



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)

Version 2.1 March 2022

DOI: 10.5281/zenodo.6346797

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

ACME-FAIR is a set of guides produced in the FAIRsFAIR project, 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 (for short we refer to this as ‘FAIR-enabling practice’). 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 targets their strategic-level management, **ACME-FAIR aims to support 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 the recommendations of the European Commission’s Expert Group on FAIR data, *Turning FAIR into Reality*.³

Covering key practical issues

ACME-FAIR covers 7 key issues for FAIR-enabling practice themes highlighted by FAIRsFAIR, in response to recommendations from the *Turning FAIR into Reality* report, and issues covered by the Science Europe *Guide to Sustainable Research Data*. The table below shows how the FAIRsFAIR and Science Europe guides complement each other.

<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="text-align: center;">} Technical preparedness</p>
--	---

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

The ACM-FAIR guides are a series, with one guide for each of the issues in Table 1. Each includes a brief introduction, together with the explanation above, followed by a checklist describing the scope of the capabilities covered. Each guide then offers a rubric or set of tables describing maturity and community engagement dimensions of these capabilities.

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 European Commission, Directorate-General for Research and Innovation, (2018). *Turning FAIR into reality : final report and action plan from the European Commission expert group on FAIR data*, Publications Office. <https://data.europa.eu/doi/10.2777/54599> (p.57)

Why use ACME-FAIR?

The ACME-FAIR aims to be useful to services providing researchers with support on FAIR implementation. Its fundamental role is to offer a framework for discussion within and between organisations. 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 communication and adoption of community practices and the organisational maturity of the support offered for these.
2. Provide a basis for dialogue with colleagues to set out a roadmap for improving on current support, e.g. through training and skills development to improve the communication and adoption of community practices.
3. Support sharing of consistent information between peer organisations about their current levels of maturity and community engagement around FAIR-enabling practices, e.g. with national or international coordination and facilitation.

Organisations that perform research vary a great deal, both in how they are organised internally, and the environments they operate in. No capability model can take all of these factors into account, so anyone involved in planning a roadmap for their organisation's services in this area is likely to want or need more specific guidance on the topics covered. The ACME-FAIR guides will be developed further to reference some of these. FAIRsFAIR also offers a set of examples in the form of 'Implementation Stories' that cover the same themes.⁴

Background

ACME FAIR is partly based on the Digital Curation Centre's *RISE* self-evaluation framework for research data service development⁵, and partly on the guide '*Do I-PASS for FAIR*', which was produced in the context of the Dutch Coordination Point Research Data Management.⁶

ACME FAIR uses a two-dimensional scale, comprising 0-3 maturity levels for each of the 7 issues, and 0-3 levels of communication and adoption of practice. The **maturity levels** are a simplified version of the first 3 levels of the widely adopted *CMMI* (Capability Maturity Model Integration) framework⁷.

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, or areas to target for further action.

⁴ <https://fairsfair.eu/implementation-adoption-stories>

⁵ 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

⁶ Taco de Bruin, Sarah Coombs, Jutta de Jong, Irene Haslinger, Henk van den Hoogen, Frans Huigen, Mijke Jetten, Jacko Koster, Margriet Miedema, Sjeff 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>

⁷ CCMI. e.g. https://en.wikipedia.org/wiki/Capability_Maturity_Model_Integration

Capability dimensions: maturity and community engagement

The maturity and community engagement dimensions both indicate progression from no activity (level 0), through ad-hoc coverage of some practice areas (e.g. varying widely across research projects), through to more standardised approaches across the organisation. The maturity and community engagement dimensions are described in more detail as follows:

Maturity

0. **Not addressed.** The relevant professional services for research support do not coordinate any support capability for researchers in this area of focus. Some staff may help but it is not a formally recognised part of their job.
1. **Initial.** May be incomplete and falling short of the intent of the area of focus. Aware of and addressing performance issues.
2. **Managed.** Complete coverage delivering the full intent of the area of focus, 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.

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.

0. **Not addressed.** The relevant professional services for research support do not coordinate any support capability for researchers in this area of focus. Some staff may help but it is not a formally recognised part of their job.
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.

Please give us your feedback

The Digital Curation Centre (DCC) maintains ACME-FAIR. Feedback on this guide was gathered in the FAIRsFAIR project, and changes have been made to reflect that. DCC very much welcomes your thoughts on how to improve it further, especially suggestions of guidance to reference on each of its themes. Please give your feedback using this [short questionnaire](#). It asks how far you agree with 4 simple statements, and invites you to add any comments you wish. Please note that it collects no personal information.

ACME Checklist: Defining Data Interoperability Frameworks

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.

ACME Rubric: Defining Data Interoperability Frameworks

Defining Data Interoperability	Maturity			
	1) Initial	2) Managed	3) Defined	Maturity level (0-3)
	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 organisation's 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 how these are cited.	
Establishing persistent identifiers (PIDs) for research objects, people and organisations	We offer guidance on the relevant forums and infrastructures for assigning persistent identifiers, and acquire the knowledge needed for providing services to integrate these at organisational level.	We have defined a PID Policy to guide researchers on optimal PID usage. The PID policy contains an overview of the types of objects that require a persistent identifier and the PID services that are available to be used for this purpose.	Our organisation has established PID services to identify all relevant research staff and outputs, integrating these services with local research information systems. We monitor that these are used according to the PID policy.	

	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 (0-3)
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 organisation's 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 relevant infrastructures that are available to our organisation and its researchers.	Our services have agreed on standard vocabularies, a minimum metadata model and crosswalks to ease discovery of the organisation's research output. We ensure semantic interoperability of datasets by referring to external resources, such as Data Type Registry, recommended data schemas, concepts and metadata standards.	Our organisation has 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 (0-3)
Supporting community-relevant standards for interoperability	We ensure shared awareness of standards for interoperability and 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 new standards proposed by the relevant community fora. We actively engage with research groups on using standards that are endorsed by relevant communities, both for their specific domain and for cross-domain purposes.	Our organisation enables 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.	

