



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

D3.4 Recommendations on practice to support FAIR data principles

Work Package	WP3, FAIR Data Policy and Practice
Lead Author (Org)	Laura Molloy (CODATA)
Contributing Author(s) (Org)	Josefine Nordling (CSC), Marjan Grootveld, René van Horik (DANS), Angus Whyte, Joy Davidson, Patricia Herterich (DCC), Ivan Martin , Eva Méndez (UC3M), Pedro Principe, André Vieira (MINHO), Ari Asmi (UH)
Due Date	30.06.2020
Date	30.06.2020
Version	1.0
DOI	https://doi.org/10.5281/zenodo.3924131

Dissemination Level

<input checked="" type="checkbox"/>	PU: Public
<input type="checkbox"/>	PP: Restricted to other programme participants (including the Commission)
<input type="checkbox"/>	RE: Restricted to a group specified by the consortium (including the Commission)
<input type="checkbox"/>	CO: Confidential, only for members of the consortium (including the Commission)

Abstract

Building upon an analysis of the research data practice landscape in 2019, FAIRSF AIR has prepared a series of recommendations for practical actions to support the realisation of a FAIR ecosystem. These recommendations will be used to inform the development of guidance resources to support further adoption of FAIR data standards and practices by research communities. They are released as a living document that will be refined to reflect the forthcoming work in FAIRSF AIR, other projects funded under the INFRAEOSC-05-2018-2019 call, and other relevant initiatives.

Versioning and contribution history

Version	Date	Authors	Notes
0.1 -0.2	03/2020	Laura Molloy (CODATA), Angus Whyte (DCC), Joy Davidson (DCC)	Initial drafts of document
0.3	29/05/2020	WP members Ari Asmi (UH), Joy Davidson (DCC), Marjan Grootveld (DANS), Patricia Herterich (DCC), Ivan Martin (UC3M), Eva Méndez (UC3M), Laura Molloy (CODATA), Josefine Nordling (CSC), Pedro Principe (MINHO), René van Horik (DANS), André Vieira (MINHO), Angus Whyte (DCC)	Draft of substantive sections by WP members
0.4	29/05/2020	Laura Molloy (CODATA), Angus Whyte (DCC), Joy Davidson (DCC)	Draft for internal review
0.5	02/06/2020	Laura Molloy (CODATA), Angus Whyte (DCC), Joy Davidson (DCC)	Draft for internal review, edited
1.0	30/06/3030	Laura Molloy (CODATA), Angus Whyte (DCC), Joy Davidson (DCC), Marjan Grootveld (DANS), Patricia Herterich (DCC), Josefine Nordling (CSC), René van Horik (DANS)	Revisions to address internal review comments

Disclaimer

FAIRSF AIR has received funding from the European Commission's Horizon 2020 research and innovation programme under the Grant Agreement no. 831558 The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.

Abbreviations and Acronyms

Competence centre	FAIR Data Stewardship Competence Centre: A shared hub of expertise in implementing FAIR data stewardship principles, offering leadership, coordination and cataloguing services to connect relevant people, guidance, learning resources and curricula (FAIRSF AIR D6.1)
CORDIS	Community Research and Development Information Service
DMP	Data Management Plan
EGFC	European Group of FAIR Champions
EOSC	European Open Science Cloud
ESFRI	European Strategy Forum on Research Infrastructures
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
HEIs	Higher Education Institutions
HLAC	High Level Advisory Committee
IPR	Intellectual Property Rights
KOS	Knowledge Organisation System
RDA	Research Data Alliance
RPO	Research Performing Organisation e.g. university, research institute etc.
RSE	Research Software Engineer
TDR	Trusted Digital Repository
TFIR	<i>Turning FAIR Into Reality report and action plan (European Commission, 2018)</i>

Executive Summary

Previous work by FAIRSF AIR Work Package 3 (Data Policy and Practice) has identified areas in which contemporary research culture could benefit from further support or clarification to help make more research data and other digital research outputs 'FAIR', i.e. findable, accessible, interoperable and reusable. This deliverable draws on these findings, and follows the structure and themes of the 2018 European Commission report, *Turning FAIR Into Reality*, to set out recommended, practical actions that meet those gaps in provision.

Four thematic sections in the report identify recommended actions to enable FAIR data:

A. Develop and implement data sharing and interoperability frameworks: The recommendations made here encourage collaboration on research community-specific and inter-domain agreements and uptake of standards, shared vocabularies, ontologies and metadata schemas to stimulate new research collaborations based upon cross-community discovery, retrieval and reuse of research datasets and other digital research outputs.

B. Ensure data management is supported by data management plans (DMPs): This section encourages the enhanced use of data management plans by researchers as dynamic, human- and machine-readable research support resources.

C. Develop professional support for FAIR data: Here, we recognise the invaluable contribution of emerging data professional roles such as the data steward and research software engineer and advocate for the emergence of related training and qualification opportunities, career path development, and professional societies. We also recommend the development of a self-assessment framework for research institutions and infrastructures to gauge progress towards building their FAIR data stewardship capabilities.

D. Ensure trusted curation of data: These recommendations encourage research communities to explore the use of trusted digital repositories and to find out more about how they can help their research by providing suitable, secure places of deposit for all types of research data including sensitive data.

Each aim is supported by proposed actions which set out practical steps for research communities, research support personnel (including data stewards and software engineers) and - where appropriate - research support actors such as infrastructures and institutions to take. Each recommendation is explained in non-specialist language, so that research communities and research support personnel can work together on these recommendations.

Table of contents

Executive Summary	4
1. Introduction	6
1.1 Aim of these recommendations	6
1.2 Target audience	7
1.3 Background concepts for FAIR	9
1.4 Scope and structure of the recommendations	10
1.4.1 Key themes	10
2. Recommendations on implementing FAIR for research communities and research support actors	12
Theme A: Develop and implement data sharing and interoperability frameworks	12
Theme B: Ensure data management is supported by data management plans (DMPs)	17
Theme C: Develop professional support for FAIR data	21
Theme D: Ensure trusted curation of data	28
3. Next steps	32
Annex 1. Selected Recommendations from Turning FAIR into Reality	33
Bibliography	36

1. Introduction

1.1 Aim of these recommendations

This report sets out practical recommendations on making data FAIR, aimed at research communities and the organisations and roles that support them in that endeavour. Creating, managing, accessing, and using FAIR data depends upon a number of factors in the ecosystem. Researchers need to be made aware of the value and benefits of FAIR data. They must have access to knowledgeable support staff. They need the support of data services and infrastructures that enable FAIR data as well as other good research practice.¹ These need to be aligned with the harmonisation of research policies, frameworks and governing legislation to better support FAIR data.

The fostering of a FAIR data ecosystem is the joint effort of many projects and initiatives. We aim to support these collective efforts by drawing together actions for an easier transition to a culture of making research data FAIR. This report highlights useful examples of actions already being taken to implement and embed FAIR data management and proposes further actions towards that goal for research communities, support roles, and organisational actors that support research.

The FAIRsFAIR project amplifies and communicates existing work from other initiatives in order to draw these useful resources to the attention of a wider audience, and to avoid unnecessary duplication and redundancy. This current set of recommendations draws on the European Commission Expert Group on FAIR Data (2018) publication, *Turning FAIR Into Reality report and action plan* (abbreviated to TFIR throughout this document).

The TFIR recommendations shaped FAIRsFAIR's previous analysis in D3.2 FAIR Data Practice (Whyte et al. 2019). That report offers a broad cross-disciplinary analysis of practices to support FAIR data production. A snapshot of the state of play in 2019 across the European data-intensive research space, the report was based on desk research, interviews with EOSC cluster project representatives, a workshop discussion, and survey results. It came to a number of conclusions that inform this report, including the following:

1. Awareness of FAIR principles among the six research communities we reviewed was reported to be very low, and the level of culture change towards FAIR data production challenging to assess.²
2. To improve monitoring, and encourage and promote collaborations to implement FAIR, a suitable capability model for organisations to self-assess their FAIR enabling practice(s) is needed.
3. To guide implementation and inspire change towards FAIR-enabling practices an inventory of FAIR good practice is needed. Short case studies we refer to as 'implementation stories'

¹ FAIRsFAIR M2.7 report offers guidance on assessing whether services 'enable' FAIR, 'respect' FAIR, or actually 'reduce' FAIR. See Koers, Hylke, Gruenpeter, Morane, Herterich, Patricia, Hooft, Rob, Jones, Sarah, Parland-von Essen, Jessica, & Staiger, Christine. (2020). Assessment report on 'FAIRness of services' (Version 1.0). Zenodo.<https://doi.org/10.5281/zenodo.3688761>

² As a proxy, we reviewed the extent of FAIR initiatives and their take-up in each of six disciplinary areas, representing the ESFRI clusters. These are described at <https://eosc-portal.eu/news/five-new-esfri-cluster-projects-eosc-panorama>

should illustrate how communities collaborate to produce FAIR data, enabled by any relevant research instruments³ and service providers.

4. To lead by example, FAIRSFair commits to support change in 6 areas we selected as initial priorities, based on the available evidence and the capabilities of the project partners. These 6 areas covered the following:
 - a. Methods for building consensus on metadata and interoperability frameworks within and across communities
 - b. Machine-actionable DMP templates and guidelines on using them to inform subsequent data management activities
 - c. Assigning PIDs at the point of data creation e.g. to instruments
 - d. Terminology for competence centres to annotate and retrieve training materials on enabling FAIR
 - e. Managing FAIR support costs, including resource models for coordinating research support teams
 - f. Good practice for researchers, repositories and ethics committees on selecting and preparing sensitive data to be FAIR

Based on this work and related initiatives, FAIRSFair has prepared this initial set of recommendations on practice to implement the FAIR data principles. These recommendations supplement the work presented in D3.3 Policy Enhancement Recommendations (Davidson et al. 2020) – naturally, there is much interplay between policy and practice as the two dimensions ideally inform and improve each other.

FAIRSFair is primarily about implementation, and the focus in the current deliverable is on highlighting useful examples of actions that can be taken now by researchers and others who support researchers with data management and stewardship. The examples we provide draw on the recommendations of related projects including EOSC-hub, EOSCpilot, RDA Europe, OpenAIRE, FREYA, and others who have an interest in and/or are actively working to support changes in research practice towards FAIR data production. Neither this list of projects nor the actions we recommend are intended to be exhaustive and FAIRSFair welcomes additions from the wider community. It is intended that this deliverable is a ‘living document’ during the life of the FAIRSFair project; all versions produced will be made publicly available in the FAIRSFair collection on the Zenodo repository.⁴

1.2 Target audience

This report and its recommendations are aimed at people in emerging roles that support researchers in the production and management of data, code and related research objects. Broadly speaking this is the practice of data stewardship, which has been described as a shared responsibility of researchers and others to make data FAIR and keep it FAIR.⁵ More specifically it

³ This refers to instruments in the broadest sense, including any plans, protocols, equipment, sensors, or devices used to produce data.

⁴ <https://zenodo.org/communities/fairsfair/>

⁵ Whyte, A., Leenarts, E., de Vries, J. et al. (2019) Strategy for Sustainable Development of Skills and Capabilities, EOSCpilot D7.5 <https://eoscpiot.eu/content/d75-strategy-sustainable-development-skills-and-capabilities>

covers “The responsible planning and executing of all actions on (digital) data before, during and after a research project, with the aim of optimising the usability, reusability and reproducibility of the resulting data”.⁶ This involves two support roles we target in particular, those of Data Stewards and Research Software Engineers, which can be described as follows.

Data stewards: Support staff from research communities and research libraries, and those managing data repositories. Data stewards typically vary in the tasks they cover in three main areas:

- embedded and operational
- generic and advisory
- policy, strategy, and coordination

Research software engineers: Occupy a role combining professional software engineering expertise with an intimate understanding of research (TFIR, p. 49).⁷

Individuals in these roles cannot put FAIR into practice on their own, but they can be catalysts for action by other relevant stakeholders, and our recommendations also call for support from those described in the TFIR report as follows:

- **Research communities:**
 - **Research practitioners** from all research fields, clustered around disciplinary interests, data types or cross-cutting grand challenges;
- **Research support actors:**
 - **Data service providers:** Domain repositories, research infrastructures (e.g. ESFRIs) and e-infrastructures, institutional, community and commercial tools and services.
 - **Coordination fora:** Global and national bodies such as the Research Data Alliance, CODATA, WDS Communities of Excellence, GO-FAIR, German Data Forum (RatSWD), Dutch Coordination Point (LCRDM), and similar initiatives.
 - **Institutions:** Universities and research performing organisations (RPOs).
 - **Publishers:** Not-for-profit and commercial, Open Access and paywall publishers of research papers and data.

(Definitions from TFIR, p. 61)

FAIRSFair has a focus on ‘semantic artefacts’, i.e. data standards and terminologies, and experts on these may be found in any of the above stakeholder categories. But for the sake of clarity we include in the targets of this report the ontologists and knowledge engineers, informatics professionals, and others across the semantic web community who are engaged in defining and maintaining these artefacts.

⁶ Christine Staiger, Mijke Jetten, Jasmin Böhmer, Inge Slouwerhof, Marije van der Geest, Celia W.G. van Gelder, & Salome Scholtens. (2019, June 12). Data stewards function landscape and its stakeholders. Zenodo. <http://doi.org/10.5281/zenodo.3460552>

⁷ Society of Research Software Engineering. <https://society-rse.org/about/>

FAIRSFAR also has a focus on ‘Competence Centres’ a relatively new term but one that may helpfully summarise what the above stakeholders can do to collaboratively support these FAIR practice recommendations.⁸

In practice, close cooperation between all the above stakeholders will be needed to act on the recommendations. And although we highlight the new Data Steward and Research Software Engineer roles, these (and other) research support actors who are responsible for managing service delivery at organisational level also play a crucial part in implementing the recommendations.

There are of course other stakeholders who can exert a major influence on practice, including research funders, EOSC governance bodies, and policy makers at national and institutional level. These stakeholders are addressed by recommendations of the EOSC FAIR Working Group,⁹ as well as by the FAIRSFAR policy recommendations.¹⁰

1.3 Background concepts for FAIR

The FAIR Guiding Principles for Scientific Data Management and Stewardship (henceforth ‘the FAIR principles’ / ‘the principles’) emerged in 2016 (Wilkinson et al. 2016) from a cross-organisational, interdisciplinary effort to refine the availability and usability of digital research data and other digital outputs of research, regardless of their public availability (European Commission 2018). At their highest level, the FAIR principles advocate for making research data, software and other digital outputs of research as Findable, Accessible, Interoperable and Reusable as possible.

The FAIR principles are created for the consideration of all those who search for, create, manage, share, and use research data. Given the extent to which data handling is - or is becoming - automated, the FAIR principles are also written with machine-actionability in mind. Machine-actionable is defined as “the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention” (GO-FAIR, 2017).

A major step forward in progress towards achieving these goals was made with the publication of the *Turning FAIR Into Reality report and action plan* (European Commission 2018), from the European Commission Expert Group on FAIR Data. This action plan is comprised of a set of six pillars or groups of recommendations, arranged as follows:

Actions to define:

1. Concepts for FAIR implementation

Actions to implement:

2. FAIR culture

⁸ FAIRSFAR D6.1 offers the following definition: ““FAIR Data Stewardship Competence Centre: A shared hub of expertise in implementing FAIR data stewardship principles, offering leadership, coordination and cataloguing services to connect relevant people, guidance, learning resources and curricula” (p.7) <https://doi.org/10.5281/zenodo.3549790>

⁹ Deniz Beyan, Oya, Chue Hong, Neil, Cozzini, Stefano, Hoffman-Sommer, Marta, Hooft, Rob, Leminen, Liisi, ... Teperek, Marta. (2020, June 23). Seven Recommendations for Implementation of FAIR Practice. Zenodo. <http://doi.org/10.5281/zenodo.3904139>

¹⁰ Davidson, Joy, Grootveld, Marjan, Whyte, Angus, Herterich, Patricia, Engelhardt, Claudia, Stoy, Lennart, & Proudman, Vanessa. (2020). D3.3 Policy Enhancement Recommendations (Version 1.0). Zenodo. <https://doi.org/10.5281/zenodo.3686900>

3. FAIR ecosystem
4. Skills for FAIR

Actions to embed and sustain:

5. Incentives and metrics for FAIR data and services
6. Investment in FAIR

(TFIR p. 17)

The FAIRsFAIR project “aims to supply practical solutions for the use of the FAIR data principles throughout the research data life cycle” (FAIRsFAIR, 2020). FAIRsFAIR Work Package 3, ‘FAIR Data Policy and Practice’, aims to directly inform data-related research in order to help make and keep data FAIR and embed these practices in research communities. Accordingly, here we primarily engage with TFIR pillar 2, ‘Actions to implement FAIR culture’.

This work is taking place at the same time as a number of other projects also concerned with the adoption of the FAIR principles. Amongst these, the FAIR Practice task force of the EOSC FAIR working group notes that “most scientific communities are still struggling with understanding and implementing FAIR principles” (Marttila, et al. 2020). It is hoped that these recommendations can ultimately help to make the FAIR principles more actionable, and will support the ability of research communities to make their data FAIR, and to deposit and retrieve FAIR data through the use of trusted digital repositories in order to stimulate reuse and widen the impact of FAIR research data.

1.4 Scope and structure of the recommendations

The FAIRsFAIR report *D3.2 FAIR Practice Analysis* recommended establishing an ‘inventory of exemplars’ (D3.2, p. 71) to inspire further implementation of the TFIR recommendations by research communities and the professional services that support them (e.g. research data management services, legal services).

The recommendations in this short report aim to help people in these groups to identify some practical commitments they can make towards producing FAIR data in collaboration with repositories. Repositories play key roles in the appraisal, curation and preservation of FAIR data. The recommendations are grouped under four main ambitions, outlined as themes A - D below¹¹, with recommendations numbered accordingly, e.g. A1 designates Theme A, Recommendation 1. Each recommendation contains an explanatory discussion including action points picked out in bold type. Each recommendation box concludes with the summary of actions for various stakeholder groups.

1.4.1 Key themes

The themes and their recommendations draw on the TFIR recommendations and develop topics proposed by the previous D3.2 report (FAIR Practice Analysis) (Whyte, et al. 2019) for further action

¹¹ For convenient reference, Annex 1 lists the relevant TFIR recommendations and their corresponding action points that have informed the follow-up recommendations in this report.

by FAIRSF AIR, in collaboration with relevant communities and projects as specified within the recommendation discussion.

Theme A: Develop and implement data sharing and interoperability frameworks

A1: Describe research outputs using agreed terminologies and metadata standards to make data FAIR

A2: Build a culture of data citation

Theme B: Ensure data management is supported by data management plans (DMPs)

B1: Formalise and support appropriate data management plans (DMPs) for FAIR data

B2: Develop roadmaps, guidance and workflows for machine-actionable data management plans (DMP) to inform FAIR data stewardship

Theme C: Develop professional support for FAIR data

C1: Define and manage FAIR support costs and resources

C2: Develop and implement models for coordinating and supporting data stewards and research software engineers

C3: Develop and implement terminology for competence centres to annotate and retrieve training materials on enabling FAIR

C4: Develop and implement a self-assessment framework for Research Infrastructures, institutions, and other FAIR competence centres

Theme D: Ensure trusted curation of data

D1: Develop and implement guidance and support for selection of appropriate trusted digital repositories (TDRs)

D2: Develop and implement guidance and support for making sensitive data FAIR for reuse

2. Recommendations on implementing FAIR for research communities and research support actors

This section consists of four themes, labelled A-D, that elaborate on an aspect of making data and research outputs more FAIR. Each of the four themes is supported by several recommendations which are explained in accessible language. These are numbered for clarity. The actions to be taken by research communities, research support roles and, where relevant, actors such as Research Infrastructures, are also set out.

Theme A: Develop and implement data sharing and interoperability frameworks

Interoperability frameworks are various sorts of agreements about how data and other research outputs are described, formatted, shared, and used. These include agreements on standards, terminologies, processes, and technologies to make and keep data as FAIR as possible and to encourage data sharing and data reuse.

Recommendation A1:

Describe research outputs using agreed terminologies and metadata standards to make data FAIR

As pointed out in the *Turning FAIR Into Reality* report (European Commission 2018), vocabularies, ontologies and category definitions (together known as semantic artefacts) should be FAIR themselves so they can completely fulfil their role in the FAIR data ecosystem. FAIRSFAR considers semantic artefacts to vary in their ‘strength’.¹² For example artefacts with weak semantics include lists, glossaries and controlled vocabularies. Taxonomies or XML schema are semantically stronger as they are hierarchically organised. A thesaurus is stronger as it indicates terms that are related to each other. A formal ontology is stronger still, as it adds relationships of different types. Metadata standards may embody any of these types of structure, and there is a diverse range of data models and formats available to ‘serialise’ or describe data using them.

TFIR recommendation 7, ‘Support semantic technologies’, also stresses the importance of harmonising the formats, styles and definition languages to allow both researchers and machines to find, access and reuse them.

- **Research software engineers and data stewards have an important role to play in making semantic artefacts FAIR and should work together with research communities on defining and promoting domain standards (to ensure that semantic artefacts reflect shared community terminology)**

¹² Le Franc, Yann, Parland-von Essen, Jessica, Bonino, Luiz, Lehvälaiho, Heikki, Coen, Gerard, & Staiger, Christine. (2020). D2.2 FAIR Semantics: First recommendations (Version 1.0). FAIRSFAR. <https://doi.org/10.5281/zenodo.3707984>

Researchers can greatly benefit from using domain-specific terminologies and standards that are grounded in so called foundational ontologies. These are independent of a particular domain, and resolve ambiguity in what their terms mean. They also allow researchers to formulate their hypotheses in terminology that is both domain-specific and also understandable across disciplines. Researchers' process of working with concepts from different standards and terminologies is better supported if these standards and terminologies are well documented including details about their provenance. Using standard terminologies also improves the interoperability of metadata descriptions. This helps to make data more Findable to other users and machines while searching across different repositories. By using consistent terminology, data can be made more understandable and retrievable, and therefore more Reusable. Standard terminologies and ontologies should be published so that the communities who need to be involved in their maintenance can more easily find and use them.

Furthermore, standards and terminologies need to have well-documented digital rights management (DRM) interoperability¹³ as this is a prerequisite for enabling FAIR reuse. Applying appropriate licenses to digital objects¹⁴, such as Open Government licensing¹⁵ or CC BY 4.0¹⁶, and adding licence information in both human and machine readable formats, can ease reuse by clarifying the legal conditions that apply.¹⁷

To help researchers make informed choices about standards that others in their community are using, GO-FAIR is developing a *FAIR Implementation Profile (FIP)*¹⁸. This aims to guide research communities to document their FAIR related choices and publish these choices in a machine-actionable and reusable form. Each FIP will be registered in the FAIR Convergence Matrix.¹⁹ This will include profiles of standards identified by a range of research communities. This opens up new possibilities for exchange of FAIR data practices within and across disciplines.

- **Data stewards, working with researchers, should use resources available to help produce FAIR-compliant metadata**

There are resources available to help support researchers to improve the quality of their data by choosing community-specific and community-validated metadata standards. One resource that emphasises working collaboratively on producing machine-actionable metadata is the *Metadata for Machines (M4M)*²⁰ event format initiated by members of GO-FAIR and RDA. The metadata is assessed in an approach similar to the familiar hackathon format, where domain experts work

¹³ Kamperman F. (2007) Digital Rights Management Interoperability. In: Petković M., Jonker W. (eds) Security, Privacy, and Trust in Modern Data Management. Data-Centric Systems and Applications. Springer, Berlin, Heidelberg

¹⁴ See <https://www.dcc.ac.uk/guidance/how-guides/license-research-data> for guidance on choosing an appropriate licence for data.

¹⁵ Open Government License <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

¹⁶ Creative Commons BY 4.0 <https://creativecommons.org/licenses/by/4.0/>

¹⁷ FAIRSFAR (2020). D2.2 FAIR Semantics: First recommendations report. DOI: 10.5281/zenodo.3707984

¹⁸ Described in <https://osf.io/c4fth/>

¹⁹ <https://www.go-fair.org/today/fair-matrix/>

²⁰ <https://www.go-fair.org/resources/go-fair-workshop-series/metadata-for-machines-workshops/>

together towards a mutual goal. Once a cross-disciplinary scientific community has identified what terms it needs, based on available domain metadata standards, these can be addressed separately for each letter of FAIR. The selected definitions are formulated into machine-actionable templates and are made openly available and accessible for reuse. Another easy and accessible approach to improving researchers' understanding of the need for quality metadata is through an interactive LEGO® Metadata game.²¹ By trying to reproduce a LEGO design based only on the metadata provided, researchers can see very quickly that a lack of good documentation hampers reproducibility. The game provides rules for planning the metadata, standards available, documenting the research and communicating the different stages of the research process. The game also showcases the importance of adhering to standards and demonstrates how researchers within the same scientific discipline understand each other's data.

Registries of semantic resources are another key resource for supporting the production of research data that is more interoperable. Vocabularies and Knowledge Organisation Systems (KOS) such as classifications, thesauri and authority files are listed in BARTOC (Basel Register of Thesauri, Ontologies and Classifications²²). This is a comprehensive registry of Knowledge Organisation Systems, which centrally gathers and promotes existing KOSs to facilitate searches and knowledge sharing. BARTOC is a rich source of information that can be used as a basis for improving the quality of the descriptions of research outputs. All the KOS used in FAIR data production should be machine-readable, and the broader adoption of Networked KOS (NKOS) should be encouraged.²³

Research communities	Researchers, working with data stewards, should make use of domain-specific terminologies and standards grounded in foundational ontologies and keep these up to date. This will improve the interoperability of the metadata descriptions and allows other researchers to find and reuse their data.
Research support roles	Data stewards should assist researchers in finding the most commonly used domain specific terminologies and standards and contribute to keeping these up-to-date. With the help of research communities, research software engineers and data stewards should develop machine-readable FAIR semantic artefacts and domain standards and ensure these are properly maintained, documented and used.
Further work by FAIRSF AIR	<ul style="list-style-type: none"> FAIRSF AIR will provide examples of good practice in building community consensus around interoperability frameworks

²¹ Donaldson, M. and Mahon, M. (2019); LEGO® Metadata for Reproducibility. University of Glasgow. DOI: <http://dx.doi.org/10.36399/gla.pubs.196477>

²² www.bartoc.org

²³ <https://nkos.slis.kent.edu/>

	<ul style="list-style-type: none"> ● FAIRsFAIR will contribute to further development of guidance materials on running M4M workshops. ● FAIRsFAIR will offer good practice guidance for research communities on creating, finding and using terminologies that can make their data more FAIR, and on contributing to registries of these (e.g. BARTOC).
--	---

**Recommendation A2:
Build a culture of data citation**

A Persistent Identifier (PID) is a long-lasting reference to an object such as a publication, dataset, project, person or organisation. Several identifier formats for data objects have emerged as well as systems and services to support these. Such systems generate unique identifiers and ensure they resolve, i.e. connect to metadata about the objects to which they refer (the object’s ‘landing page’). PIDs are fundamental building blocks for data citation, because they offer a trustworthy way to refer to data and potentially connect it to other people or objects involved in its production. This can increase the visibility of people who create, curate and share data as well as the datasets and other research outputs themselves. In all scientific disciplines it is common to create references to information sources, such as literature or archival records, for which several widely used citation systems are available. These types of citations have become de facto standards in scientific communication and are fundamental e.g. to assess the quality of scientific results.

The role, function and value of PIDs in the science landscape are clear, but their use is still not widespread in research practice. FAIRsFAIR aims to support research communities to build a culture of data citation by using PIDs in their scientific activities with support from Research Infrastructures. Based on an assessment of the current maturity of available PID systems, FAIRsFAIR will support a number of practices. The work and results of the FREYA project form the basis for this support.²⁴ FREYA extends the infrastructure for persistent identifiers (PIDs) as a core component of open research in the EU and globally. Four types of objects have well-established PID services that FREYA interconnects:

1. Publications, using DOI²⁵
2. Datasets, using DOI, URN:NBN or Handle
3. Researchers, using ORCID²⁶
4. Research organisation, using ROR²⁷

²⁴ <https://www.project-freya.eu/en>

²⁵ DOI handbook and FAQs available at <https://www.doi.org/faq.html>; it is worth noting that some repositories such as many institutional repositories and Zenodo assign DOIs to digital research objects such as publications and datasets without charge to the user.

²⁶ Researcher-friendly guidance and sign-up is available at <https://orcid.org/>

²⁷ <https://ror.org/>

A PID service requires specific actions to be taken by its user - including the researcher that creates, uses and refers to different types of objects. In all cases the user has to create metadata about the object that will be connected to the PID or use a service that helps to create the metadata: e.g. a title of a publication, a description of a dataset and the name of a researcher.

PIDs for publications, datasets, researchers and organisations are among the most mature available PID systems. PIDs for other objects, such as scientific instruments, grants, services, and projects are emerging. FAIRSF AIR will assess how these types of PIDs can be implemented in scientific discourse, and offer examples and guidance.

Research communities	<p>Researchers should get an ORCID identifier and use this identifier in publications and other communication activities. This will improve the disambiguation of their research from that of other researchers with similar names, and improve the findability of their research outputs.</p> <p>Research communities should agree on the PID policies and practices that fit their needs for version control of the objects they work with, and for the research to be reproducible. These policies and practices should be documented and kept up to date.</p>
Research support roles	<p>Data stewards should advocate and support the use of PIDs for objects such as publications, data sets, organisations, and researchers.</p>
Research support actors	<p>Institutions and service providers should ensure that they adhere to the PID policies of the communities they serve. This means implementing PID systems where feasible (instruments, etc) and offering training and services that support reasonable PID use, and sustainable solutions e.g. on how researchers should use ORCIDs , or use handle servers.</p>
Further work by FAIRSF AIR	<ul style="list-style-type: none"> ● FAIRSF AIR aims to support research communities, working with research infrastructures, to build a culture of data citation by using PIDs in their scientific activities. ● Based on an assessment of the current maturity state of available PID systems, FAIRSF AIR will offer additional guidance, building on the work and results of the FREYA project form.

Theme B. Ensure data management is supported by data management plans (DMPs)

Data management plans (DMPs) help research groups to agree and confirm arrangements and procedures around how data is collected, who is responsible for its management, where and how it will be kept, and how access to data will be provided. DMPs are increasingly required in research funding bids to demonstrate how data will be made and kept FAIR.

Recommendation B1:

Formalise and support appropriate data management plans (DMPs) for FAIR data

The FAIR principles apply to data management throughout a project, from the DMP at the beginning to outputs produced through implementing the plan. Making DMPs effective tools for producing data that is more FAIR depends on community efforts to agree what the DMP should contain, and how they should be updated throughout the research lifecycle. According to *Turning FAIR into Reality* (rec. 5), “any research project producing or collecting research data must include data management as a core element necessary for the delivery of its scientific objectives, and should address this in a Data Management Plan. The DMP should include all the relevant project outputs and be regularly updated to provide a hub of information on FAIR Digital Objects” (European Commission, 2018, p. 65).

- **Research communities, with support from data stewards, need to agree templates for DMPs that identify useful steps towards making data FAIR in a domain-relevant way.**

By agreeing on what they consider a good choice of (e.g.) standards or repositories, a community can reduce the burden on individual researchers to reinvent how to be FAIR. Many funders (including the European Commission) now expect a DMP to be produced, require the research data coming from supported projects to be FAIR, and leave it to the research team to make that happen with the help of their organisation and all relevant support actors. Scientific societies and professional associations can also help to articulate the needs of communities they represent. However many researchers lack such support, and data stewards can help to fill this gap.

Templates may be put together in a bottom-up way, based on examples relevant to the community. Useful DMP examples may be found in the LIBER DMP catalogue²⁸ and DMP Online²⁹. Nevertheless, they also need to reflect funders’ expectations of what a template needs to cover. These expectations are becoming increasingly harmonised. Science Europe and other stakeholders provide useful templates for a DMP, to identify steps towards FAIR.³⁰ Their 2018 report on Domain Data Protocols offers a snapshot of efforts to contextualise DMP guidance for different domains.³¹ FAIRSFAR D3.2 maps these to Research Infrastructures as representatives of the communities featured in the report.

²⁸ <https://libereurope.eu/dmpcatalogue/>

²⁹ https://dmponline.dcc.ac.uk/public_plans

³⁰ https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf

³¹ <https://www.scienceeurope.org/our-resources/guidance-document-presenting-a-framework-for-discipline-specific-research-data-management/>

DMP authoring platforms offer help in working with these templates, including community-relevant guidance.³² Taking this further, the GO-FAIR network proposes the *FAIR Implementation Profile* (FIP). According to this approach, which is currently under development, a FIP is a community-specific set of default choices that represents good practice in using standards, services and technologies that will give FAIR results.³³

- ***Researchers, data stewards and academic journals can use FAIR data criteria to review whether a DMP includes appropriate steps to make the data FAIR.***

Peer review principles can be applied to DMPs as to any kind of research output, with similar benefits. It can be worthwhile asking colleagues to review a plan at any stage of the project. The FAIR criteria can be used to identify whether the plan is likely to result in data that is FAIR, just as they can be used to assess whether a dataset meets those same criteria. Data stewards can support this assessment role and offer constructive advice over the life of the project on how to better prepare data to be FAIR. Some funders now also expect institutions that host the researchers they fund to assure that this happens, and take responsibility for making the data outputs FAIR. This makes FAIR data assessment tools a useful complement to Data Management Planning tools. Service providers should better support their complementary role by integrating these tools. Academic journals and publishers can also contribute; some accept DMPs for peer review, and there are potential rewards for the publishing sector in being able to better connect different outputs via the project DMP.

Research communities	<ul style="list-style-type: none"> ● Researchers should use their DMP as a guide for producing data that is FAIR, turning the DMP into a working document that records how the plan meets reality. ● Research communities should identify guidance and model DMPs that provide domain-relevant responses to the general questions about data management they are asked to complete in DMP templates.
Research support roles	<ul style="list-style-type: none"> ● Data stewards should work proactively with research communities in their organisations to build cooperation around DMP production, and work with other research support colleagues as much as possible to create common practices. ● Data stewards can benefit by using FAIR data criteria to offer constructive feedback on DMPs, helping ensure plans contain appropriate steps to prepare data that will be FAIR.
Research support actors	<ul style="list-style-type: none"> ● Institutions and other research performing organisations should agree on templates and recommendations applicable across their organisations.

³² <https://dmponline.dcc.ac.uk/> and <https://dmptool.org/> are well-established examples.

³³ <https://www.go-fair.org/today/fair-matrix/>

	<ul style="list-style-type: none"> • Service providers, including providers of DMP platforms, can potentially offer more comprehensive support for data management by using FAIR data assessment tools to review the steps a research project is taking towards meeting the FAIR criteria.
<p>Further work by FAIRSFAR</p>	<ul style="list-style-type: none"> • FAIRSFAR is developing assessment tools for FAIR data. We will investigate options for integrating such tools with DMP authoring platforms, to help data stewards review DMPs on the steps being planned to meet the FAIR criteria. • In collaboration with GO-FAIR we will offer use cases and examples to illustrate how a FAIR Implementation Profile can help to put a DMP into action.

Recommendation B2:

Develop roadmaps, guidance and workflows for machine-actionable DMPs to inform FAIR data stewardship

The *Turning FAIR into Reality* report advocates that DMPs should be FAIR outputs in their own right. Making DMPs ‘machine-actionable’ means making their content findable and accessible, exchanging that content with other systems in standardised, interoperable ways, and potentially reusing that content. Potentially this can help transform DMPs into working documents, hubs of information and tools for supporting data management more proactively. A standard for exchanging DMP content³⁴ has recently been developed by a RDA working group³⁵ with early adoption by some DMP platforms. A recent hackathon organised by the RDA working group has demonstrated the effective exchange of DMP data across several DMP platforms using the standard.³⁶

- **DMP platform providers, institutions, repositories and other data service providers should adopt the RDA Common Standard for Machine-Actionable Data Management Plans to exchange DMP content**

By supporting the RDA common standard, service providers can offer DMP authors a mechanism for their plans to trigger research support systems into responding appropriately, e.g. to further determine what is needed. In some cases, services might automatically fulfill these needs. In other cases, data stewards and other advisors may be better informed where there is a need to consult further with the research group. Data repositories can also be informed earlier regarding the projects that plan to use them, and begin a conversation - for example about anticipated storage and metadata requirements. The RDA Common Standard for machine-actionable DMPs

³⁴<https://www.rd-alliance.org/group/dmp-common-standards-wg/outcomes/rda-dmp-common-standard-machine-actionable-data-management>

³⁵ <https://www.rd-alliance.org/groups/dmp-common-standards-wg>

³⁶ <https://rda-dmp-common.github.io/hackathon-2020/>

offers potential benefit to all stakeholders. As a data exchange standard it supports data portability, and helps reduce the risk of vendor lock-in.

- **Institutions and research communities should prepare a roadmap and guidance for implementing machine-actionable DMP workflows**

Researchers and other stakeholders benefit from wider exposure of DMP content, e.g. examples of completed DMPs provide opportunities to early career researchers and students, to learn about data management approaches being adopted in their community. And by exposing DMP content early in a project, its members can receive guidance that is better tailored to their needs. Workflows, business processes and safeguards need to be established to ensure that research groups benefit by exposing DMP content within acceptable levels of risk. Where a project will use sensitive personal data, for example, a DMP that describes processing details may need to have access limitations. Risks of being ‘scooped’ are also felt. To build trusted relationships between DMP authors and service providers, get real benefits from machine-actionable DMPs, and manage any risks involved in their wider exposure there is a need for concerted action within institutions and across them, by service providers, and research communities. For example, DMP content needs to employ community-endorsed controlled vocabularies and persistent identifiers. Researchers and support roles need consultation on benefits and risks, informed by the RDA working group on Exposing DMPs, for example.³⁷ The group recommends that organisations and communities plan how to integrate DMP content into ethical approval workflows and scholarly communications systems. Some useful examples are already available³⁸.

Research communities	Researchers can benefit by learning about the possibilities that machine-actionable DMPs offer to make their DMP more useful, by connecting to other research support systems.
Research support roles	Data stewards, research software engineers and others involved in advising researchers can benefit by learning how machine-actionable DMPs can help them fulfil their support role more effectively or efficiently. They should gain familiarity with the DMP Common Standard ³⁹ , and consider giving their feedback to the draft recommendations on Exposing DMPs. ⁴⁰
Research support actors	<ul style="list-style-type: none"> ● Institutions and other research performing organisations should consider how automation may assist data stewards and others to give research projects more effective support in writing a DMP that is feasible, and in implementing their plan. ● DMP platform providers can benefit by implementing software

³⁷ <https://www.rd-alliance.org/groups/exposing-data-management-plans-wg>

³⁸ E.g. University of Manchester:

<https://www.dcc.ac.uk/blog/building-integrated-service-dmponline-university-manchester> and TU Delft

<https://www.dcc.ac.uk/blog/dmponlinetu-delft>

³⁹ <https://www.rd-alliance.org/group/dmp-common-standards-wg/outcomes/rda-dmp-common-standard-machine-actionable-data-management>

⁴⁰ This work can be tracked at <https://www.rd-alliance.org/groups/exposing-data-management-plans-wg>

	<p>support for the RDA common standard, e.g. through APIs. They should work with providers of other services, e.g. research information, data centres, repositories, and publishers, to improve efficiency and effectiveness of data management workflows. Coordination fora can further support the necessary dialogue.</p> <ul style="list-style-type: none"> ● Research ethics committees should consider how ethical approval systems can benefit from integration with DMP platforms to reduce duplication of effort and to mitigate risks to data subjects that may be associated with access to the content of the project DMP.
<p>Further work by FAIRsFAIR</p>	<ul style="list-style-type: none"> ● FAIRsFAIR will build on outputs of the RDA Exposing DMP working group, to document further examples of workflows implementing machine-actionable DMPs to support FAIR data production. ● We will contribute where appropriate to implementation of workflows with partner projects e.g. PaNOSC and ExPaNDS, to meet use cases that offer mutual benefits to repositories and the research communities that deposit data with them. ● Consulting with service providers and research communities involved in EOSC projects, we will provide guidance for institutions and infrastructures on dealing with ethical and confidentiality issues in exposing DMP content.

Theme C: Develop professional support for FAIR data

Research support roles are critical for making and keeping data FAIR. Some of the necessary skills are well-recognised competence areas including research management, IT, library, and repository specialists. Others have only recently begun to be professionalised, such as Data Stewards, and Research Software Engineers. To support their crucial role in FAIR data management, better support is needed with action at institutional level as well as on national and international levels. The further development of these roles requires shared frameworks and terminologies to ensure that research groups and organisations can budget for these roles, describe their contribution, and give them appropriate recognition. Self-assessment tools can help support professionals to work with each other to identify whether their institutional processes are effective in making and keeping data FAIR, and what skills they may need to develop further.

**Recommendation C1:
Manage FAIR support costs and resources**

Supporting researchers in making and keeping their data FAIR requires resources in the form of infrastructure and professional support. *Turning FAIR into Reality* acknowledges the need to cost data management activities and calls for funders to provide financial support (TFIR, rec 18). Most research institutions provide some research support for data management. However, funding for

those roles needs to be sustainable.⁴¹ Institutions need to help research groups to budget for the resources to make their data FAIR. This includes understanding the costs of using external services where appropriate. For instance, it may be more cost-effective to get support on complex data processing or storage from a Research Infrastructure or EOSC service. And where local support is needed, research groups may need guidance on how to budget for it in project proposals. Accurately assessing the viability of data-driven research depends on realistic cost estimates.

- **Researchers should budget explicit expert support from professional service roles in their grant applications.**

To facilitate expert research support for FAIR on a large scale and build recognition for the support contributed to research projects, this support should be more straightforward for research performing organisations (RPOs) to directly cost, so that researchers can budget appropriately for them and ensure that eligible costs are not omitted in grant applications.

- **Research support roles should develop costing models for their expert services that allow researchers to ask for appropriate resources at the start of a research project and ensure long-term FAIRness of any data created.**

As grant applications are complicated already, costing in any expert support for FAIR data management should be made as easy as possible for researchers. Thus, research support roles should provide costing models that allow for easy inclusion of dedicated support in grant applications. Resources to help cost aspects of data management are provided by, for example, the UK Data Service⁴², the Dutch national RDM network LCRDM (building on the UK Data Service checklist)⁴³ and information of the 4C project results.⁴⁴

Research communities	Researchers will benefit from expert support in managing their datasets and digital research outputs and making them FAIR, and so should cost that support in their grant application.
Research support roles	Data stewards and RSEs should provide accurate and justifiable costs for their support services and ask researchers to include these in grant proposals.
Research support actors	Data repositories and institutions should be transparent in how their services are funded to ensure these services can be sustained, based on their usage.
Further work by	FAIRsFAIR will work with data stewards and research software

⁴¹ <https://openworking.wordpress.com/2019/02/20/data-stewardship-goals-for-2019/>

⁴² <https://ukdataservice.ac.uk/media/622368/costingtool.pdf>

⁴³ https://www.lcrdm.nl/files/lcrdm/2019-04/RDM%20and%20CostTable_v20160218.pdf

⁴⁴ 4C (Collaboration to clarify the cost of curation) project results <https://www.4cproject.eu/> (project ended in January 2015)

FAIRsFAIR	engineers on case studies for costing their services that can be used to raise further awareness of the costs required to provide expert support and to keep data FAIR over time.
-----------	---

Recommendation C2:

Develop and implement models for coordinating and supporting data stewards and research software engineers

The TFIR report identified a need to professionalise data science and data stewardship support roles (TFIR rec 10). Research software engineers have already achieved a level of professional cohesion including the formation of a professional body⁴⁵ and an established job title. Data stewardship roles can learn from these achievements and join forces with research software engineers in their institutions to address any upcoming challenges. Institutional adoption of emerging standards such as the CRediT taxonomy⁴⁶ can support the recognition of the broad range of contributions made by researchers and professional services staff.

- **Institutions should develop models for collaboration between research support roles to support the creation and management of FAIR data**

Models and case studies are needed of how institutions can establish teams and coordinate the delivery of support across research faculties and groups. These would build on existing studies (e.g. LCRDM 2019; Higman, et al. 2017), and align with activities of the proposed RDA Interest Group on Professionalising Data Stewardship.

These models and case studies should reference the existing body of practical guidance on FAIR-enabling activity, for examples of how research support roles can work together with researchers in making their data FAIR. For example, the set of *23 Things: Support for Research Data* highlights relevant actions tailored for a variety of audiences,⁴⁷ as does the *Five Recommendations for FAIR software*.⁴⁸ As well as collaborating on services locally, data stewards and research software engineers can make valued contributions to community resources such as *The Turing Way*⁴⁹ guide to reproducible data science.

⁴⁵ <https://society-rse.org/>

⁴⁶ <https://casrai.org/credit/> - The CRediT taxonomy is currently not integrated with popular metadata standards such as the DataCite metadata schema. Progressing this work could be a way to address recommendation A1.

⁴⁷ Jetten, Mijke, Van den Berg, Boudewijn, Kuzak, Mateusz, Witkowska, Iza, Karvovskaya, Lena, Teijeiro Barjas, Carlos, ... Van Horik, René. (2020). *23 Things: Support for Research Data*. Audience-specific versions by the Dutch Community (LCRDM task group RDA/23 Things/RDA Europe grant). Zenodo. <https://doi.org/10.5281/zenodo.3773663>

⁴⁸ <https://fair-software.eu/>

⁴⁹ The Turing Way Community, Becky Arnold, Louise Bowler, Sarah Gibson, Patricia Herterich, Rosie Higman, ... Kirstie Whitaker. (2019, March 25). *The Turing Way: A Handbook for Reproducible Data Science (Version v0.0.4)*. Zenodo. <http://doi.org/10.5281/zenodo.3233986>

- **Data stewards should create formal career pathways and professional bodies for data stewardship roles**

Professional bodies will need to advocate for and articulate the career pathways that institutions need to put in place, to ensure that data stewards roles are sustained and mature. There may be useful lessons to be transferred from the research software engineer role, where coordination of grassroots activities has led to formation of a Charitable Incorporated Organisation - the Society of Research Software Engineers.⁵⁰ This promotes collaboration between researchers and software experts, and supports the creation of an academic career path for Research Software Engineers. The model was also expanded from the UK to start chapters in the US, the Netherlands, Germany, Australia/New Zealand, and the Nordic countries and this may serve as inspiration for data stewards to coalesce in similar ways.

There are signs of growth in the tertiary education sector towards relevant formal qualifications. For example, a network of Danish universities has made the case for postgraduate data steward education in their country⁵¹; and in Germany, a postgraduate programme is available from the Potsdam University of Applied Sciences.⁵² These national efforts through both formal and informal professional development can be brought together by the RDA Interest Group on Professionalising Data Stewardship⁵³ and the GO-FAIR implementation network of Data Stewardship Competence Centres (DSCC-IN) to inspire wider growth.

- **When established, research support staff should consider membership of and participation with representative and membership organisations to help drive forward professionalisation efforts.**

Professional organisations require their members to actively engage, reporting issues and highlighting priority areas for action. Thus, research support staff should join their relevant membership organisation when established to support establishing career paths and help address upcoming challenges. Institutions should support the active engagement of their staff in professional membership organisations.

Research communities

Researchers will benefit from local or disciplinary professional support where possible and should ask for dedicated support if that is not available yet. They can promote the importance of professional contribution e.g. through crediting them on papers using the CRediT Contributor Roles taxonomy.⁵⁴

⁵⁰ <https://society-rse.org/>

⁵¹ Wildgaard, Lorna, Vlachos, Evgenios, Nondal, Lars, Larsen, Asger Væring, & Svendsen, Michael. (2020). National Coordination of Data Steward Education in Denmark: Final report to the National Forum for Research Data Management (DM Forum) (Version 1). Zenodo. <http://doi.org/10.5281/zenodo.3609516>

⁵² <https://www.ddm-master.de/>

⁵³ <https://www.rd-alliance.org/groups/professionalising-data-stewardship>

⁵⁴ <https://casrai.org/credit/>

Research support roles	Data stewards and RSEs will benefit from synergies in providing services together and addressing challenges in career progression together.
Research support actors	<ul style="list-style-type: none"> ● Institutions should facilitate collaboration between data stewards and research software engineers and encourage their involvement in professional bodies. ● Repositories and other service providers should support the use of the CRediT taxonomy to recognise the professional contribution of data stewards and RSEs to all kinds of research outputs.
Further work by FAIRSFAR	FAIRSFAR will provide case studies and guidance for institutions, based on interviews with data stewards, research software engineers and their managers. These will cover FAIR support costs, roles and capabilities, including the use of terminologies or frameworks to describe and recognise contributions to FAIR data management .

Recommendation C3:

Develop and implement terminology for competence centres to annotate and retrieve training materials on enabling FAIR

It is essential that researchers have access to expertise and training materials about FAIR and Open data, and about data services that help to make and keep data FAIR, both within and across domains. A range of relevant training materials is already available, however the coverage is sparse, and information to help find them is fragmented.

EOSC working groups, and others e.g. in the RDA are harmonising descriptive metadata for FAIR data training material. This will allow competence centres (e.g. in institutions and infrastructures) to catalogue their training resources. This will help researchers and data stewards to find these resources, and help ensure appropriate skills are developed. Using standard terminology to identify the target audience and level of difficulty, for example, helps trainees to search across providers and their catalogues. To further assess relevance, trainees also need to know the domain context and what aspects of FAIR data stewardship they cover. Training providers can describe these consistently, according to a standard set of terms for FAIR data stewardship.⁵⁵ Including these terms in metadata for training materials should enable providers to annotate their materials, so that competence centres and training aggregator services can automatically retrieve them and make them findable for the right audience (TFIR, recs 6, 11).

⁵⁵ As per the Terms4FAIRskills terminology in development at <https://terms4fairskills.github.io/>; this project includes contributors from FAIRSFAR WP2, WP3 and WP6 .

- **Training service providers should develop a common terminology that describes the competencies, skills and knowledge associated with making and keeping data FAIR.**

For this, they can build upon the Terms4FAIRskills terminology which is being developed by a diverse group of domain experts.⁵⁶ This is based on work in EOSCpilot proposing a competence framework for FAIR data stewardship in science and scholarship (FAIR4S). Building on that scope, which is described as the activities to ‘make data FAIR and keep it FAIR’, the terms4FAIRskills terminology provides statements defining knowledge, skills and aptitudes. The use cases include facilitating the annotation, discovery and evaluation of FAIR-enabling materials (e.g. training) and resources.⁵⁷

- **Research Infrastructures and training providers should contribute towards the development and implementation of common standards and terminology.**

The emerging consensus on standards for describing training materials will benefit skills development only to the extent that providers use them, and maintain them. FAIR competences terminology is immature, and its development would benefit from wider participation by Research Infrastructures and training providers. This means annotating FAIR-enabling materials with the relevant terms, and providing these to training aggregator services. Guidance materials and training will therefore also be needed for providers to get acquainted with the terminology and apply it consistently. This will reflect the emerging consensus that training materials also need to be FAIR.⁵⁸

Research communities	Researchers will find relevant materials for self-study about FAIR data when competence centres, knowledge portals etc, use a shared terminology for describing the knowledge and skills they can acquire.
Research support roles	Data stewards who coordinate or support training provision should expect to find FAIR-enabling training resources more easily from EOSC catalogues, whether for their own use or to refer others to, complementing any locally available training.
Research support actors	Training service providers and research infrastructures should apply standard metadata and terms to help improve the discovery of FAIR-enabling training materials across their catalogues and repositories.
Further work by FAIRsFAIR	<ul style="list-style-type: none"> ● FAIRsFAIR will contribute to further development of minimum standards for describing training materials, and to

⁵⁶ <https://github.com/terms4fairskills/FAIRterminology>

⁵⁷

⁵⁸ See <https://eoscpilot.eu/content/d75-strategy-sustainable-development-skills-and-capabilities> and <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007854>

	<p>guidance for training providers on implementing them in the EOSC context.</p> <ul style="list-style-type: none"> ● FAIRsFAIR will contribute to further development of the terms4FAIRskills initiative, and facilitate its testing with the FAIRsFAIR competence centre (WP6).
--	--

<p>Recommendation C4: Develop and implement a self-assessment framework for Research Infrastructures, Institutions and other FAIR competence centres</p>	
<p>Research performing organisations and research infrastructures would benefit from help to monitor and plan their actions to enable FAIR. A framework to support this could be developed based on the TFIR action plans, focusing on the extent to which support on FAIR is offered to their data producer communities in delivering FAIR data, or put differently, to assess how FAIR-enabling they are (TFIR, recs 6, 24).</p> <ul style="list-style-type: none"> ● A self-assessment framework may help Research Institutions and Infrastructures to develop a roadmap for developing their FAIR-enabling activities. <p>The Consortium of European Social Science Data Archives (CESSDA)⁵⁹ has carried out a self-assessment of it’s activities against the recommended action plan presented in the TFIR report (2018). Thanks to this, the consortium is able to picture where CESSDA is already performing the relevant actions to enable FAIR, where it plans to do so, and where there is a lack of planning. The DCC RISE model for research data management services is another example of such a framework, offering organisations a set of levels to self-assess whether their capabilities are at a basic, intermediate or advanced level. This should be adapted to the TFIR action plan, focusing on the support with FAIR that Institutions and Infrastructures provide to data producer communities.⁶⁰</p> <ul style="list-style-type: none"> ● Research Infrastructures and Research Performing Organisations should test whether such a framework helps their planning. <p>The Research Infrastructures, Institutions and other Research Performing Organisations partnering in ESFRI cluster and EOSC projects are well- placed to test a self-assessment framework. These projects have networks for exchanging first-hand experiences and for gradually improving a self-assessment framework. Competence centres (see also recommendation C3) can be useful partners in this.</p>	
<p>Research communities</p>	<p>Research communities will benefit from research infrastructures that have a clear notion of the FAIR support that they offer, and</p>

⁵⁹ R. Dekker: Social Data: CESSDA Best Practices. In: Data Intelligence, Volume 2 | Issue 1-2 | Winter-Spring 2020, p.220-229. https://doi.org/10.1162/dint_a_00044

⁶⁰ <https://www.dcc.ac.uk/guidance/how-guides/RISE>

	about any aspects they can improve on.
Research support roles	Data stewards should be proactive in using self-assessment frameworks to bring together research support colleagues and the communities they work with, and plan better provision of support to prepare FAIR outputs across the research data lifecycle.
Research support actors	Research infrastructures and research performing organisations can benefit by self-assessing the concrete steps they are taking towards enabling their communities to meet the FAIR data principles.
Further work by FAIRSFAR	FAIRSFAR will liaise with RIs, Institutions and competence centres to co-design a framework, based on DCC RISE, and will test its utility for planning support in implementing the TFIR recommendations. Coordination is also needed to ensure alignment with work on maturity models for FAIR-enabling repositories (see the relevant FAIRSFAR recommendation). ⁶¹

Theme D: Ensure trusted curation of data

Trusted digital repositories (TDRs) exist to provide a reliable and effective means of curating (i.e. preserving and providing appropriate access to) research data. TDRs can provide services for secure storage and managed access to all types of data, including sensitive data. Researchers benefit from familiarity with the repository landscape as it allows them to deposit data appropriately, and search for data effectively.

Recommendation D1:

Develop and implement guidance and support for selection of appropriate trusted digital repositories (TDRs)

According to *Turning FAIR into Reality* (TFIR, rec. 20), “Research data should be made available by means of Trusted Digital Repositories, and where possible in those with a mission and expertise to support a specific discipline or interdisciplinary research community” (European Commission 2018, p. 72).

- **Researchers should be supported in the deposit of data in Trusted Digital Repositories.**

In order to keep data FAIR over the long term, researchers should deposit datasets in a trusted digital repository (TDR), where available. A TDR is a data archive that is certified against a set of commonly agreed criteria such as CoreTrustSeal (CTS). The certification status of a repository will give the researcher confidence that data will be archived in a sustainable way and will also take data protection issues into consideration.

⁶¹ H. L'Hours, I. von Stein et al. (2020): Repository Certification Mechanism: a Recommendation on the Extended Requirements and Procedures. <https://doi.org/10.5281/zenodo.3835698>

Using certified repository services is already recommended in data management policies for several research infrastructures in a wide range of disciplines. In outreach, training and instruction activities, researchers can be encouraged to become aware of and use trusted repositories. For example, the data management training provided by the CESSDA research infrastructure for social science research⁶² includes this encouragement; as do the repository certification activities provided by the SSHOC cluster project, the first of which is addressing service providers.⁶³ Convincing researchers to share data by using TDRs should be supported by leaflets, short video clips, social media posts, and other easily digestible advocacy resources that contain encouragement and guidance for researchers in the use of a TDR.⁶⁴ Guidance on completing Data Management Plans should cover how to deposit data in a TDR where there are applicable examples in the relevant domain.

The depositor - whether researcher or data steward - should be aware of TDR requirements related to the characteristics of the data objects deposited. These requirements relate to the features of the data objects to be deposited, such as the data format, and the documentation of the data objects. The depositor should apply common community standards identified in Data Management Planning guidance.⁶⁵ It is good practice to consider specific requirements associated with the data and/or repository selected for deposit as part of data management planning.

- **Research communities should adopt common agreements on which TDR or TDRs are most suitable to be used in their community.**

Many types of TDR exist. These can be domain specific, operated by individual research institutes, or provided at the national level. The Re3data directory⁶⁶ contains over 2500 repositories and allows potential depositors and data users to search for appropriate data repositories, including those certified against CoreTrustSeal and other certification schemes. FAIRSFAR has also developed a filter in the Repository Finder tool to support searching for repositories that support FAIR.⁶⁷ OpenAIRE also offers researcher-friendly guidance on data repository selection for depositing research data.⁶⁸

Journals and publishers are also working with research communities they serve to highlight relevant TDRs, or recommend suitable repositories where there is a lack of certified TDRs in the relevant domain. The FAIRsharing.org service is a useful source for finding such recommendations, and worked with relevant RDA groups to coordinate the criteria used by

⁶² See: <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide>

⁶³ <https://doi.org/10.5281/zenodo.3725868>

⁶⁴ See e.g. "Sharing data: good for science, good for you - 2013 (<https://www.youtube.com/watch?v=HJbo-OAaJ1I>)

⁶⁵ E.g. the metadata formats and data formats supported by the TDR. See e.g.

https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/before-depositing/file-formats?set_language=en

⁶⁶ <https://www.re3data.org/search>

⁶⁷ <https://repositoryfinder.datacite.org/>

⁶⁸ <https://www.openaire.eu/find-trustworthy-data-repository>

journals and publishers. ⁶⁹	
Research communities	Researchers will benefit from selection of an appropriate TDR for their specific usage. Research communities should decide on repository choice together, after consulting existing guidance resources and in consultation with their data steward.
Research support roles	Data stewards will benefit from having state-of-the-art knowledge on the TDR landscape so they can provide appropriate advice to researchers. This guidance can be referred to during DMP creation.
Research support actors	<ul style="list-style-type: none"> • Providers of TDRs should proactively define and communicate domain-specific FAIR data management needs to other relevant service providers. • Journals and publishers should ensure that their repository recommendations keep pace with the progress of repository certification in their relevant domain.
Further work by FAIRSFAR	FAIRSFAR will provide guidance for the selection of suitable TDRs to manage the research data. For this, registry services such as the Re3data registry and the Repository Finder tool will be used.

Recommendation D2:

Develop and implement guidance and support for making sensitive data FAIR for reuse

According to *Turning FAIR into Reality* (TFiR, rec. 17) “Policies should be aligned and consolidated to ensure that publicly-funded research data are made FAIR and Open, except for legitimate restrictions. The maxim ‘as Open as possible, as closed as necessary’ should be applied proportionately with genuine best efforts to share” (European Commission 2018, p. 70). “Concrete and accessible guidance should be provided to researchers to find the optimal balance between sharing whilst also safeguarding privacy” (European Commission 2018, p. 71).

Sensitive data is information that must be protected against unauthorized access. The General Data Protection Regulation (GDPR) and other legal regulations such as intellectual property rights (IPR) must be taken into consideration when making sensitive data FAIR and usable for authorized users. The GDPR concerns data protection and privacy in the EU and European Economic Area. It contains provisions and requirements related to the processing of personal data of individuals. Creative Commons (CC) licences are among the public copyright licenses that allow conditional use of copyrighted work.⁷⁰

⁶⁹ <https://fairsharing.org/communities>

⁷⁰ <https://creativecommons.org/>

- **Researchers and data stewards should work within GDPR and IPR regulations to make data FAIR and “as open as possible, as closed as necessary”⁷¹**

Stakeholders must be informed on details of GDPR and IPR issues in order to be aware of how the right balance can be achieved between making data openly FAIR, or FAIR with restricted access. This guidance can have several forms and starts with informing researchers on how GDPR regulations influence their research in practice. Data stewards can advise, e.g. on pseudonymisation or full anonymization of datasets or controlled access. They also can provide an overview of available data license models and in which situations they are applicable. OpenAIRE offers research-friendly guidance on how to deal with sensitive data and how to preserve it safely.⁷² OpenAIRE also published a checklist on how to license repositories that data stewards should use to assess IPR issues in relation to data repositories.⁷³

OpenAIRE provides a toolkit for researchers on legal issues⁷⁴, which presents a set of user FAQs for researchers on copyright, open science and data. This offers an up-to-date, legally advanced but still manageable set of rules, guidelines, and resources to enable the full potential of open science in the EU research field with a view to addressing copyright and related rights issues. Researchers can find useful guidance on how to license research data⁷⁵, how to reuse research data⁷⁶, and about protection of research data by intellectual property rights.⁷⁷

Besides resources that inform researchers on GDPR and IPR issues, services and tools should also be used to classify or process datasets. A relevant initiative in relation to the sharing and use of sensitive data by researchers in a standardized and responsible way is the DataTags project that standardizes the classification of protection levels of data sets.⁷⁸ DataTags or labels represent human-readable and machine actionable data policies and are based on applicable laws, contracts and best practices. When using dataset tagging services the applicable context must be taken into consideration, such as national regulations or specific policies. Data stewards or research communities should be aware of this context and advise researchers accordingly.

Other types of services in relation to GDPR and IPR issues concerns tools to pseudomize or anonymize data sets⁷⁹ and tools to help to select appropriate licenses that can be connected to digital research objects and outputs.⁸⁰

Research communities	Researchers should get acquainted with state-of-the-art information on GDPR and IPR, and informed on related tools and services.
-----------------------------	--

⁷¹ TFIR p. 15

⁷² <https://www.openaire.eu/sensitive-data-guide>

⁷³ <https://www.openaire.eu/making-your-repository-open>

⁷⁴ <https://zenodo.org/record/2574619#.Xs6BqGhKiUI>

⁷⁵ <https://www.openaire.eu/how-do-i-license-my-research-data>

⁷⁶ <https://www.openaire.eu/can-i-reuse-someone-else-research-data>

⁷⁷ <https://www.openaire.eu/how-do-i-know-if-my-research-data-is-protected>

⁷⁸ <http://datatags.org/>

⁷⁹ E.g. the Amnesia anonymization tool, see: <https://amnesia.openaire.eu/>

⁸⁰ E.g. the Creative Commons License Chooser: <https://creativecommons.org/choose/>

Research support roles	Data stewards should get informed on information, tools and services relevant for processing (sensitive) data (e.g. data tags).
Further work by FAIRSF AIR	FAIRSF AIR will provide guidance on how to make sensitive data FAIR (taking into consideration the applicable GDPR and IPR regulations) synthesising work in this area e.g. from SSHOC, FAIR4Health, FAIRplus, and EOSC-life, and collaborating where appropriate with the proposed RDA working group on Guidelines for Health Research Performing Organizations to implement a FAIR Data Policy.

3. Next steps

These recommendations are released as a living document with the potential to be refined to reflect the forthcoming work of other projects funded under the INFRAEOSC-05-2018-2019 call, and other relevant initiatives to the extent that they contribute to addressing the practical steps recommended here.

FAIRSF AIR will work collaboratively to provide practical support to a range of communities – including research communities, data stewards, research software engineers, and the infrastructures and institutions with which they operate – to support practical steps towards making data FAIR in their everyday research activities. We will develop practical guidance, templates and training resources to support stakeholders in implementing these recommendations for practice. Wherever possible, we will seek to work closely with other initiatives active in the FAIR data landscape to develop and implement the recommendations. We also seek to collaborate with the working groups of the Research Data Alliance and the EOSC Governance. We aim to work in similar directions for our respective audiences, and we hope this practical collaboration continues to engender action towards more production of FAIR data.

Annex 1. Selected Recommendations from Turning FAIR into Reality

The recommendations and action points below are the main basis for this report, and are drawn from the EC FAIR Data Expert Group report, Turning FAIR into Reality (European Commission 2018).

Theme A: Develop and implement data sharing and interoperability frameworks

- Develop interoperability frameworks (TFIR recommendation 4)
 - Enabling mechanisms must be funded and implemented to support research communities to develop and maintain their disciplinary interoperability frameworks (action 4.1)
 - Examples of FAIR use cases and success stories should be developed to convince reluctant research communities of the benefits in defining their disciplinary interoperability framework. (action 4.2)
 - Disciplines and interdisciplinary research programmes should be encouraged to engage with international collaboration mechanisms to develop interoperability frameworks. Common standards, intelligent crosswalks, brokering mechanisms and semantic technologies should all be explored to break down silos between communities and support interdisciplinary research.(action 4.3)
 - Mechanisms should be facilitated to promote the exchange of good practices and lessons learned in relation to the implementation of FAIR practices both within and across disciplines. Case studies for cross-disciplinary data sharing and reuse should also be collected, shared and used as a basis for the development of good practice.(action 4.4)
 - The components of the FAIR ecosystem should adhere to common standards to support disciplinary frameworks and to promote interoperability and reuse of data across disciplines.(action 4.5)
- Support semantic technologies (TFIR recommendation 7)
 - To achieve interoperability between repositories and registries, common protocols should be developed that are independent of the data organisation and structure of various services. (action 7.2).

Theme B: Ensure data management is supported by data management plans (DMPs)

- Ensure data management via DMPs (TFIR recommendation 5)
 - Research communities must be required, supported and incentivised to consider data management and appropriate data sharing as a core part of all research activities. They should establish a Data Management Plan at project outset to consider the approach for creating, managing and sharing all research outputs (data, code, models, samples etc.) (action 5.1)
 - Data Management Plans should be living documents that are implemented throughout the project. A lightweight data management and curation statement should be assessed at project proposal stage, including information on costs and the track record in FAIR. A

sufficiently detailed DMP should be developed at project inception. Project end reports should include reporting against the DMP. (action 5.2)

- Data Management Plans should be tailored to disciplinary needs to ensure that they become a useful tool for projects. Research communities should be inspired and empowered to provide input to the disciplinary aspects of DMPs and thereby to agree model approaches, exemplars and rubrics that help to embed FAIR data practices in different settings.(action 5.3)
- Use information held in DMPs (TFIR recommendation 22)
 - DMPs themselves should conform to FAIR principles and be Open where possible. (action 22.4)
 - Information gathered from the process of implementing and evaluating DMPs relating to conformity, challenges and good practices should be used to improve practice. (action 22.5)

Theme C: Develop professional support for FAIR data

- Recognise and reward FAIR data & stewardship (TFIR recommendation 6)
 - Credit should be given for all roles supporting FAIR data, including data analysis, annotation, management and curation, as well as for participation in the definition of interoperability frameworks, whether contributing to existing resources or developing new (action 6.2).
 - Evidence of past practice in support of FAIR data should be included in assessments of research contribution. Such evidence should be required in grant proposals (for both research and infrastructure investments), among hiring criteria, for career advancement and other areas where evaluation of research contribution has a legitimate role to play. This should include assessment of graduate students (action 6.3).
 - Contributions to the development and operation of certified and trusted infrastructures that support FAIR data should be recognised, rewarded and appropriately incentivised in a sustainable way (action 6.4).
- Professionalise data science & data stewardship roles and train researchers (TFIR recommendation 10)
 - Key data roles need to be recognised and rewarded, in particular, the data scientists who will assist research design and data analysis, visualisation and modelling; and data stewards who will inform the process of data curation and take responsibility for data management (action 10.1).
 - Formal career pathways must be implemented to demonstrate the value of these roles and retain such professionalised roles in support of research teams (action 10.2).
 - Professional bodies for these roles should be created, consolidated when they exist, and promoted. Accreditation should be developed for training and qualifications for these roles (action 10.3).

- Data skills, including an appropriate foundational level in data science and data stewardship, should be included in undergraduate and postgraduate training across disciplines, and in the provision of continuing professional development (CPD) credits for researchers (action 10.4).
- Implement curriculum frameworks and training (TFIR recommendation 11)
 - Sharing and reuse of Open Educational Resources and reusable materials for data science and data stewardship programmes should be encouraged and facilitated (action 11.2).
- Cost data management (TFIR recommendation 18)
 - Research institutions and research projects need to take data management seriously and provide sufficient resources to implement the actions required in DMPs, while ensuring that financial resources are written into proposals as eligible costs (action 18.2).

Theme D: Ensure trusted curation of data

- Align and harmonise FAIR and Open data policy (TFIR recommendation 17)
 - Concrete and accessible guidance should be provided to researchers to find the optimal balance between sharing whilst also safeguarding privacy. There are many exemplars of good practice in providing managed access to sensitive data on which researchers can draw (action 17.8)
- Deposit in Trusted Digital Repositories (TFIR recommendation 20)
 - 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 (action 20.3).
 - Outreach is required via scholarly societies, scientific unions and domain conferences so researchers in each field are aware of the relevant disciplinary repositories (action 20.4).

Bibliography

- Davidson, J., Grootveld, M., Whyte, A., Herterich, P., Engelhardt, C., Stoy, L., Proudman, V. (2020). D3.3 Policy Enhancement Recommendations (Version 1.0). <https://doi.org/10.5281/zenodo.3686900>
- European Commission (2018). Turning FAIR Into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data. <https://op.europa.eu/s/n1qc>
- FAIRsFAIR (2020). 'About'. Retrieved from <https://www.fairsfair.eu/>
- EOSC-Nordic (2020). FAIR Recommended Actions Based on FAIR Maturity. Retrieved from https://www.eosc-nordic.eu/content/uploads/2020/04/EOSC-Nordic_WP4_-_FAIR_Recommended_actions_based_on_FAIR_Maturity_Evaluation.pdf
- GO-FAIR (2017) FAIR Principles. Retrieved from <https://www.go-fair.org/fair-principles/>
- Higman, R., Teperek, M., Kingsley, D. (2017). Creating a community of data champions. *BioRxiv*, 104661.
- LCRDM (2019). Data stewardship on the map: a study of tasks and roles in Dutch research institutes. https://www.lcrdm.nl/files/lcrdm/2019-05/LCRDM%20rapport%20datastewardship_EN_online.pdf
- Le Franc, Y., Parland-von Essen, J., Bonino, L., Lehvälaiho, H., Coen, G., Staiger, C. (2020). D2.2 FAIR Semantics: First recommendations (Version 1.0). FAIRsFAIR. Retrieved from <https://zenodo.org/record/3707985#.XoYgCpMzYWo>
- Marttila, J. et al (2020). Towards FAIR practice in EOSC. Blogpost dated 2020-04-21. Retrieved from <https://www.eoscsecretariat.eu/news-opinion/towards-fair-practice-eosc>
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Whyte, A., Engelhart, C., Bangert, D., Kayumbi-Kabeya, G., Lambert, S., Thorley, M., O'Connor, R., Herterich, P., Davidson, J. (2019). D3.2 FAIR Data Practice Analysis (Version 1.0) <https://doi.org/10.5281/zenodo.3581352>