable FAIR semantic resources and domain standards. In addition, the recommendations include good data citation practices. Researchers are highly advised to make use of ORCID to unambiguously identify them with their research outputs and help to make these findable. Research communities have a responsibility to agree on PID policies and practices to be used, and to maintain documentation of these. Institutions have a role to play in guiding researchers

¹ Molloy, Laura, Nordling, Josefine, Grootveld, Marjan, van Horik, René, Whyte, Angus, Davidson, Joy, Herterich, Patricia, Martin, Ivan, Méndez, Eva, Principe, Pedro, Vieira, André, & Asmi, Ari. (2020). D3.4 Recommendations on practice to support FAIR data principles (1.1 DRAFT). Zenodo. <https://doi.org/10.5281/zenodo.3924132>

on the PID policy set for their specific scientific domain and lastly, data stewards can advocate and support appropriate use of PIDs for different types of research objects.

Introducing ACME-FAIR

The document sets out a draft FAIRsFAIR framework, whose main purpose is to help those managing and delivering relevant professional services to self-assess how they are enabling researchers, and colleagues who support them, to put the FAIR principles into practice. We refer to this as ‘FAIR-enabling practice’. We welcome your comments on this draft, and responses to the specific consultation questions you can find below at the end of this introduction.

ACME-FAIR can be used independently, or it can be used to complement Science Europe’s *Practical Guide to Sustainable Research Data*.² Both guides include ‘capability maturity’ matrices (or ‘rubrics’), for Research Performing Organisations (RPOs) e.g. universities, research institutes. While Science Europe’s guide is aimed at strategic-level management of the organisation, **ACME-FAIR targets the operational levels of the organisation**. It can optionally be used to follow up an assessment based on the Science Europe maturity matrices. ACME-FAIR is also strongly informed by *Turning FAIR into Reality*³ (henceforth TFIR), the recommendations of the European Commission’s Expert Group on FAIR data.

Covering key practical issues

ACME-FAIR covers 7 key issues. These address the FAIR-enabling practice themes highlighted in a number of FAIRsFAIR deliverables, together with recommendations from the *Turning FAIR into Reality* report. The table below shows the corresponding areas covered by the Science Europe *Guide to Sustainable Research Data*.

| | |
|---|--|
| <ol style="list-style-type: none"> 1. Defining the policy environment 2. Developing sustainable business models 3. Professionalising roles through training, mentoring, and recognition 4. Supporting data management planning 5. Defining data interoperability frameworks 6. Selecting data, services, and repositories for FAIR 7. Ensuring trusted curation | <ul style="list-style-type: none"> - Policy environment - Financial aspects - Training <p style="font-size: 2em;">}</p> <p>Technical preparedness</p> |
|---|--|

Table 1. Mapping key issues addressed in ACME-FAIR (left) to Science Europe’s guidance (right)

Why use ACME-FAIR?

ACME-FAIR aims to be useful for services providing support to researchers on FAIR implementation in Research Performing Organisations (RPOs). It has 3 main use cases:

1. For the service to self-assess its readiness to support FAIR, by establishing current and desired levels of engagement with research community practices, and the organisational maturity of the support offered for FAIR data.
2. To aid colleagues’ in identifying areas of improvement in an organisation’s support for FAIR data management.

² Tommaso Boccali, Anne Elisabeth Sølvsnes, Mark Thorley, Stefan Winkler-Nees, & Marie Timmermann. (2021). Practical Guide to Sustainable Research Data. <https://doi.org/10.5281/zenodo.4769703>

³ Collins, S., Genova, F., Harrower, N., Hodson, S., Jones, S., Laaksonen, L., ... & Wittenburg, P. (2018). Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data.

3. For national or international coordination initiatives to facilitate sharing of consistent information between peer organisations about their current levels of maturity, and to encourage community engagement around FAIR-enabling practices.

The ultimate aim of ACME-FAIR is to improve the availability of information on the implementation of support for FAIR data across disciplines and communities of practice. ACME-FAIR is partly based on the Digital Curation Centre's *RISE* self-evaluation framework for research data service development⁴ and partly on the guide '*Do I-PASS for FAIR*', which was produced in the context of the Dutch Coordination Point Research Data Management.⁵

How ACME-FAIR is structured

ACME FAIR uses a scale comprising, for each of the 7 issues, the following dimensions: -

- 3 levels of **maturity**
- 3 levels of **community engagement**

The maturity levels are a simplified version of the first 3 levels of the widely adopted *CMMI* (Capability Maturity Model Integration) which has been widely adopted as a tool to guide process improvement, especially in software development contexts.⁶

In ACME-FAIR the levels of community engagement are separated out from maturity for the following reasons:

- Community engagement is essential for all of the practice areas covered;
- While the maturity goal of optimising alignment with *organisational* standards and practice is relevant to Research Performing Organisations, for research data support it is equally important to align with *community* standards, as defined by research domains and professional communities of practice;
- Identifying areas where maturity and engagement are at differing levels may be helpful to identify pockets of good practice in one or the other dimension, or areas to target for further action in your organisation.

The maturity and community engagement dimensions both indicate progression from ad-hoc project-level coverage of practice areas, through to organisation-wide coverage. These levels are:

Maturity

1. **Initial.** May be incomplete and falling short of the intent of the area of focus. Aware of and addressing performance issues.
2. **Managed.** Coverage delivering the full intent of the area of focus, minimally in some aspects, or lacking full alignment with overall organisational standards and practice. The approach identifies and monitors performance objectives. Includes and builds on level 1.

⁴ Rans, J and Whyte, A. (2017). 'Using RISE, the Research Infrastructure Self-Evaluation Framework' v.1.1 Edinburgh: Digital Curation Centre: www.dcc.ac.uk/guidance/how-guides

⁵ Taco de Bruin, Sarah Coombs, Jutta de Jong, Irene Haslinger, Henk van den Hoogen, Frans Huigen, Mijke Jetten, Jacko Koster, Margriet Miedema, Sief Öllers, Inge Slouwerhof, Ingeborg Verheul, & Jacqueline Ringersma. (2020). Do I-PASS for FAIR. A self assessment tool to measure the FAIR-ness of an organization (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.4080867>

⁶ See e.g. 'Capability Maturity Model Integration' Wikipedia article (accessed 24.11.2021) https://en.wikipedia.org/wiki/Capability_Maturity_Model_Integration

3. **Defined.** Complete coverage that delivers the full intent of the area of focus and aligns with overall organisational standards and practice. Identifies and monitors performance objectives that expand alignment to the whole organisation. Includes and builds on level 2.

Community engagement: practice awareness, adoption, and collaboration

This dimension identifies the level of engagement the organisation (or the relevant services it offers) has with the communities it serves, about maintaining and updating data stewardship practices and identifying new areas for the development of policy and implementation standards. It includes actively communicating and promoting existing and emerging approaches to the immediately impacted communities and the wider data infrastructure landscape.

1. **Awareness:** the service monitors data stewardship practice in the community or communities it serves, and makes local practitioners aware of it.
2. **Adoption:** the service or its host organisation also supports practitioners to embed community practice locally.
3. **Collaboration:** the service also engages with the design, development, and review of community practice. Consults and collaborates widely, potentially also taking a community coordination and leadership role.

ACME covers the issues listed in Table 1, each with a two-dimensional rubric (maturity x community engagement).

Consultation questions

Please use [this form](#) to give your feedback. It asks how far you agree with 4 simple statements, and invites you to add any comments you wish. Please note that the form collects no personal information.

You are also welcome to add comments directly to [this google doc](#) (these may identify you by your Google ID). If you prefer, please email the FAIRsFAIR task lead Dr Angus Whyte (a.whyte@ed.ac.uk) or the Project Coordination Office (pco@fairsfair.eu).

ACME Checklist

The ACME-FAIR checklist identifies five main capability areas under this theme. Four capability areas are assessed on the *maturity* scale, measuring integration of the capability with organisation-level standards and practices. The fifth capability area is assessed on the *community engagement* scale, measuring adoption of broader community standards and practices.

The Science Europe *Practical Guide to Sustainable Research Data* includes a capability maturity matrix that complements ACME-FAIR at a high level. The relevant capabilities it describes include:

- Policy environment: articulating the principles and practices on RDM established by the RPO and to be followed by its researchers, together with the necessary support to its researchers.
- Organisational engagement and commitment: acknowledging the need to develop solutions for sustainable research data and being committed to seek alignment of approaches with other research stakeholders (such as other RPOs, funders, infrastructures, research communities).

The scales used in the Science Europe guide are broadly consistent with ACME-FAIR. It may be helpful to use it prior to using ACME-FAIR, but this is not necessary to use ACME-FAIR effectively.

As a first step, consider the capabilities in the checklist below that are relevant to your organisation. This may help you narrow down your goals in using ACME-FAIR, which might include assessing only those capabilities already under development, only those under consideration, or both.

Which capabilities is your organisation developing or considering doing in future?

| Maturity | Current | Considering |
|---|--------------------------|--------------------------|
| 1) Supporting data citation standards? | <input type="checkbox"/> | <input type="checkbox"/> |
| 2) Establishing persistent identifiers (PIDs) for research objects, related people and organisations? | <input type="checkbox"/> | <input type="checkbox"/> |
| 3) Applying standards for metadata and semantic resources to organisational systems? | <input type="checkbox"/> | <input type="checkbox"/> |
| 4) Enabling data interoperability in research workflows for core services and facilities? | <input type="checkbox"/> | <input type="checkbox"/> |
| Engagement | | |
| 5) Supporting community-relevant standards for interoperability? | <input type="checkbox"/> | <input type="checkbox"/> |

These capabilities might be developed by a single unit within a Research Performing Organisation, for example by a Library or Research Office. More likely, several areas of the organisation's governance will also be involved, e.g. Research Committee, Research Ethics Committee, Intellectual Property and Commercialisation Unit, and any Research Data Management service.

The next step in using ACME-FAIR is to discuss with the relevant colleagues what can realistically be achieved to meet needs of researchers, other stakeholders such as funders, and the organisation. To inform that, you may find the scope notes below helpful. They describe each capability for this theme covered in the framework.

Scope

We define capabilities as follows below, and then describe levels of maturity and engagement.

Supporting data citation standards

- Making staff, and relevant professional services to support research, aware of the need for data to be citable, and of the relevant metadata to be recorded.
- Services to make users aware of how to cite research data, and to collect and share citation metadata.

- Integration of data citation metadata into library and research information systems, to track relationships between research outputs and how these are cited.

Establishing persistent identifiers (PIDs) for research objects, people, and organisations

- Engaging in relevant forums and infrastructures for assigning and tracking persistent identifiers, to gain the knowledge needed to provide support.
- Defining a PID policy that sets out the kinds of objects that require a persistent identifier, and the PID services the organisation will make available to use for each object.
- Making relevant PID services easily available and monitoring their use according to the organisation's PID policy.

Applying standards for metadata and semantic resources to organisational systems

- Advocacy and training about the value of metadata standards and semantic resources, including recommendations about which standards are relevant to local research practice.
- Applying community-endorsed metadata standards and semantic resources in repositories and research information systems.
- Supporting standards and practices defined by relevant research communities, to ensure that metadata employ community-endorsed semantic terms, are machine-actionable, and interoperable with national and international infrastructures.

Enabling data interoperability in research workflows for core services and facilities

- Learning about relevant vocabularies and minimum metadata models, and their application through the infrastructures available to the organisation and researchers.
- Agreeing on standard vocabularies, a minimum metadata model, and crosswalks to ease discovery of research output, ensuring semantic interoperability by referring to external resources.
- Using clear and precise, publicly-available definitions for all concepts, metadata and data schemas, allowing for disciplinary variation in these, and facilitating the documentation of data provenance.

Supporting community-relevant standards for interoperability

- Ensuring a shared awareness of standards and relevant infrastructures, organisations and groups developing these, and advocating for community-endorsed approaches.
- Encouraging adoption of new standards proposed by relevant community fora, and actively engaging with research groups on using standards that are endorsed by relevant communities.
- Enabling staff to participate in activities to develop interoperability standards with the infrastructures, organisations, and community groups involved in developing these.

Defining Data Interoperability Frameworks - ACME Rubric

| Defining Data Interoperability | Maturity | | | Maturity level (1-3) |
|--|---|---|---|----------------------|
| | 1) Initial | 2) Managed | 3) Defined | |
| | May be incomplete and falling short of the intent of the area of focus. Aware of and addressing performance issues | Delivering the full intent of the area of focus, though minimally in some aspects. Lacking full alignment with overall organisational standards and practice, but identifies and monitors performance objectives. Includes and builds on level 1. | Complete coverage that delivers the full intent of the area of focus and aligns with overall organisational standards and practice. Identifies and monitors performance objectives that expand alignment to the whole organisation. Includes and builds on level 2. | |
| Supporting data citation standards | Relevant professional services to support research in our organisation are aware that data needs to be citable. We are learning about citation standards, and the infrastructures that make data citation possible. We inform staff about data citation, and about recording the relevant metadata. | Our services make users aware of how to cite research data in standard citation formats. Relevant library and research information systems share metadata to enable citation and connect with national or international infrastructures for sharing data citations. | Our services fully integrate research dataset citation metadata into library and research information systems. These give feedback to data producers and other stakeholders to help track the relationships between research data and other outputs, and whether these are cited. | |
| Establishing persistent identifiers (PIDs) for research objects, people and organisations | We get engaged in relevant forums and infrastructures for assigning and tracking persistent identifiers, and get the knowledge needed for providing support. | We define a PID Policy to guide researchers on optimal PID usage. The PID policy contains an overview of the objects that require a persistent identifier and the PID services used for this. | We make relevant PID services easily available and monitor that these are used according to our organisation's PID policy. We make the services that are formulated in the PID policy easily available. | |

| | 1) Initial May be incomplete and falling short of the intent of the area of focus. Aware of and addressing performance issues | 2) Managed Delivering the full intent of the area of focus, though minimally in some aspects. Lacking full alignment with overall organisational standards and practice, but identifies and monitors performance objectives. Includes and builds on level 1. | 3) Defined Complete coverage that delivers the full intent of the area of focus and aligns with overall organisational standards and practice. Identifies and monitors performance objectives that expand alignment to the whole organisation. Includes and builds on level 2. | Maturity level (1-3) |
|---|---|--|---|-----------------------------|
| Applying standards for metadata and semantic resources* to organizational systems (*a semantic resource defines logical or other relations between terms, using a standardised format e.g. taxonomy, thesaurus, ontology) | We advocate the value of metadata standards and semantic resources in advancing interoperability and standardisation, and take active part in training. We recommend standardised data formats and vocabularies to be used, and inform staff about metadata relevant to their practice. | We apply community-endorsed metadata standards and semantic resources in our repositories and research information systems. These are mainly cross-domain standards, with limited support for some domain-specific metadata. | Our repositories and research information systems are supporting domain metadata standards and practices defined by relevant research communities. These systems apply linked open data, to ensure the metadata are using community-endorsed semantic terms, are machine-actionable, and interoperable with national and international infrastructures. | |
| Enabling data interoperability in research workflows for core services and facilities | We are learning about relevant vocabularies and minimum metadata models and how to apply these in practice through the infrastructures that are available to our organisation and researchers. | Our services have agreed on a set of standard vocabularies supporting a minimum metadata model, and offer crosswalks to ease discovery. We ensure semantic interoperability of datasets by referring to external resources, such as Data Type Registry, recommended data schemas, concepts and metadata standards. | We use clear and precise, publicly-available definitions for all concepts, metadata and data schemas used in our services. These can be extended to allow for disciplinary variations. We facilitate documentation of data provenance in research workflows and processes. | |

| | Community engagement: Practice awareness, adoption and collaboration | | | |
|---|---|--|--|--------------------------------|
| | 1) Awareness: the organisation monitors community practice and makes local practitioners aware of it. | 2) Adoption: the organisation also supports practitioners to embed community practice locally. Includes and builds on level 1. | 3) Collaboration: the organisation also engages with the design, development, and review of community practice. Consults and collaborates widely, potentially also taking a community coordination and leadership role. Includes and builds on level 2. | Engage-ment level (1-3) |
| Supporting community-relevant standards for interoperability | We ensure there is shared awareness of standards for interoperability and of the relevant infrastructures, organisations and groups involved in their development. We advocate for community- endorsed PIDs, citation approaches, and for using domain-relevant metadata and semantics for key terms. | We encourage research-active staff to adopt standards suggested by the relevant community fora. We actively engage with research groups on using community-endorsed interoperability standards both for their specific domain and for cross-domain purposes. | We enable staff to participate in activities to develop interoperability standards with the infrastructures, organisations, and community groups involved in developing these for specific domains and cross-domain purposes. | |

