Selecting data, services, and repositories for FAIR

Angus Whyte (DCC), Marjan Grootveld, Maaike Verburg, René van Horik (DANS)

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

Selecting data, services, and repositories for FAIR

Introduction

To put the FAIR principles into practice research projects need to make choices, assisted by the professional support staff in their organisations. Research performing organisations (RPOs) therefore need to offer a supportive environment for selecting data, services, and repositories. Consider these choices from the investigators' point of view. As a researcher, how should my project decide which data, of all that it produces, it should focus effort to make FAIR? Of all the services available, how do we identify some that will help manage data in a FAIR way? And which of the multitude of repositories we could potentially use to safeguard the data at the end of the project should we choose to do that? This guide aims to help organisations assess the support they offer researchers, data stewards and others whose role involves making such choices.

The report 'Turning FAIR into Reality' (TFIR) issued by the European Commission in 2018 provides guidance for practitioners and research performing organisations (RPOs) on making choices about implementing FAIR. TFIR summarises the relevant recommendations in: Rec. 19: *"Select and prioritise FAIR Digital Objects: Research communities and data stewards should develop and implement processes to assist the appraisal and selection of outputs that will be retained for a significant period of time and made FAIR."* Also in Action 20.2 the report states that *"Mechanisms need to be established to support research communities to determine the optimal data repositories and services for a given discipline or data type."* Related recommendations made in the FAIRsFAIR project[1], which also calls on organisations to develop and implement guidance and support for making sensitive data FAIR for reuse, and for selection of appropriate trusted digital repositories (TDRs).

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

[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

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

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.

1. Defining the policy environment	- Policy environment
2. Developing sustainable business models	- Financial aspects
3. Professionalising roles through training, mentoring, and recognition	- Training
4. Supporting data management planning	
5. Defining data interoperability frameworks	} Technical preparedness
6. Selecting data, services, and repositories for FAIR	
7. Ensuring trusted curation	

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.

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

---

[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)

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 development[5], 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) framework[7].

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 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

---

[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, Sjef Öllers, Inge Slouwerhof, Ingeborg Verheul, & Jacquelijn 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

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: Selecting data, services, and repositories for FAIR

The ACME-FAIR checklist identifies seven main capability areas under this theme. Five capability areas are assessed on the *maturity* scale, measuring integration of the capability with organisation-level standards and practices. A further two 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) Data selection guidance?	☐	☐
2) Collecting metadata to inform future use?	☐	☐
3) Assessing data privacy and ethical approval?	☐	☐
4) Providing FAIR enabling services?	☐	☐
5) Recommending data repositories for long-term access?	☐	☐
Engagement		
6) Appraising data value?	☐	☐
7) Consulting on tools and services for FAIR?	☐	☐

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, IT Services, Archive or 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.

Data selection guidance

- Guidance on selecting data to comply with funder requirements, or for legal and regulatory reasons.
- Criteria for research projects to select data and related outputs of long-term value, to be made FAIR as a priority, and for estimating costs of FAIR data preparation and the benefits that may be gained.
- Measurable objectives for building collections of FAIR data and related research outputs that meet identified use cases for data reusers.

Collecting metadata to inform future use

- Information gathering to identify and describe research data that must be kept for contractual, legal and regulatory reasons
- Applying standards for data citation and discovery metadata, and assigning persistent identifiers (PIDs) to data, with metadata inter-linked with related outputs
- Publishing (meta)data on all publicly-funded research outputs of long-term value, using repositories that apply community-recognised controlled vocabularies.

Assessing data privacy and ethical approval

- Offering general guidance on the GDPR and on Codes of Conduct for Responsible Research and Innovation

- Ensuring that relevant guidance is offered to projects on reusing FAIR data, and on the privacy and ethical risks involved in preparing their collected data to be FAIR
- Offering fully integrated processes for data management, GDPR compliance, and research ethics, applying standard operating procedures with committee oversight.

Providing FAIR enabling services

- Offering pointers to sources of tools and services recommended by Research Funding Organisations and Research Data Infrastructures
- Monitoring the availability of FAIR-enabling services and applying frameworks for assessing these, to offer staff guidance on a range of services that comply with local standards and policies
- Providing a choice of FAIR-enabling services, whether internally developed or provided by others through service-level agreements, and enabling integrated access to these through single-sign-on processes and infrastructure.

Recommending data repositories for long-term access

- Providing guidance to researchers about where they can find data repositories for both finding and archiving research data, including the benefits of doing so
- Recommending discipline-specific repositories as well as cross-disciplinary ones that are certified as trustworthy, and offering these recommendations as part of the support for data management plans
- Maintaining a list of selected trustworthy repositories or community-endorsed data publishing platforms that are considered leading in all the various disciplines within the organisation.

Appraising data value

- Criteria for selecting data likely to be of long-term value, and which should therefore be made FAIR
- Offering researchers domain-relevant guidance or dedicated support to work with Research Data Infrastructures, journals and repositories to identify data to prioritise for FAIR
- Enabling researchers and research support professionals to participate in community efforts to define the data and metadata characteristics of most value to their community, and contribute to its standards for evaluating FAIRness of data.

Consulting on tools and services for FAIR

- Offering tools to raise awareness of FAIR when writing (or updating) Data Management Plans, as well as to evaluators of DMPs and to all colleagues involved in publishing research data
- Offering support to research projects with adoption of domain-relevant metadata standards and identifiers in the tools and services they use to manage their workflows
- Enabling researchers and research support professionals to help establish and maintain profiles of the tools and services their community should use to implement FAIR principles.

Selecting data, services, and repositories for FAIR - ACME Rubric

Selecting Data Services and Repositories for FAIR	Maturity			Maturity level (0-3)
	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.	
Data selection guidance	Our guidance identifies which research project data and related outputs need to be made FAIR for compliance with funder requirements, or for legal and regulatory reasons.	Our guidance offers criteria for research projects to select data and related outputs of long-term value, to be made FAIR as a priority. We offer guidance on the factors likely to affect the estimated costs of FAIR data preparation and the benefits that may be gained.	Based on its research strategy, our organisation defines measurable objectives for building collections of FAIR data and related research outputs that meet identified use cases for data reusers.	
Collecting metadata to inform future use	We gather enough information from research projects to identify and describe research data that must be kept for contractual, legal and regulatory reasons. We understand what our organisation needs to do to make this information findable and accessible, to guide decisions about making project data FAIR.	To share metadata about our organisation's research we apply standards for data citation and discovery metadata, and assign persistent identifiers (PIDs) to data and related outputs. We are taking steps to ensure our systems can inter-link the metadata from the various types of outputs (data, articles, code etc.)	Our organisation's service publishes (meta)data on all publicly-funded research outputs of long-term value. The repositories we use are applying controlled vocabularies and other semantic resources that either conform to published standards or are broadly accepted in the respective communities.	

Selecting Data Services and Repositories for FAIR	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)**
Assessing data privacy and ethical approval	We are aware of researchers' needs for guidance on how to comply with data privacy regulations, and on ethical data collection and management. We offer general guidance on the GDPR and on Codes of Conduct for Responsible Research and Innovation.[8]	We liaise with colleagues who provide advice to researchers on the privacy and ethical risks involved in their projects. We work together to ensure that relevant guidance is offered to projects on reusing FAIR data and in preparing their collected data to be FAIR. We include relevant guidance in our support processes for research data management, e.g. on Data Management Plans.	Our organisation fully integrates its processes for data management, GDPR compliance, and research ethics. There is a committee to oversee these processes, which include standard operating procedures for assessing, monitoring and controlling the privacy and ethical risks.	
Providing FAIR enabling services	We are aware of recommended sources of tools and services to help researchers make data FAIR during the lifecycle of their project, e.g. from Research Funding Organisations and Research Data Infrastructures. We advise researchers to consider these recommended sources when they select services to make data FAIR	We monitor the services available nationally, e.g. through Research Data Infrastructures, and apply frameworks for assessing FAIR-enabling services, to offer staff guidance on a range of services that comply with local standards and policies.	Our organisation offers researchers a choice of FAIR-enabling services, whether internally developed or provided by Research Data Infrastructures and other service providers recommended by Research Funding Organisations. We negotiate service-level agreements for our recommended services, and integrate access to these through our single sign-on processes and infrastructure.	

---

[8] e.g. ALLEA: https://www.allea.org/wp-content/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf

Selecting Data Services and Repositories for FAIR	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)
Recommending data repositories for long-term access	We provide guidance to researchers about where they can find data repositories for both finding and archiving research data. The guidance identified benefits of providing long-term access to data by depositing in disciplinary and cross-disciplinary repositories, including trustworthiness criteria, and how to find appropriate repositories.[9]	We recommend specific repositories for selected disciplines with significant representation in our organisation, as well as a cross-disciplinary repository that is certified as trustworthy. Guidance on these recommendations is made available to researchers through our support for Data Management Plans.	Our organisation maintains a list of trustworthy data repositories or community-endorsed data publishing platforms that are considered leading in all the various disciplines within the organisation. This information is integrated with our support for Data Management Plans.	

---

[9] e.g. using Repository Finder (https://repositoryfinder.datacite.org) or FAIRsharing.org (https://fairsharing.org)

	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)**
Appraising data value	We offer generic guidance to researchers on criteria for selecting data likely to be of long-term value, and which should therefore be made FAIR. We make this guidance available to researchers early in their projects, e.g. when supporting them with their Data Management Plans.	We support researchers to work with Research Data Infrastructures, journals and repositories to identify data of value in their community, to prioritise effort on making data as FAIR as it can be before the data is published. We offer researchers domain-relevant guidance or dedicated support to achieve this.	Our organisation encourages researchers and research support professionals to participate in community efforts to define the data and metadata characteristics of most value to their community, and contribute to its standards for evaluating FAIRness of data.	
Consulting on tools and services for FAIR	We offer tools to raise awareness and support researchers and colleagues to apply FAIR from the beginning of research projects, including when writing (or updating) Data Management Plans, as well as to evaluators of DMPs and to all colleagues involved in publishing research data.	We offer support to research projects to encourage their adoption of domain-relevant metadata standards and identifiers in the tools and services they use to manage their workflows.	Our organisation encourages researchers and research support professionals to help establish and maintain profiles of the tools and services their community should use to implement FAIR principles,. e.g. by employing domain-relevant metadata and identifiers in project workflows	