



**FAIRSFair**  
Fostering Fair Data Practices in Europe

## Ensuring trustworthy curation

### Key issue #7 in Assessing Capability Maturity and Engagement with FAIR-enabling Practices (ACME-FAIR)

Marjan Grootveld, Ricarda Braukmann, René van Horik, Maaïke Verburg (DANS), Angus Whyte (DCC)

Version 2 March 2022

DOI: 10.5281/zenodo.6345294

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	<a href="http://www.fairsfair.eu">www.fairsfair.eu</a>

# Ensuring trustworthy curation

## Introduction

Ensuring sustainable access to the data collected and produced in research processes is a critical concern for governments and research funding bodies in Europe and internationally. Research Performing Organisations (RPOs) such as universities and research institutes are key players in this endeavour. This requires data to be produced and managed according to the FAIR data stewardship principles, to be *Findable, Accessible, Interoperable, and Reusable*. Curating the data involves keeping it FAIR, and this requires services capable of applying the TRUST principles.<sup>1</sup> These involve providing *Transparency* about data holdings, taking *Responsibility* for the data integrity, maintaining *User focus* to serve communities, ensuring *Sustainability* of services to preserve data, and utilizing *Technology* to fulfil these principles. To make this happen, RPOs can partner with Trustworthy Repositories to achieve a level of technical preparedness that will ensure long-term accessibility to publicly-funded data holdings. This guide aims to offer RPO staff help to identify an appropriate level of preparedness for their circumstances. It complements FAIRsFAIR guidance to repositories on certification standards for trustworthy services.<sup>2</sup>

The 2018 European Commission ‘Turning FAIR into Reality’ report and action plan (TFIR) is a key reference for, amongst other stakeholders, the communities of researchers and professional staff in RPOs who are looking for guidance in this area. The report makes recommendations relevant to RPOs on sharing via repositories, including the following.

- Recommendation 17. Align and harmonise FAIR and Open data policy: Concrete and accessible guidance should be provided to researchers to find the optimal balance between sharing whilst also safeguarding privacy.
- Recommendation 20. Deposit in Trustworthy Digital Repositories: Concrete steps need to be taken to ensure the development of domain repositories and data services for interdisciplinary research communities so the needs of all researchers are covered.

These messages from TFIR are amplified in FAIRsFAIR *D3.4 Recommendations on practice to support FAIR data principles*.<sup>3</sup> These include:

- Researchers should be supported in the deposit of data in Trustworthy Digital Repositories, e.g. by data stewards.
- Researchers and data stewards should work within GDPR and IPR regulations to make data FAIR and “as open as possible, as closed as necessary”.

The emphasis in the current guide is on providing the environment for trustworthy curation, rather than on dealing with specific details such as GDPR. It covers taking responsibility for curation, in response to a

---

<sup>1</sup> Lin, D., Crabtree, J., Dillo, I. *et al.* The TRUST Principles for digital repositories. *Sci Data* **7**, 144 (2020). <https://doi.org/10.1038/s41597-020-0486-7>

<sup>2</sup> Hervé L'Hours, Ilona von Stein, Jerry deVries, Linas Cepinskas, Joy Davidson, Patricia Herterich, Robert Huber, & Benjamin Jacob Mathers. (2021). M4.3 CoreTrustSeal+FAIRenabling, Capability and Maturity (1.0). Zenodo. <https://doi.org/10.5281/zenodo.5346822>

<sup>3</sup> 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>

high-level policy commitment, and the engagement with trustworthy repositories that may be needed to help fulfil the commitment.

Support for these recommendations is also available in related ACME-FAIR guides to be available in the ACME-FAIR [collection](#) - *Defining the Policy Environment*, and *Selecting data, services, and repositories for FAIR*.

## 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*.<sup>4</sup> 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*.<sup>5</sup>

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

**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.

<sup>4</sup> 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>

<sup>5</sup> 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.<sup>6</sup>

## Background

ACME FAIR is partly based on the Digital Curation Centre's *RISE* self-evaluation framework for research data service development<sup>7</sup>, and partly on the guide '*Do I-PASS for FAIR*', which was produced in the context of the Dutch Coordination Point Research Data Management.<sup>8</sup>

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) framework<sup>9</sup>.

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.

## 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

---

<sup>6</sup> <https://fairsfair.eu/implementation-adoption-stories>

<sup>7</sup> 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](http://www.dcc.ac.uk/guidance/how-guides)

<sup>8</sup> 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>

<sup>9</sup> CCMI. e.g. [https://en.wikipedia.org/wiki/Capability\\_Maturity\\_Model\\_Integration](https://en.wikipedia.org/wiki/Capability_Maturity_Model_Integration)

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: Ensuring trustworthy curation**

The ACME-FAIR checklist identifies six 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. Another two capability areas are 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) Monitoring the scale of data production?	<input type="checkbox"/>	<input type="checkbox"/>
2) Ensuring all retained datasets are consistently assigned Persistent Identifiers (PIDs)?	<input type="checkbox"/>	<input type="checkbox"/>
3) Being responsible for data curation?	<input type="checkbox"/>	<input type="checkbox"/>
4) Providing (access to) a data repository?	<input type="checkbox"/>	<input type="checkbox"/>
Engagement		
5) Engaging with domain standards?	<input type="checkbox"/>	<input type="checkbox"/>
6) Engaging with the trustworthy repository community?	<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.

**Scope**

We use the following definition of data curation:

*“The activity of managing and promoting the use of data from their point of creation to ensure that they are fit for contemporary purpose and available for discovery and reuse. For dynamic datasets this may mean continuous enrichment or updating to keep them fit for purpose. Higher levels of curation*

*will also involve links with annotation and with other published materials.”*  
<https://casrai.org/term/curation/>

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

### Monitoring the scale of data production

- Having some knowledge of the amount of datasets produced.
- Developing a process to monitor dataset production in the organisation.
- Monitoring the amount of datasets produced and the amount that is deposited in a repository.

### Ensuring all retained datasets are consistently assigned Persistent Identifiers (PIDs)

- Promoting that data are deposited in a repository that assigns PIDs.
- Providing and tracking PIDs.
- Ensuring through policy and documented process that all datasets are deposited in a repository that assigns PIDs.

### Being responsible for data curation

- Providing guidance about the minimal contextual information that researchers should provide for their data.
- Curating research data on researchers' request.
- Carrying out data curation based on a documented process.

### Providing (access to) a data repository

- Running a repository (or referring to an external repository) that ensures minimal data and metadata preservation.
- Running a repository (or referring to an external repository) that ensures preservation including file migration and logging of all actions the repository performs on the data.
- Running a repository (or referring to an external repository) that commits to maintaining the significant properties of the data, for required retention periods and identified user groups.

### Engaging with domain standards

- Promoting awareness of research data and metadata standards that are relevant to specific domains.
- Providing guidance on such standards through various research support activities.
- Engaging in the development of such standards, and supporting research groups to engage in developing and maintaining domain-relevant standards.

### Engaging with the trustworthy repository community

- Informing researchers about how trustworthy repositories can support FAIR compliance.
- Promoting the use of certified repositories.
- Encouraging collaboration with external trustworthy repositories that complement our curation services.

## ACME Rubric: Ensuring trustworthy curation

Ensuring Trustworthy Curation	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.	
<b>Monitoring the scale of data production</b>	We are aware that our organisation needs a structured approach to monitor how many datasets are produced. We have some knowledge of projects with extensive data production.	We have developed a process to monitor how many datasets are produced, and apply this process in our priority areas	Our organisation has a documented process to monitor datasets that are produced, and can consistently identify the extent to which datasets of value are deposited in a repository, recording this for example using a CRIS (Current Research Information System).	
<b>Ensuring all retained datasets are consistently assigned Persistent Identifiers (PIDs)</b>	We are aware that our organisation needs a structured process to keep track of PIDs. We promote that data should be deposited in a repository that assigns PIDs.	We are developing a process to ensure researchers are provided with PIDs, and to keep track of PIDs assigned to datasets produced. These are mostly standard PIDs (e.g. DOIs).	Our organisation has both a data policy and a documented process in place to ensure that all datasets are deposited in a repository that assigns them a PID.	



Ensuring Trustworthy Curation	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)
Being responsible for data curation	We are informing researchers about the minimal contextual information they should provide for data that they deposit in a repository.	We support researchers upon request with curating data, e.g. to provide the relevant metadata, documentation, and file formats.	Our organisation has a process for consistently recording how our organisation's responsibilities for curating data have been met, within the scope identified by its data policy.	
Providing (access to) a data repository	We run a repository service or use an externally-provided repository and/or storage service that ensures continued bit-level integrity of the data collections it holds, its metadata, and its links to any related information submitted with it.	The repository service we use puts preservation plans into action e.g. by ensuring standard file formats are provided. The service records all actions, file migrations and administrative processes. Our organisation has a written agreement with any external providers of curation services, describing the responsibilities each provider has for curation.	The repository service we use commits to deploy the tools and expertise needed to maintain the significant properties of data, metadata and related information, for required retention periods and identified user groups.	

	Community engagement: Practice awareness, adoption and collaboration			
<b>Ensuring Trustworthy Curation</b>	<b>1) Awareness:</b> the organisation monitors community practice and makes local practitioners aware of it.	<b>2) Adoption:</b> the organisation also supports practitioners to embed community practice locally. Includes and builds on level 1.	<b>3) Collaboration:</b> 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.	<b>Engagement level (0-3)</b>
<b>Engaging with domain standards</b>	We are aware of the importance of using research data and metadata standards relevant to specific domains, and promotes awareness of current developments in these.	Through Data Management Plans and other research support, we steer researchers towards standards for their domain, and relevant to the repositories they intend to use, providing guidance on applying these standards.	Our organisation engages in relevant fora that develop cross-domain standards for curating research data. It also supports research groups to get involved in developing standards for their specific domain.	
<b>Engaging with trustworthy repository community</b>	We are aware of the role that trustworthy data repositories can play in making and keeping data FAIR, and we include that in our support for researchers.	Through Data Management Plans and other research support, we steer researchers towards depositing their data in a certified repository.	Our organisation encourages collaboration agreements with external certified repositories that complement our own curation services.	

