FAIRSFAIR
Fostering Fair Data Practices in Europe

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

D3.2 FAIR Data Practice Analysis

Work Package	WP3, FAIR Data Policy and Practice
Lead Author (Org)	DCC
Contributing Author(s) (Org)	DCC, STFC, EUA, UGOE
Due Date	30.11.2019
Date	19.12.2019
Version	1.0
DOI	https://doi.org/10.5281/zenodo.3581352

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

☐ RE: Restricted to a group specified by the consortium (including the Commission)

☐ CO: Confidential, only for members of the consortium (including the Commission)

Abstract

This document provides an analysis of practices to support FAIR data production within a broad selection of research disciplines and research data repositories. It aims to inform the priorities of stakeholders interested in embedding those practices in research communities. Those stakeholders include policy makers, data librarians and others providing data services to research communities, as well as champions of FAIR principles in those communities. It also identifies priority themes for initial work in FAIRsFAIR to support ESFRI cluster and EOSC projects in FAIR culture change. These include developing a self-assessment framework for research infrastructures and institutions on their progress to support FAIR enabling practices in the communities they serve. This will underpin further work to build capabilities, describe good practice and address the highly uneven awareness of FAIR principles and the lack of information on research community implementation.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Versioning and contribution history

Version	Date	Authors	Notes
0.1	21.09.2019	Angus Whyte (DCC)	ToC
0.2	14.10.2019	Joy Davidson, Angus Whyte (DCC)	Initial content sections 1- 3
0.3	07.11.2019	Joy Davidson, Angus Whyte (DCC) Claudia Engelhart, Daniel Bangert (UGOE)	Further content sections 1-3
0.4	21.11.2019	Gabin Kayumbi-Kabeya, Simon Lambert (STFC), Patricia Herterich, Joy Davidson, Angus Whyte (DCC)	Further contributions to section 2- 3
0.5	02.12.2019	Claudia Engelhart, Daniel Bangert (UGOE) Gabin Kayumbi-Kabeya, Simon Lambert, Mark Thorley (STFC), Ryan O'Connor, Patricia Herterich, Joy Davidson, Angus Whyte (DCC)	Draft for internal review including section 4
1.0	13.12.2019	Angus Whyte (DCC)	Edits following internal review, addition of annex, abstract

Disclaimer

FAIRsFAIR 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.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Abbreviations and Acronyms

FAIR	Findable, Accessible, Interoperable, Reusable
EOSC	European Open Science Cloud
ESFRI	European Strategy Forum on Research Infrastructures
HLAC	High Level Advisory Committee
EGFC	European Group of FAIR Champions
HEIs	Higher Education Institutions
CORDIS	Community Research and Development Information Service
DSM	Digital Single Market
EOSC EB	EOSC Executive Board
RDMF	Research Data Management Forum
IDCC	International Digital Curation Conference
ENVRI-FAIR	ENVironmental Research Infrastructures building Fair services Accessible for society, Innovation and Research
ESCAPE	European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures
PANOSC	Photon and Neutron Open Science Cloud
SSHOC	Social Sciences & Humanities Open Cloud
RDA	Research Data Alliance
BBMRI	Biobanking and BioMolecular resources Research Infrastructure
CESSDA	Consortium of European Social Science Data Archives
CLARIN	Common Language Resources and Technology Infrastructure
DARIAH	Digital Research Infrastructure for the Social Sciences and Humanities
EPOS	European Plate Observing System
E-RIHS	European Research Infrastructure for Heritage Science

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

ESS	European Social Survey
ICOS	Integrated Carbon Observation System
IS-ENES	Infrastructure for the European Network of Earth System Modelling
LTER	Long Term Ecological Research
SeaDataNet	Pan-European Infrastructure for Ocean and Marine Data Management
SHARE	Survey of Health, Ageing and Retirement in Europe
DMP	Data Management Plan

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Executive summary

The overall objective of FAIRsFAIR is to accelerate the realisation of the goals of the EOSC by opening up and sharing all knowledge, expertise, guidelines, implementations, new trajectories, courses and education on FAIR matters. It seeks to establish a level playing field for all European member states (and beyond) when it comes to contributing data to scientific and scholarly communities and reusing data from scientists and scholars elsewhere. All this is made possible by the coordinated effort of twenty-two partners spanning eight member states that are working together to define guidelines towards a FAIR approach to data and service management for data repositories across disciplines.

This document provides an analysis of practices to support FAIR data production within a broad selection of research disciplines and research data repositories. It aims to inform the priorities of all stakeholders interested in embedding those practices in research communities. Those stakeholders include policy makers, data librarians and others providing data services to research communities, as well as community champions of FAIR principles. It also identifies priority themes for initial work in FAIRsFAIR to support ESFRI cluster and EOSC projects in FAIR culture change.

The analysis begins by reviewing current literature on disciplinary variation in data practices It then looks at specific examples of good practice in addressing challenges to FAIR implementation. By highlighting these, the analysis offers a basis for subsequent work in FAIRsFAIR tasks T3.2 and T3.3 to engage with their practices and support other communities to increase their production of FAIR data.

To frame the analysis, the report uses the implementation recommendations of the EC FAIR Expert Group set out in the report *Turning FAIR data into reality*. By examining how its action points for communities, service providers and data stewards are being tackled, the analysis forms a basis for practical action by these stakeholder groups. The report considers how further work in FAIRsFAIR to promote FAIR culture should define 'disciplinary practice' and descriptions of 'good' practice.

As Tasks 3.2 - 3.4 progress, FAIRsFAIR will collaborate with other projects and communities to address challenges to FAIR implementation. An inventory of FAIR data practice exemplars will be provided based on the following;-

- Engagement activity with community stakeholders to harmonise relevant policies and implement recommendations for particular standards, registries, or repositories.
- Outreach activity with relevant repositories, data facilities and their user communities to facilitate self-assessment of their activities against the *Turning FAIR* recommendations, e.g. on adoption of standards for data management.
- Guidance on providing machine-actionable statements in Data Management Plans or other operational -level policy documents that inform downstream data stewardship actions.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Table of Contents

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

1. Aims, scope and methodology

1.1. Aims and introduction

This report aims to brief policy makers, institutions, data stewards, and service providers on the status of FAIR data stewardship practice among research communities. The focus is on identifying current practices to support data production, curation and reuse that influence adoption of FAIR data principles by researchers, their communities, repositories and other service providers.

The sources for the report were desk-based literature survey, a small number of targeted interviews, workshop discussion and survey results. These are each described under 'methodology' later in this section.

The analysis has three dimensions, illustrated in Figure 1.1; *data practices* that enable FAIR implementation, *changes* to practices the EC FAIR Expert Group recommended to support a change towards FAIR culture, and the research *communities* producing FAIR data and related outputs.

Figure 1.1 Landscape analysis dimensions

1.2. Changes in practice towards FAIR implementation

The changes that are needed to implement FAIR data stewardship principles are informed by *Turning FAIR into Reality* the report of the EC Expert Group on FAIR (European Commission, 2018). That report (referred to as *Turning FAIR* throughout this deliverable) gives recommendations for the conceptualisation of FAIR, implementation, and sustaining of FAIR practices. Those recommendations and action points relating to FAIR implementation represent the main parameters for the analysis, especially those relating to changes in culture. Figure 1.2 below summarises these.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

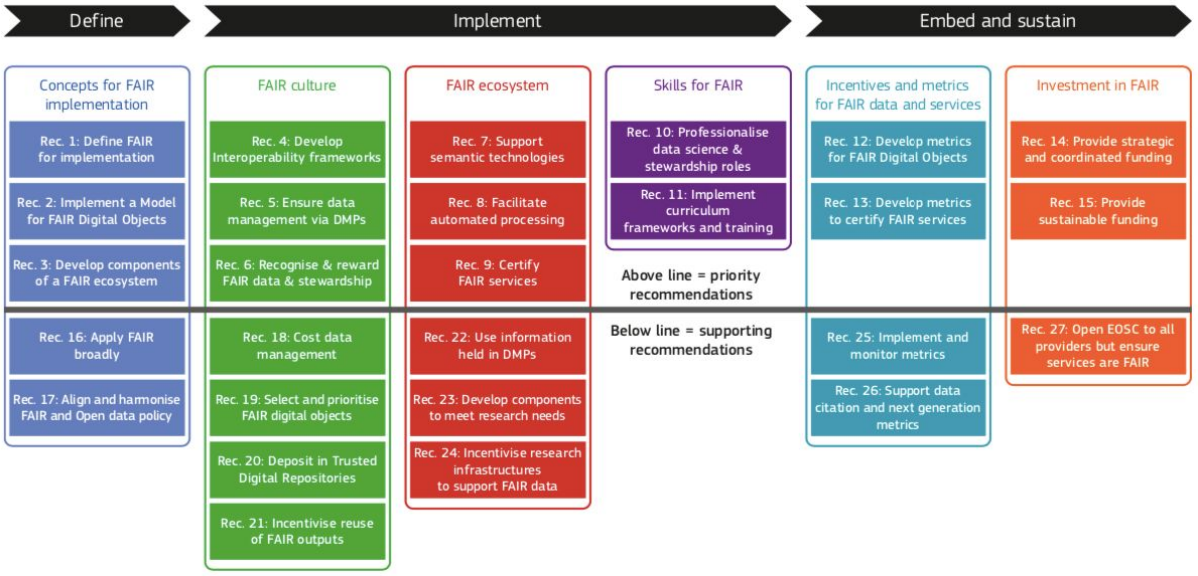

Figure 1.2. Index to FAIR Action Plan recommendations (source EC Expert Group on FAIR data)

The *Turning FAIR* recommendations and action points are described in section 8.4 of that report. Those dealing with FAIR culture (shown in green in Figure 3) are relevant for this report, especially those directed at research communities and service providers. Please note some actions are excluded where they are directed towards other stakeholders. This report focuses on actions requiring collaboration between research communities and other stakeholders, especially institutions, data stewards, and other service providers.

Other FAIRsFAIR reports deal with *Turning FAIR* recommendations on the conceptualisation and sustainability of FAIR policy outcomes (D3.1), the FAIR ecosystem including semantic technologies (D2.1), metrics for FAIR digital objects and service certification (D4.1), as well as recommendations about curriculum frameworks for FAIR (D6.1 and D7.1).

The practice analysis is based on the 9 recommendations relevant to the scope of the report, primarily those on FAIR culture.

These are as follows:

- Interoperability frameworks and metadata
  - Develop interoperability frameworks (recommendation 4)
  - Facilitate automated processing (recommendation 8)
- Data management plan support and usage
  - Ensure data management via DMPs (recommendation 5)
  - Use information held in DMPs (recommendation 22)
- Managing costs, incentives and rewards
  - Cost data management (recommendation 18)

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

    - Encourage and incentivise reuse of FAIR outputs (recommendation 21)
    - Recognise and reward FAIR data & stewardship (recommendation 6)
- Selecting and depositing FAIR outputs
    - Select and prioritise FAIR digital objects (recommendation 19)
    - Deposit in Trusted Digital Repositories (recommendation 20)

Section 3 of the report considers each of the recommendations and the 21 action points that relate to them.

1.3 Communities

To characterise research communities the analysis adopts the disciplinary groupings used for the ESFRI roadmap[1], and associated cluster projects (funded under the European Commission's INFRAEOSC- 04 call). These are as follows:

- Physical science and engineering
- Energy
- Environment
- Health and food
- Social and cultural innovation
- Data computing and digital research infrastructures

The practice analysis refers to these headings on the basis that they are useful to ESFRI projects and clusters. Several alternative classifications are available to describe research disciplines, including, e.g. the Re3data schema, Narcis, US Library of Congress, and Australian and New Zealand Standard Research Classification.

For our work, it is crucial to identify practices connecting researchers working in one or more research domain with the Infrastructures, Institutions and other service providers interested in FAIR implementation. It is less critical which classification system is used to describe research domains, as we are not attempting in this analysis to provide an exhaustive or representative sample of the many research sub-domains that fall within these very high-level domain categories.

Rather than viewing domains as 'silos' of practice, the analysis treats the ESFRI categories and disciplinary categories more generally as one of the various factors influencing practice. As the FAIRsFAIR project is working across domains, and aims to find and promote cross-domain synergies, it is useful to apply other lenses to look at data practices and the

[1] ESFRI Roadmap 2018 Strategy Report. http://roadmap2018.esfri.eu/media/1048/rm2018-part1-20.pdf

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

communities that may form around them. These may include, for example, communities sharing data management challenges (e.g. handling of sensitive data on human subjects), data centres associated with physical facilities (e.g. synchrotrons) or cross-disciplinary research challenges (e.g. heritage science). Section 2 considers the unit of analysis for 'communities' in more depth.

1.4 Data practices

To characterise FAIR data enabling practices, further work leading from this analysis uses FAIR4S, a framework developed in the EOSCpilot project to describe competencies for FAIR data stewardship (Whyte et al, 2019). FAIR data enabling practices can be seen as the activities that apply these competencies, which are described as those needed "to make data FAIR and keep it FAIR". Doing so will help link practical support offered to repositories and communities with the project's training and professional development activity, in Task 3.3.

The FAIR4S framework has been used in the terminology development initiative terms4FAIRskills, which is utilised in two FAIRsFAIR work packages that have skills improvement as their main focus – WP6 (FAIR Competence Centre) and WP7 (FAIR Data Science and Professionalisation). These work packages will support the professionalisation of data science and stewardship, while WP3 analysis focuses on actions stakeholders are already taking towards that professionalisation and any relevant disciplinary factors.

The framework comprises 6 data management activities forming a project-level data lifecycle (plan and design, collect and process, integrate and analyse, appraise and preserve, publish and release, expose and discover). In addition, the framework includes 3 areas of organisational practice to sustain the production of FAIR outputs (govern and assess, scope and resource, advise and enable). These are illustrated in Figure 1.3 and described in Table 1.1 (both sourced from Whyte et al (ibid.)

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Figure 1.3. FAIR4S data stewardship framework

Table 1.1 Data stewardship activities enabling FAIR data

Plan and design: Planning and design of data, research software and other outputs, including the associated documentation. This will include all relevant steps including identifying requirements of research output users, the organisation and research funders, establishing effective approaches to meet their requirements, then reviewing this planning.
Capture and process: Capturing and processing of data or related materials to enable research evidence to be prepared for analysis; provisioning of secure managed access to networked storage, scalable to meet demands, plus resources, tools, standards and workflows for collaboration between research team members, and relevant third parties.
Integrate and analyse: Developing and applying appropriate methods to enable lines of enquiry to be formulated and pursued towards the research objectives, by assembling and integrating selected data, software, systems, or other resources, and enabling relevant knowledge and techniques to be applied in their analysis and transformation into research outputs.
Appraise and preserve: Developing and applying appropriate methods to appraise research outputs for their compliance with ethical, FAIR and research integrity principles, their value to the organisation and to research reproducibility, and their potential to serve new purposes or communities; planning and taking action to mitigate risks to long-term access for further appraisal.
Publish and release: Describing research products and their inter-relationships, providing access to meet the needs of their providers, users, and other stakeholders, in order to maintain or enhance their value and comply with ethical, FAIR and research integrity principles and policies.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Expose and discover: Ensure that processes and mechanisms for providing access to research products and their inter-relationships follow technology developments, community standards, and good practices for compliance with ethical, FAIR and research integrity principles.

Govern and assess: Developing and maintaining the organisation's strategies, policies, and processes on FAIR/open research outputs, and associated documents and processes that enable these to be implemented, and relevant laws or regulations to be complied with. Continually reviewing these strategies, policies and processes through stakeholder consultation, communication, and impact monitoring.
Scope and resource: Identifying the scope of research data services and stewardship activities and securing the resources to sustain these. Continually reviewing the business case considering the service value propositions, processes, and relevant costs and benefits, taking into account governance processes and timelines, and the need for cost recovery mechanisms to comply with funder requirements.
Advise and enable: Providing training, advice and support for data stewardship and open research, including the online or face-to-face training and mentoring that service customers need to make effective use of them. This will include interacting with relevant professional service units, building and maintaining stakeholder relationships.

1.5 Methodology

The analysis set out to look for examples of FAIR data initiatives relevant to each ESFRI discipline, but open to cross-disciplinary communities of practice that may be a focus of subsequent work in FAIRsFAIR tasks 3.2, 3.3. and 3.4. These aim to identify opportunities to promote policy harmonisation (T3.2), measures to embed changes in practice (T3.3) and support repositories to increase the supply of FAIR data (T3.4). The criteria used to select exemplars of good practice were drafted in a milestone report (3.2) as follows:

I. Involve repositories or data centres that are collaborating with EOSC projects or national funders to support FAIR policy development and implementation.

II. Relevant to one or more of the EFSRI clusters and disciplinary groups.

III. Involve recent development of policy/ practice that additional FAIRsFAIR support is likely to increase production and use of FAIR data over the 2 year period from M6-M30. That support could take the form of:

- Working with stakeholders to support their harmonisation of policies, through drafting and consulting on policy texts, or monitoring policy uptake.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- Drafting policy-related guidance materials on standards or related resources, e.g. registries, repositories to help promote their uptake in the relevant communities
- Supporting implementation of machine-readable policy resources, e.g. data management plans, to assist FAIR data production in downstream workflows

An important consideration is that results should facilitate the exchange of good practice in FAIR implementation. For that purpose, the project team consulted with ESFRI clusters to identify specific communities of practice likely to offer lessons to others. The initial engagement will be followed up in tasks 3.2, 3.3 and 3.4, and further good practice examples will be identified through participation in EOSC events, discussion with EOSC projects, and responses to open calls.

The selection of 'good practices' in this report have been identified through an approach combining the following:

I. Desk research on current literature and related projects.

II. Interviews with ESFRI cluster project contacts to nominate communities.

III. Survey of Research Infrastructures and Institutions, primarily those participating in ESFRI cluster projects, and INFRAEOSC-5 projects, on their perceptions of FAIR policy and practice in communities they support.

IV. Internal liaison with FAIRsFAIR WP2, WP4, WP6 and WP7, and on cross-project synchronisation (via WP5).

1.5.1. Interviews

Interviews were conducted with representatives of ESFRI 'cluster projects' and related EOSC projects. These projects run roughly parallel with FAIRsFAIR and involve consortia that comprise Research Infrastructures and other partners, each with a focus on broad scientific disciplines. The interviews were carried out through online calls in September-October 2019. Each interview was summarised, and these summaries were then checked with interviewees for accuracy.

1.5.2. Open Consultation

The FAIRsFAIR Policy and Practice Consultation open consultation was carried out online over approx. 6 weeks (August - September 2019). The consultation targeted members of the research support community in research infrastructures and institutions, via ESFRI cluster contacts and relevant email lists, aiming to consult respondents on their views and experiences about implementing the FAIR principles.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

The open consultation questions were grouped under five broad themes: 1) practice 2) policy 3) repositories 4) skills and 5) competence centres. The 'practice' and 'repositories' questions are relevant to this report. Responses to questions on policy are analysed in the related FAIRsFAIR report D3.1 (FAIR Policy Analysis) and those on skills and competence centres in D6.1 (Competence Centre Requirements).

Responses to the open consultation and interviews are analysed in sections 2 and 3.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

2. Understanding FAIR data practice

2.1. Introduction

This section begins in section 2.2 with an overview of the FAIRsFAIR Policy and Practice interview and survey-based consultation with Research Infrastructure and Institutional providers of support on FAIR to researchers.

Section 2.3 then reviews the literature on factors influencing research culture and initiatives to enable FAIR data production. Some of these factors and initiatives are strongly disciplinary, especially those enabling interoperability and reusability. These correspond respectively to the *Turning FAIR* action points on interoperability frameworks and development of trustworthy repositories.

By contrast, initiatives relevant to the FAIR culture change recommendations on data management planning and costs, benefits and incentives for FAIR data management tend to be undertaken by funders and institutional networks. Although trans-disciplinary in scope, take-up is more evident in some disciplines than others. Institutional-level support is also influential, and the section highlights recurring themes in recent studies that indicate the points in the data lifecycle where support arrangements are critical to downstream production of FAIR data.

Section 2.4 gives further consideration to research disciplines as a unit of analysis. Studies of research data management practices consistently show that these vary at a finer-grained level than the discipline. They are shaped by a variety of factors that engender formal, standardised approaches such as high costs of data collection, extensive specialisation of research tasks, strong methodological consensus, and use of data from human subjects. Communities that share common data practices may also form around data types and formats. The concepts of ‘data community’ and ‘repertoire’ are proposed as a more useful way to discuss FAIR practice than discipline alone.

2.2 Current practice and gaps - consultation responses

2.2.1. Interviews with ESFRI cluster representatives

The FAIRsFAIR work packages WP3 (Data Policy and Practice) and WP6 (Competence Centres) collaborated on interviewing the following EOSC and ESFRI-supported ‘cluster’ projects. Interviews aimed to gather views and experiences of supporting FAIR data, and identify opportunities for collaborative action to improve practices.

FAIRsFAIR “Fostering FAIR Data Practices in Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- ENVRI-FAIR [2]– Europe's environmental research infrastructures
- EOSC-life [3]– creating EOSC for the life sciences
- ESCAPE[4] – astronomy & particle physics cluster
- PaNOSC[5] – Photon and Neutron Open Science Cloud
- SSHOC[6] – Social Sciences and Humanities Open Cloud

Three of the clusters were available for interview (ENVRI-FAIR, ESCAPE, PaNOSC, and SSHOC), although ESCAPE were interviewed too late to include in this report. The EOSC projects FREYA,[7] which provides infrastructure for persistent identifiers, also participated in the interviews as did ExPaNDS,[8] which supports the data management services at RIs based around Photon and Neutron facilities.

This summary sets out the key points from these interviews relating to policy and practice. Other points relating to the potential needs for support from a FAIRsFAIR competence centre are included in D6.1.

FAIR awareness and implementation

Making data FAIR requires a cohesive set of policies and systems to stimulate changes in data-related practices. The extent to which these support researchers' FAIR data practices can be understood as a form of 'FAIR maturity' for the organisations involved. The interviews sought cluster project representatives' views at an early stage in FAIRsFAIR of the current level of awareness of FAIR, and maturity of the support in the communities that they are engaged with.

Accordingly, project contacts were first asked, 'What, in your opinion, is the current level of maturity in the domains you support with regards to researchers making their data FAIR?' Suggested options ranged from 'no awareness', to 'general awareness but not actively putting into practice', 'active' or 'leading in several areas'.

Participants did not find it straightforward to align the suggested options with the researcher communities they support. Representatives from ExPaNDS and PaNOSC were generally in agreement that FAIR awareness and implementation is generally considered to be low to moderate in the various disciplines they are working with. Awareness of FAIR is stronger within the research infrastructure (RI) facility-based teams in these areas. PaNOSC noted some confusion in their user community between comprehension of Open data and

[2] http://envri.eu/envri-fair/
[3] http://www.eosc-life.eu/
[4] https://projectescape.eu/
[5] https://www.panosc.eu/
[6] https://www.sshopencloud.eu/
[7] https://www.project-freya.eu/Plone/en
[8] https://www.panosc.eu/related-projects/expands/

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

FAIR data. FREYA - echoed that in their experience the levels of awareness varied greatly across the domains, whilst noting that there are very few domains where FAIR data awareness does not exist at all.

Representatives from projects including PaNOSC, FREYA and ENVRI-FAIR representatives observed that FAIR data awareness and practice is sometimes due to the existence of agreed community standards that pre-date the FAIR principles, but are intended to make data (and as PanOSC noted, some software) findable, accessible, interoperable or reusable.

The interviews covered potential drivers for domains to develop relatively mature practices. Example answers were provided, including high data collection costs, common methodologies, mature data standards, technical specialisation of research tasks, high levels of researcher collaboration and interdependence, availability of specialist data repositories, extensive industry collaboration, and strong policy mandates. The FREYA representatives perspective was that all of these suggested drivers were relevant, and that change was unlikely to be driven solely by a strong policy mandate. It was likely that sharing data in the high-energy physics domain was, in their view, driven by the technical specialisation of research tasks. Data sharing is an essential part of doing research in this domain, as the data originates in highly centralised facilities such as the Large Hadron Collider, and needs to be made accessible to specialised computing facilities elsewhere.

Similarly, PanOSC outlined how the costs of transporting large data volumes in photon and neutron research led to a drive for improved research data management (RDM) awareness and practice, which in turn drove data policy development and moves towards standardisation of data container formats for interoperability.

In the environmental sciences, methodologies can also play a role. Some environmental studies are longitudinal, and as such data relating to observations at specific periods cannot be collected again (an issue also present in the social sciences). ENVRI-FAIR noted that funders in the environmental sciences are now requesting DMPs and supporting FAIR-related projects. The need for FAIR is recognised, to meet the challenge that environmental scientists face in comparing data across multiple sources and fragmented infrastructures. ENVRI-FAIR noted that where RIs don't already have sustainably operating systems , they are keen to understand and where possible reuse what other RIs have already developed.

FAIR data policy at the operational level

Interviewees were asked about the extent to which FAIR data is specifically dealt with in RI or institutional policy. The FAIR principles seem to have had relatively extensive influence on data policy development in this group. FREYA, as a project rather than an institution, does not have its own data policy, but its partners do. FAIR has probably influenced data policy

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

development – and definitely data-handling practice – at FREYA project host institution The British Library.

Where there is not a specific data policy at cluster level, as with ENVRI-FAIR, the FAIR principles still have influence in the challenging work of harmonising or aligning across partner RI data policies. PanOSC partners are also interested in data policy harmonisation. ExPaNDS partner institution, the Paul Scherrer Institut (PSI), has a data policy. ExPaNDS noted that the cluster values data policy harmonisation in the interest of giving researcher users across the community a consistent policy message. There is a perception that a RI may be disadvantaged by appearing to have a more demanding set of requirements than the others. EC-funded work is noted as being helpful for such harmonisation work.

Interviewees found barriers to implementing FAIR include lack of funding for data analysis and data management efforts. Monitoring compliance is challenging, specifically if data access is off-site, as noted by ENVRI-FAIR. PanOSC suggested that DMPs could be used in future to help assess compliance, and ENVRI-FAIR reported ongoing work to develop methodologies for compliance assessment.

Participants were asked about areas of operational level policy they would like to make machine-actionable, for example around DMPs or repository ingest, to make downstream data management processes more efficient in making data FAIR.

FREYA reported aspirations to develop automated PIDs linked to data at the moment of creation, including via the use of specific instruments such as microscopes. This would help with the smooth running of downstream RDM processes, and also help to inculcate the idea within the researcher user communities that data are citable from the point of creation.

There was support in ExPaNDS for data management to be automated, and interest in exploring APIs to help integrate partner services where this would be helpful. ENVRI-FAIR also supported making operational level policy more machine-actionable for increased interoperability.

Areas for potential collaboration

Interview questions returned to collaboration on RDM and/or FAIR data policy harmonisation. FREYA and ENVRI-FAIR noted they are expected to collaborate – including at the policy level – with other EU-funded infrastructures such as OpenAIRE and the EOSC clusters. In particular, the FREYA team has been working on the importance of PIDs for FAIR, for example, with the Social Sciences and Humanities Open Cloud (SSHOC). There is also interest in using PIDs in skills capacity development, for example, with OpenAIRE. Policy work is also important for FREYA when contributing to the EOSC FAIR Data WG and EOSC Architecture WG.

FAIRsFAIR “Fostering FAIR Data Practices in Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Interviewees were also asked about how FAIRsFAIR might amplify any FAIR-related codes of practice for making data FAIR in communities they work with. In response, ExPaNDS pointed to the crystallography community's FAIR data practices as potentially already useful to the photon and neutron facility communities, and in particular of use to the IT groups within those research domains.

Most participants responded to this question with ideas for practices or resources that would be helpful to the research communities with which they work. FREYA would support amplification of the awareness and appropriate use of PIDs across research domains; ENVRI-FAIR is interested in further support from FAIRsFAIR for the environmental sciences, including training provision; PanOSC would be interested in the provision of example data policies that could be shared, and adapted for use as templates.

ExPaNDS and ENVRI-FAIR reported collaboration opportunities at the European level and also beyond Europe, with some indication of interest in FAIR from similar facilities internationally, including the USA (ExPaNDS) and other participants in the Research Data Alliance (ENVRI-FAIR).

FAIRsFAIR values the potential for collaboration with all aligned efforts, particularly within the European research space. Although interviews with EOSC-life and ESCAPE were not possible in time for this report, further collaboration with EOSC-life is planned through that project's working group on FAIR.

2.2.2. Responses to FAIRsFAIR Policy and Practice Open Consultation

There were 106 responses to the Policy and Practice open consultation, provided by representatives of a wide range of organisations.[9] The majority of responses were received from staff working in Universities (44%) followed by responses from Research Infrastructure staff (30%), Research Performing Organisations (10%) and 'other' (9%). Those selecting 'other' included financing organisation, think tank; University Medical Center; Funding bodies, e-infrastructures, Institute of Technology; Cross-disciplinary trustworthy digital repository; cluster of Research Infrastructures; not-for-profit organisation.

We received responses to the open consultation from across Europe. The majority of responses were received from respondents based in the UK (19%), followed by the

[9] The anonymised data is available in Zenodo [Davidson, Joy, and Angus Whyte. 2019. 'FAIRsFAIR Policy and Practice Survey 2019 data for D3.1_D3.2_D6.1'. FAIRsFAIR. https://doi.org/10.5281/zenodo.3550529) Note that responses are given equal weight although some represent individual views, while others were submitted as a collective response on behalf of an EOSC project. A number of clusters responded this way in addition to being interviewed.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Netherlands (13%), Germany (12%), Finland and France (9% each), Ireland, Italy, Spain and Sweden (5% each).

Respondents identified with a variety of roles, 49% indicated that they fill several roles at their institutions. Of the 51% of respondents that have a unique role, the majority (14% of all respondents) working in research support and liaison followed by policy makers or senior managers (11% of the 106 respondents) and data stewards or research data librarians (9% of total answers).

There were ten questions, including seven asking about community expectations of repositories. The responses to these repository questions are described later in section 3.4. The remaining three questions related to the research disciplines the respondents were working with, and are described below.

Disciplines respondents work with

Most respondents said they work with 'data computing and digital research infrastructures', as shown in Figure 2.1. Other domains were more or less equally represented in the responses, with except for energy. Typically the 30 respondents who said they worked across two disciplines were referring to data computing (etc.) and one of the others. There were as many respondents working across all six disciplines as there were working in a single discipline (26 and 27 respectively).

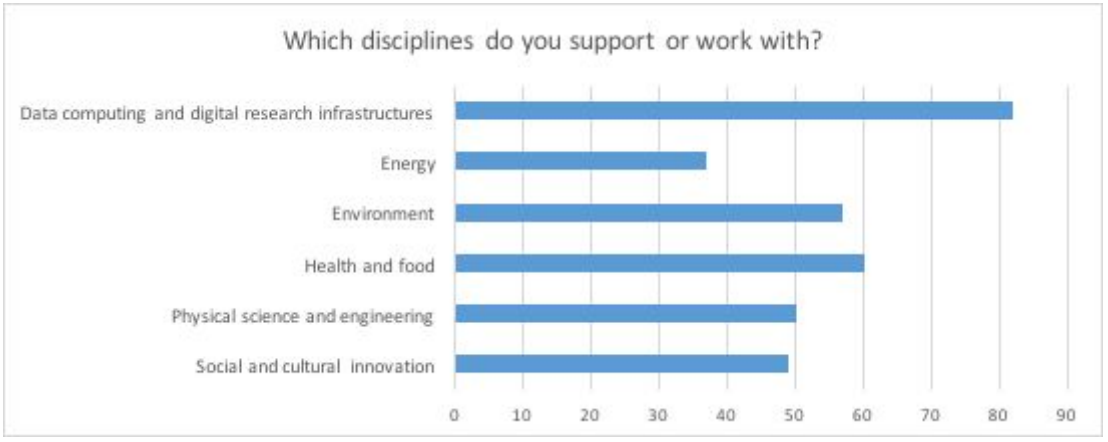

Figure 2.1 Disciplines that consultation respondents work with

Terms of access and reuse

Respondents were asked, "When researchers that you support deposit the research data what restrictions are commonly placed on access and reuse by others?" They were asked to identify how commonly access and reuse were restricted according to the options shown in Figure 2.2.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

The responses indicate that unrestricted access and reuse is common for around half of respondents. More respondents (72%) had found it common in their communities to restrict access to specific groups. The responses do not tell us anything about the dependencies in individual responses, e.g. individuals could respond 'very commonly' to as many options as they found applicable. However the balance between open and restricted access is interesting to compare with the recent State of Open Data report, illustrated in Figure 2.3. Most researchers responding to that survey indicated they make data openly available more than half of the time.

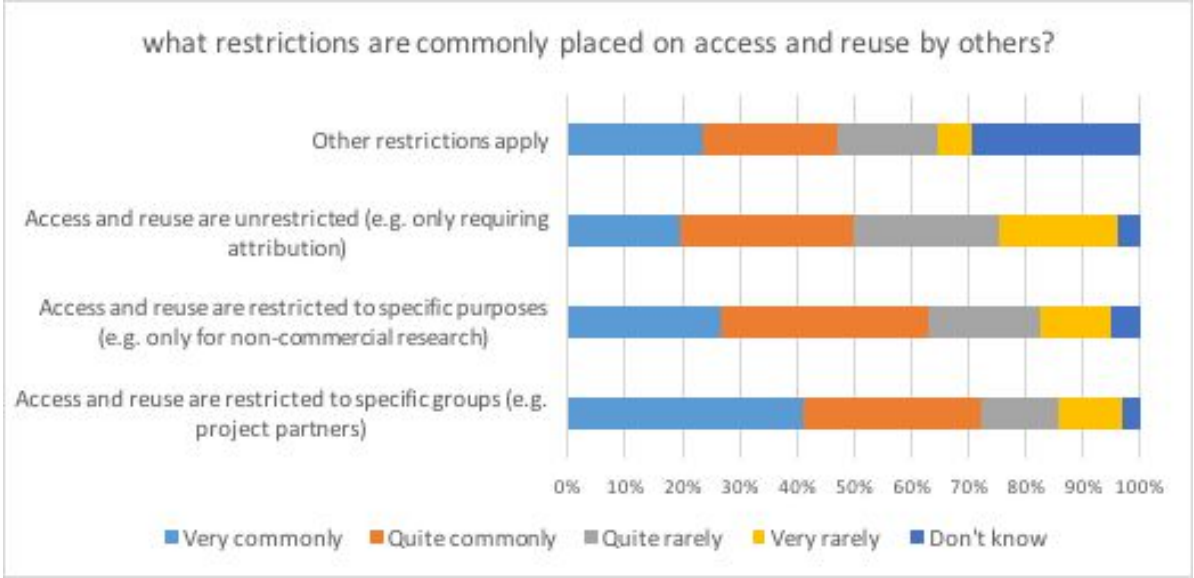

Figure 2.2 Respondents perception of common data access and reuse conditions

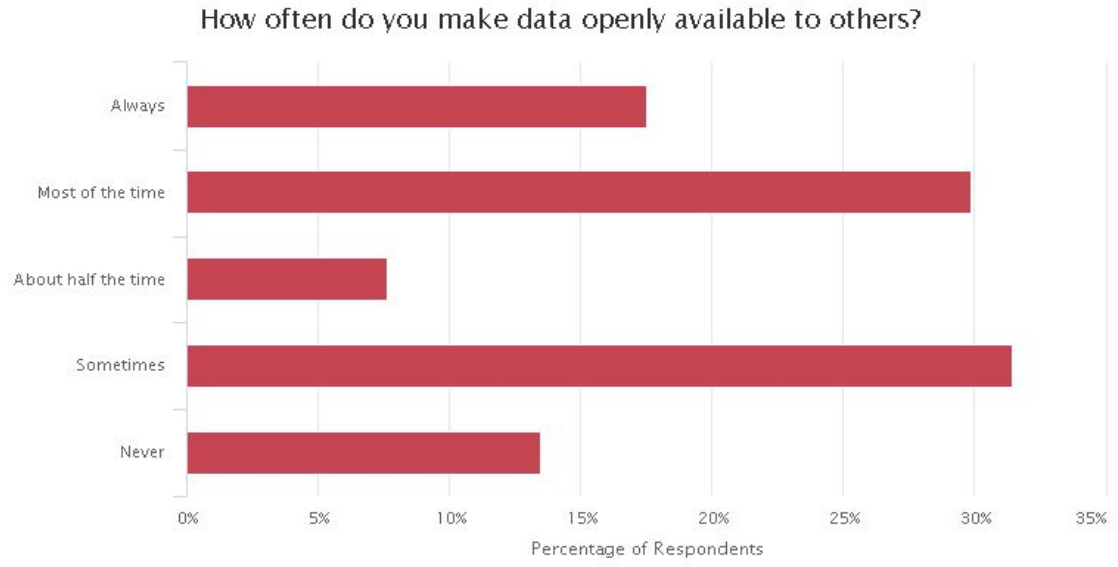

Figure 2.3 How often do you make your data openly available to others? State of Open Data 2019 data visualisation by Martin John Hadley (Figshare, 2019)

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Which disciplines are seen as 'advanced' on FAIR?

Respondents were asked to name any domains they consider relatively advanced in making their data interoperable and reusable, 'based on the researchers you work with/support'. There were 85 responses to the question on interoperability and 75 to the question on reusability.

To illustrate the relative differences across the main ESFRI disciplinary clusters, the responses were re-coded, as shown in Figures 2.4 and 2.5.

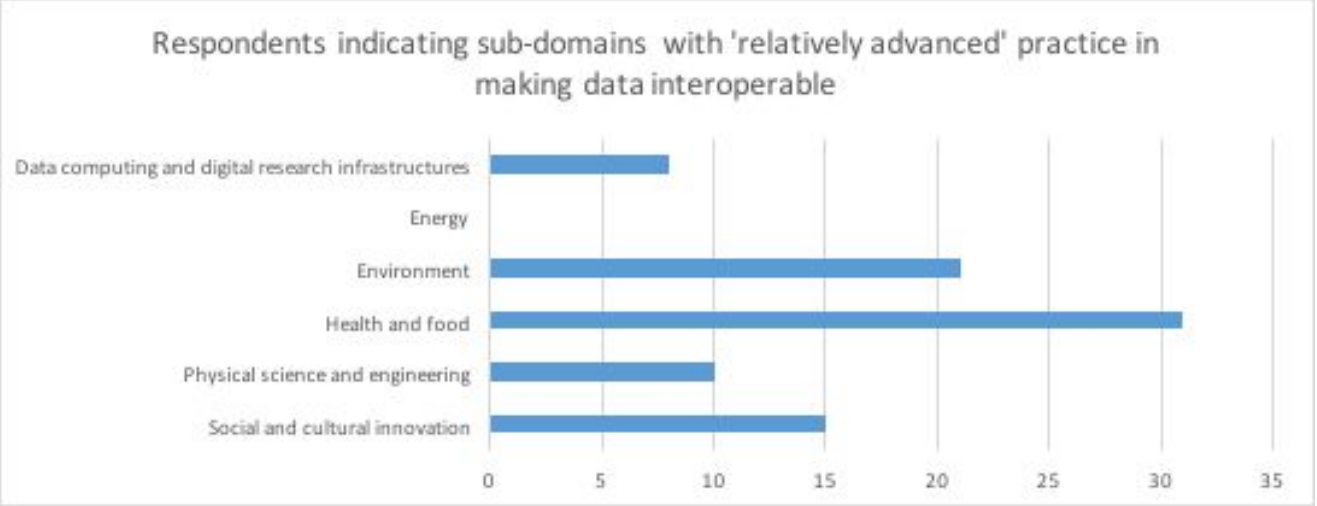

Figure 2.4 Number of respondents indicating sub-domains with 'relatively advanced' practice in interoperability, by disciplinary group

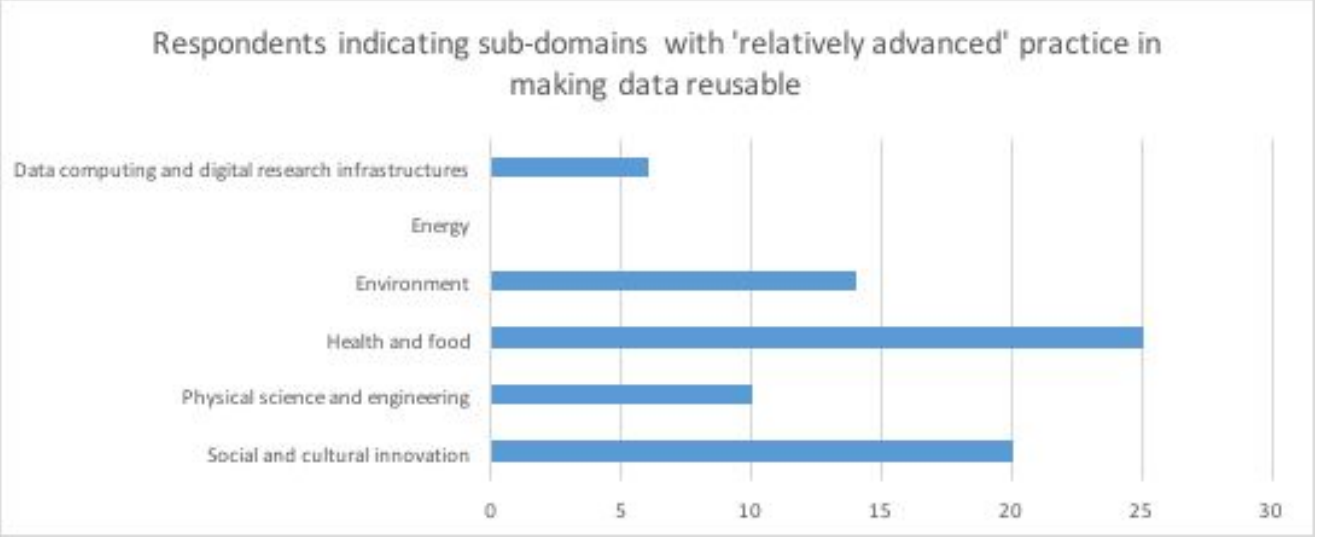

Figure 2.5 Number of respondents indicating sub-domains with 'relatively advanced' practice in reusability, by disciplinary group

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Responses to the open questions on sub-disciplines that are seen as relatively advanced in making data interoperable and reusable described these at a high-level e.g. 'crystallography' or 'oceanography' or 'proteomics'. Other responses identified communities that share specific practices or data characteristics, e.g. '[software] code based research' e.g. fields with 'relatively complexity and variability in their data'. One respondent identified "those with quantitative data, and who have a long history of building shared repositories in their research communities."

They also show that more respondents see Health and Food disciplines as relatively advanced than any other. More respondents see Environmental disciplines as advanced on interoperability than on reusability. Conversely, more respondents see Social and Cultural innovation disciplines as advanced on reusability than on interoperability.

Many of the respondents who suggested 'advanced' disciplines mentioned disciplines that were not in the areas the respondents associated themselves with. None of those who said they were working with Energy communities included disciplines in that area. Data management in Energy and Engineering is reported to be highly fragmented in the first findings of the CESAER Taskforce on Open Science, which is investigating FAIR RDM in that area. [10]

2.3 Factors affecting FAIR data practices

In this section, we review disciplinary, demographic and institutional factors affecting the data management and stewardship practices that enable FAIR data. Studies on these factors mostly focus on data sharing and predate the FAIR principles, and their findings demonstrate the relevance of the *Turning FAIR* recommendations.

The Springer Nature - Digital Science annual survey of research authors on their attitudes and behaviour around research data provides one indication of the need to raise awareness of the FAIR principles. The recently released State of Open Data 2019 survey, based on a survey of around 8000 respondents in 190 countries, demonstrates the lack of awareness of the principles among research communities.

[10] CESAER Taskforce (2019) https://rdm.engineering/fair-data-engineering-first-findings/

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

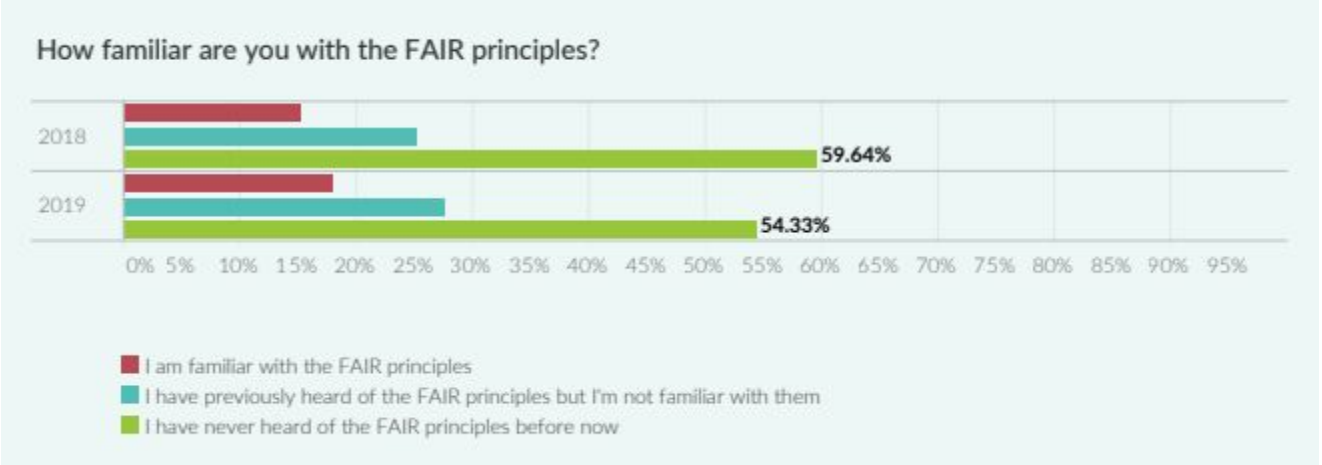

Figure 2.6. Familiarity with the FAIR principles 2018 versus 2019. Source: Figshare (2019)

As shown in Figure 2.6, there has been a slight increase between 2018-19 in the percentage of respondents indicating familiarity with the principles, but this is still under 20% of an audience that might be expected (as likely users of Figshare) to have a relatively high interest in research data management.

Research communities have been managing and sharing data for many years despite their low awareness of FAIR principles. There is a growing body of literature on the factors influencing practice in these areas. Fecher et al (2015) carried out a systematic review of approx. 100 sources and a fresh survey, to provide a conceptual model shown in Figure 2.7

The categories in Figure 2.7 cover data sharing and the RDM activities leading to sharing. They are useful for considering other studies that can similarly help us understand why FAIR-enabling activities are more extensive in some communities than others.

2.3.1. Researcher characteristics

The individual-level ('data donors' in Figure 2.7) factors include individual motivations for data sharing, i.e. the perceived rewards from making data FAIR. The most commonly identified benefits are quality improvements, peer visibility, and formal recognition/reward. Other important factors include the higher impact of their research, new contacts/opportunities for cooperation, and possibilities for data to be cited (O'Caroll et al 2017).

Researchers' *willingness* to share does not necessarily correspond to *actual* sharing behaviour. Different enablers and barriers influence researchers with specific socio-demographic characteristics differently. Such aspects as career seniority, gender, openness, and nationality (suggesting national policy influence) are related to the actual sharing of research data (Linek et al, 2017).

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

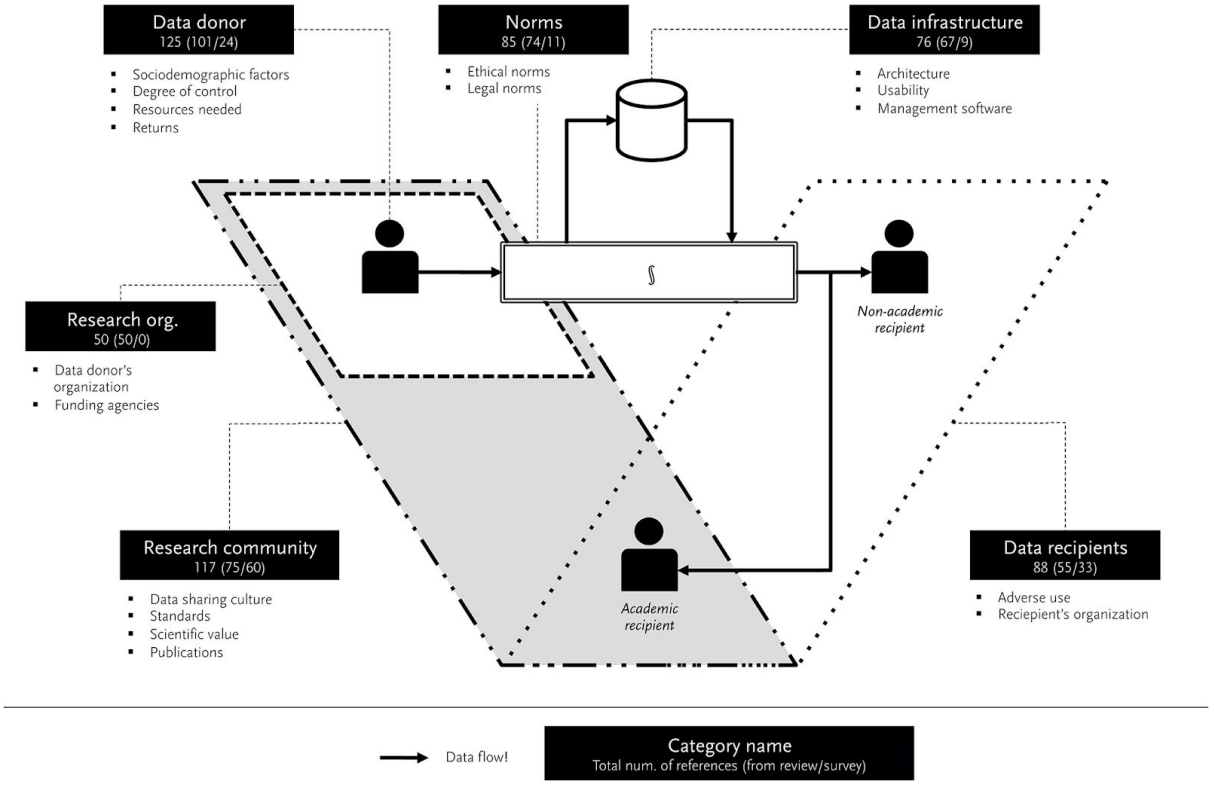

Figure 2.7. Factors influencing RDM and sharing. Source: Fecher et al. (2015)

2.3.2. Research organisation context – institutional support

The research organisation context refers to characteristics of the host institution, together with the relevant funding body's policy and grant conditions. Some funders and publishers have formulated their data policy around the FAIR Principles. The sister report to this (D3.1 FAIR data policy analysis) considers the current state of play on this.

Lack of institutional support to establish career structures and rewards for data management is a critical disincentive. According to O'Carroll et al. institutional support for RDM is generally lacking in institutions legal support. Intellectual Property Rights (IPR) and the technical infrastructure to facilitate Open Science has also been reported (Pryor et al, 2013).

Institutional support is commonly designed to match a lifecycle of data management actions that match relevant stages of a research project, commencing with a planning stage and ending with data publication. According to an international survey of libraries providing RDM services, carried out by Cox et al. in 2014 and repeated in 2018, libraries are beginning to recognise the FAIR principles as relevant, but they are only just beginning to gain ground in explicit policies (Cox et al 2019).

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Tenopir et al. (2017) surveyed 119 European academic research libraries from 22 countries in 2016 and found that they are more likely to offer consultative-type services than hands-on/ technological services. Consultative services are those frequently involving a personal client-librarian relationship about such things as how to find information on data management plans, metadata standards, or data citation practices. According to these authors, "the lower and slower uptake of technical services compared to consultative services may reflect the fact that these services require a substantial investment in time, resources, and new technical knowledge"(ibid.) Technical service take-up is shown in Figure 2.8.

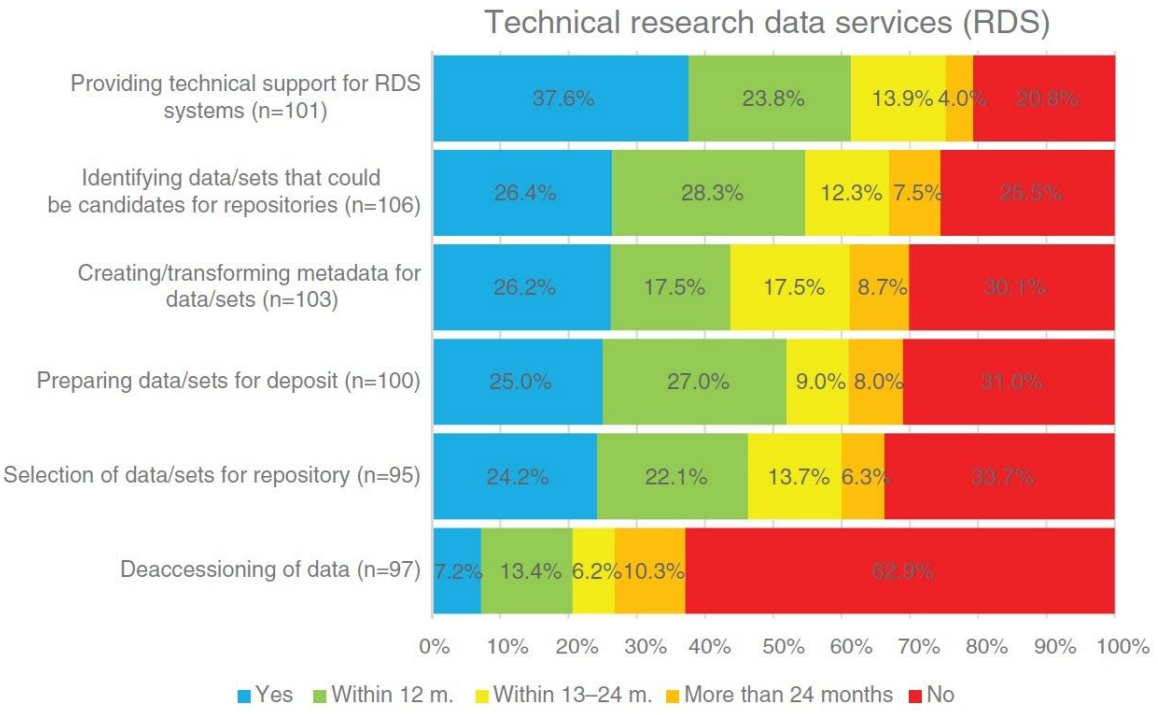

Figure 2.8 European academic library provision of technical research data services. Source: Tenopir et al. (2017)

Responses to the FAIRsFAIR policy and practice open consultation also indicate the importance of support to researchers. Support was considered the most favourable policy factor, influencing researchers' behaviour with more than 93% of respondents rating this factor as 'very' or 'quite' positive (Figure 2.9). More than three-quarters of respondents working in Universities provide in-house support (77%), followed by 66% of Research Infrastructures, and 64% of Research Performing Organisations. However, these figures should be viewed in light of the 2017-2018 EUA Open Access survey which revealed that

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

only 13% of European HEIs had developed institutional guidelines for open access to research data.[11]

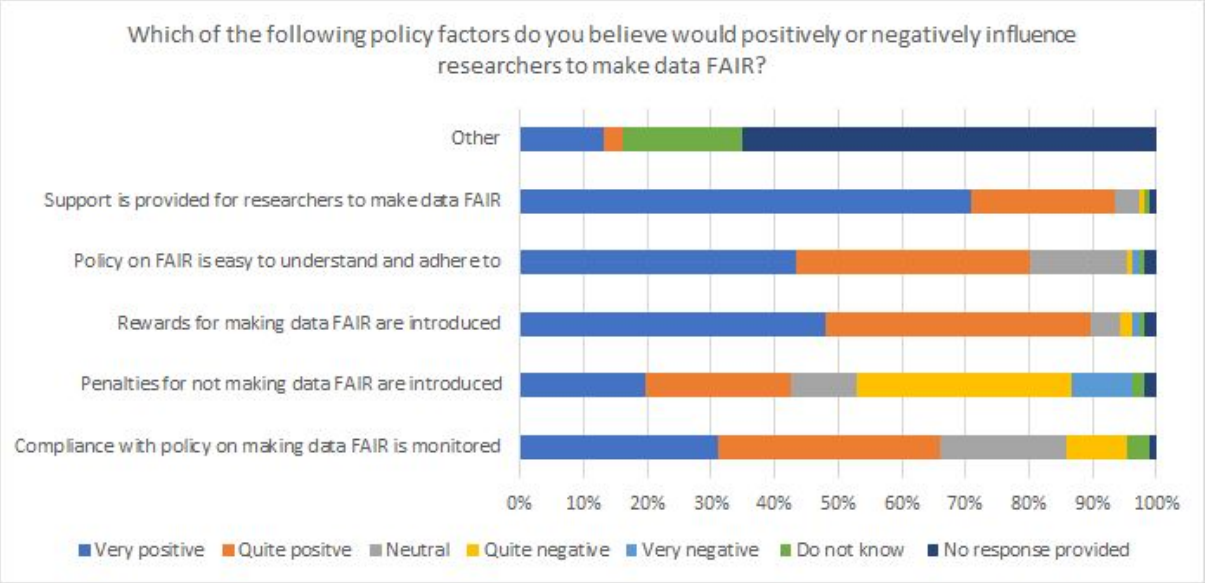

Figure 2.9. Influence of policy factors on researchers' practice, FAIRsFAIR open consultation

Institutional cultures have shaped the development of RDM support services, for example, referring to the diversity of cultures in UK institutions. Pryor (2013) describes the diversity in institutional cultures and claims that 'institutional inertia' can inhibit more established, research-intensive institutions from offering RDM support to researchers.

A critical issue for effective practice is to identify the points in the data lifecycle where intervention by RDM support providers is most effective in enabling FAIR data production and stewardship. Recent work by Yakel et al. (2019) examined collaborative data sharing, curation and reuse practices among eleven zooarchaeologists and two curators in the context of a data reuse project. This study examined how factors at one point in the data lifecycle impact on other points, forming virtuous (positive) and vicious (negative) circles. They conclude that "data producers partnering with data curators to steer the data production process, after data management planning and before data deposit, is critical to avoid the proliferation of vicious circles and enable meaningful data reuse. Key actions they highlight to promote good practice include the following:

- Curators should facilitate learning about higher level disciplinary-wide standards and best practices for consistent data entry, standardised training of those recording data, and systems that provide some automatic checks on data entry. In doing so

[11] https://eua.eu/resources/publications/826:2017-2018-eua-open-access-survey-results.html

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

they can mitigate idiosyncratic recording practices, although these may also be necessary for local interpretation.

- Data producers should communicate their reasons for selection decisions, as these provide important contextual information for reuse.
- Curators should actively steer decisions on standards, data structure solutions, and data recording norms throughout the data lifecycle.

2.3.3. Research community norms

Research community norms include the application of trans-disciplinary concepts such as research integrity, transparency and reproducibility. They also include discipline-specific practices such as the development of metadata standards to describe data and analytic procedures relating to the research questions commonly pursued in the discipline.

Disciplinary differences in data management practices relevant to making data FAIR have long been acknowledged. Case study evidence indicates that the benefits of formalised practices are more strongly felt in fields that exhibit specific characteristics. These characteristics include factors identified with research culture (Lyon et al 2013, Cragin et al, 2010). The study by Tam (2019) of differences within a single data-intensive domain, namely Geography, shows that generalising practices to the domain may give a false picture.

Tam's study compares the twin strands of physical and human geography.It describes sub-disciplines in these categories that are oriented respectively to the physical sciences, or the social sciences and humanities. The study is based on a mix of qualitative interviews with researchers, and content analysis of departmental websites, coupled with bibliometric analysis of co-authorship patterns at sub-disciplinary level. The study uses a framework based on earlier educational research by Becher and Trowler (2001). This describes disciplines and sub-disciplines on various dimensions as follows;

- Hard/soft: 'hardness' is based on features including the production of testable predictions, controlled experimentation, quantifiable data and mathematical models, where a high degree of accuracy, objectivity cumulativeness, and replicability are valued.[12]
- Pure/applied: this refers to the extent to which the research is applied towards finding solutions to practical issues.
- Urban/rural: this refers to the ratio of researchers to research problems, i.e. in 'densely populated' disciplines a substantial number of researchers work on a single research problem.

[12] The hard/soft science distinction has been widely used in other studies of science practice, see e.g. Wikipedia entry: https://en.wikipedia.org/wiki/Hard_and_soft_science

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- Convergent/ divergent: this refers to the degree of consensus on the discipline's research topics, standards, methods and techniques.

According to Tam's study, there is a propensity to share data in sub-domains that are relatively hard, urban, and convergent. The pure/applied dimension is not relevant to sharing, and Tam suggests this is because research in Geography typically involves both pure and applied aspects.

Survey and bibliometric studies suggest similar domain characteristics have a role to play in realising benefits from data management and sharing. Survey evidence points to similar characteristics. For example, the report *'Open Data: the researcher perspective'* (Berghmans et al., 2017) found that among their survey respondents "... the fields of computer science, physics, and astronomy...have the most positive view of data sharing" (p.21). The authors state that fields, where data tends to be managed collaboratively and shared within the research group, are more likely to use cloud-based archives and repositories to share publicly.

Research in fields where the transfer of data amongst collaborators is not essential for data analysis or interpretation are more likely to hold data in personal, departmental or institutional archives. They add that "collaborative research is a common driver of data sharing in all fields. Our study suggests that the concept of open data speaks directly to basic questions of ownership, responsibility, and control." (ibid.p.5).

Bibliometric studies have been carried out in a number of fields indicating that authors receive more citations to their articles if they share the underlying data. These have tended to be in domains that may be described as relatively 'hard' and 'convergent' using the Becher and Trowler typology, for example microarray studies in life science, or experimental political science.

2.3.4. Legal and ethical parameters

Legal and ethical parameters affect the entire lifecycle of research data from collection through to reuse. Common issues for researchers include the following :

- Protection of personal data to implement the General Data Protection Regulation
- Freedom of information (FOI) and environmental information regulations (EIR)
- Intellectual property rights (IPR) in data and databases, particularly the licensing conditions for data sharing and reuse
- Cloud storage service provision, and potential barriers to sharing presented by storing data across multiple legal jurisdictions

Researchers' lack of awareness of these issues is highlighted in recent surveys. Specific issues include subject consent terms and user agreements that facilitate sharing, and

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

awareness of intellectual property rights (IPR), impacting on their ability to share data and limiting the legal interoperability of research data sets that are shared, and therefore impeding data reuse. For example, according to O'Carroll et al. (2017) "It is clear that most researchers either do not issue or simply do not know about user agreements. Early-career researchers tend to issue less user agreements and say they know less about user agreements than senior researchers."

Berghmans (2017) states that "legal and ethical concerns are cited as reasons for not publishing research data alongside an article: a substantial proportion of the survey answers on this topic mention that data is proprietary or that researchers do not have consent to share data", and that "...researchers are not actively thinking about reuse licenses they can assign to their data. When asked which creative commons license they would make their data available under, 62% answered that they didn't know. Where researchers did provide an answer, they tended to favour more restrictive licenses." (ibid. p.24)

Inconsistent licensing and incompatible licensing terms are a significant barrier to research data interoperability. For example, it can become impractical to integrate and analyse multiple datasets if the terms applicable to each has to be examined on a case by case basis. To address these issues, a set of 'legal interoperability principles' has been asserted by the Research Data Alliance as high-level guidance for the research community.

Research data services that use cloud storage for data during research projects are also likely to be affected by legal interoperability issues, as are archival services that contract their storage to cloud service providers. These relate mainly to third-party rights in the data, and the need to ensure and the need for service level agreements ensure data protection and comply with IPR terms and conditions when data is placed in overseas legal jurisdictions (National Archives UK, 2014).

2.3.5. Data recipients

The potential recipients of data, including third-parties and characteristics of their organisation, are significant factors in researchers' data sharing behaviour. According to Berghmans (2017), more researchers agree that having access to other researchers' data would benefit them (73%) than agree that they are willing to share their data (64%), or have shared data (65%).

These authors' definition of 'sharing' is also very broad, including (non-public) person-to-person sharing by email, and including where that is to the researchers' direct collaborators. According to their results, only 39% share with "external parties" and "only 14% share data directly with researchers they do not know when they are working on a project." [20] Very similar figures are reported in the survey by O'Carroll et al. (2017) who say that "Almost two-thirds of researchers grant access to their data to research project/group

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

members and almost half grant access to interested persons by request" but only around 15% to "everyone".

Researchers' desire for control over sharing has been explored in several studies, including Tenopir et al. (2015) and Fecher et al. (2015) characterise the types of control sought by researchers in terms of first-use rights, control of reuse conditions, and knowledge of reusers (e.g. through registration or user agreement). Applying access and reuse conditions of these types would not necessarily contradict FAIR principles, but would only meet the "as open as possible, as closed as necessary" principle of the EC Horizon Europe if they were justifiable on ethical or confidentiality grounds (Lahti et al, 2019).

2.3.6. Data infrastructure

Available and usable infrastructure enables FAIR data production, the lack of either availability or usability impedes FAIR data practice. Fecher et al. (2015) identify three main aspects of data infrastructure from their systematic review; the *architecture* to enable sharing with access control where required; the *usability* of the technology; and *documentation and metadata management* for findability, interoperability and reuse.

Infrastructure elements include both technical or 'hard' services (for computation, networking and storage) and 'soft' services to make these more technical services work for specific research aims and contexts. These may be classified according to the data management capabilities they support across the research lifecycle. The EOSC pilot project FAIR4S framework described in section 1.2 identifies ten capabilities for making data FAIR and keeping it FAIR.

According to Berghmans (2017), 60% of their respondents strongly agreed that RDM specialists need to play a role in research data sharing". This is primarily to reduce the efforts involved; "researchers describe that research data management typically requires "some" (59%) to "a lot of" (25%) effort. The main reasons for this level of effort include the need to navigate legal issues (e.g., confidentiality, legislative issues), format the data (i.e., presenting it clearly), develop logistics (e.g., where to upload), and perform data cleaning (i.e., making the data usable)."

The use of metadata standards is critical to FAIR, and a key aspect of the time spent on preparing data for sharing. However, the adoption and use of metadata standards across domain and organisational boundaries is not straightforward. According to Edwards et al. (2013), metadata represent a form of scientific communication, and it is important to facilitate researchers' communication processes about data, even when standard products are available. "Well-codified metadata *products* increase the precision with which a dataset can be fitted to purposes for which it was not originally intended, or can be reused by people who did not participate in creating it. At the same time, ephemeral, incomplete, ad hoc metadata *processes* act as lubricants in disjointed, imprecise scientific communication.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

This latter category of metadata frequently appears alone, in the case of datasets for which no metadata products exist, but it also frequently appears in the actual use of metadata products." (ibid.) Metadata standards, like data management plans, require improvisation and social negotiation to put into practice.

2.3.7. Beyond disciplinary views of FAIR data practice

Data management practices are influenced by disciplinary boundaries, but it is clear that other factors also matter. This section briefly outlines alternative views include the Wenger's 'community of practice' (Wenger, 1998), 'research repertoire' (Leonelli and Ankenny, 2015), and 'data community' (Cooper and Springer, 2019). The community of practice is the longest-established and broadest of these sociological approaches and the latter two build on it.

According to Wenger's approach, a community of practice (*CoP)* is a group of people sharing a common activity, whether that group evolves naturally through shared experience and learning, or is constructed with that aim, as in an online community of practice. Wenger characterises a CoP according to three main dimensions:

- Mutual Engagement: members establish norms of good practice by building collaborative relationships
- Joint Enterprise: through their interactions, members create a shared understanding of what binds them together; i.e. the 'domain' of the community.
- Shared Repertoire: communal resources, including symbolic ones such as unifying concepts, that are used to pursue the domain aims. (ibid.)

The CoP concept has been extensively used in knowledge management, and the term is commonly associated with the online variant. For example, an online CoP has been established with support from OpenAIRE, for coordinators of data management training.

The 'research repertoire' has emerged more recently as a conceptual model for studying research collaborations. It results from studies in the life sciences of the 'ensemble of material and social conditions that make it possible for a short-term collaboration, set up to accomplish a specific task, to give rise to relatively stable communities of researchers.[16] The approach examines relationships between the adoption of instruments for data production, and the development of infrastructures and related community norms, such as databases and guidelines on data sharing. It encourages a focus on shared commitments to techniques, assumptions, values, institutions, funding sources, and methods.

The 'data community' is a similar concept, but has a slightly narrower focus. Cooper and Springer define a data community as "a fluid and informal network of researchers who share and use a certain type of data, such as crystallographic structures, DNA sequences, or

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

measurements relating to natural disasters". According to the authors, successful data communities are characterised by three features: bottom-up development, absence or mitigation of technical barriers to sharing, and community norms. [17]

Any of these three alternative ways of framing community may be more helpful than broad disciplinary categories for future work in FAIRsFAIR, on supporting emergent collaborations that lead to increased FAIR data production. The 'data community' approach has been adopted in the SSHOC cluster, and the 'research repertoire' approach has particular relevance to EOSC-life. Nevertheless, the ESFRI categories are a useful starting point, to help identify a suitably diverse spread of communities to conduct outreach activity with and begin practical support.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

3. Communities of data practice and action on FAIR

This section analyses research community action towards a culture of FAIR data, based on selected 'culture-related' *Turning FAIR* recommendations identified in section 1. Under each heading evidence is given of the extent of action on the relevant points for each research community. The analysis should offer a basis for further engagement.

3.1 Interoperability Frameworks and Metadata

Table 3.1. FAIR interoperability frameworks and metadata take-up

Relevant FAIR data practices	Scope and resource, Advise and enable, Capture and process, Integrate and analyse, Publish and release, Expose and discover

FAIR Action Plan progress (and relevant action points)	Phys. Science Eng.	Energy	Environ -ment	Health/ Food	Social Cultural	Data Comp. Digital
Usage of domain standards (r4.1, r4.5, r4.5, r8.2)						
Advocacy across domains, and of cross-domain standards (r4.2, r4.4)						

Key: FAIR initiatives by stakeholders are: Rare or missing, low take-up Growing, patchy take-up Common, extensive take-up Information unavailable

Relevant action points on the recommendations for interoperability frameworks and metadata cover the development and use of domain standards, and advocacy.

3.1.1. Development and use of domain standards

Turning FAIR Recommendations

- *r4.1 Enabling mechanisms must be funded and implemented to support research communities to develop and maintain their disciplinary interoperability frameworks*
- *r4.3 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.*
- *r4.5: 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.*

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- *r8.2 Metadata standards should be adopted and used consistently to enable machines to discover, assess and utilise data at scale.*

In a paper on the "State of FAIRness in ESFRI projects", Wittenburg et al.(2020) assess the activities towards FAIR implementation of five "typical examples" of distributed infrastructures (CLARIN, ICOS, EPOS, IS-ENES and BBMRI). All of them are working on improving interoperability in one way or another, e.g. by harmonising metadata descriptions, creating metadata catalogues and indexes or developing and promoting standards and ontologies.

In the following, we look at one of the disciplinary groupings used for the ESFRI roadmap - social and cultural innovation or, in other words, the social sciences and humanities (SSH) - in more detail and, in addition, briefly describe examples from other domains.

Based on 16 expert interviews complemented by desk research, the SSHOC project (2019) examined the current use of metadata standards and data formats in the SSH. It identified interoperability problems, developed recommendations on metadata standards and data formats for different SSH domains and outlines further SSHOC activities to enhance interoperability, primarily based on conversion services. The report describes the metadata landscape in the SSHOC domains as "heterogeneous and evolving", but identifies the most important standards for each domain (In brackets the infrastructure initiatives and SSHOC partners associated with the domain):

- Social Sciences (CESSDA, ESS, SHARE)[13]: DDI[14] Codebook, DDI Lifecycle, DataCite, Dublin Core
- Arts and Humanities (DARIAH): TEI[15], CIDOC CRM[16], Dublin Core
- Language science (CLARIN): CMDI[17], TEI, Dublin Core, OLAC[18]
- Heritage science (E-RIHS): EDM[19], Dublin Core.

While most of these are domain-specific, Dublin Core and DataCite are generic. Another exception is CMDI, which is not a metadata standard but rather a "framework intended to express a vast number of different metadata profiles using a common language" according to the SSHOC report. It notes the tremendous heterogeneity in controlled vocabularies used in the communities it surveyed, including the following as examples: [20]

[13]

[14] Data Documentation Initiative

[15] Text Encoding Initiative

[16] CIDOC (International Committee for Documentation) Conceptual Reference Model

[17] Component Metadata Infrastructure

[18] Open Language Archives Community

[19] Europeana Data Model

[20] Ibid., p. 14, 18.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- CESSDA: CESSDA Topic Classification, DDI Controlled Vocabularies, ELSST[21]
- CLARIN: CLARIN Concept Registry, CLAVAS[22], ISO 639-1 language list, local vocabularies
- DARIAH: GND[23], OpenGeoNames, TaDiRAH[24], TGN[25]
- E-RIHS: AAT[26], PICO Thesaurus, TGN, VIAF[27]
- ESS: DDI Controlled Vocabularies, ESS' self-defined controlled vocabularies.

Interoperability problems perceived by the informants of the SSHOC study include metadata as well as data interoperability problems. With regard to the former, one problem reported is that concepts or metadata fields are not always precisely defined in some standards, which leads to diverging interpretations by different users. In some cases, the same concept or field is even defined inconsistently across different standards. Another issue reported stems from older metadata records which are not compliant with modern standards. The flexibility of DDI and CMDI can (and is) regarded as an advantage, but also has a downside, according to the report. Both offer an extensive set of concepts, from which the user chooses a subset to create an application profile tailored to a specific use case. The results are diverging user-created profiles which may be "technically compatible in terms of validation" but "not comparable in terms of content".(ibid.)

The most frequently reported issue relating to data interoperability was that some tools only work with one single data format, thus making it necessary to use conversion tools if the data is to be used in other contexts or even just for curation and preservation. The conversion process can be associated with quality assurance problems (e.g. if data structure or content are misinterpreted by the conversion tool) or loss of information. Legacy formats, often requiring a chain of multiple conversions, are particularly challenging in this respect. Other issues include interoperability problems within a data format (caused by version dependent features) and data formats that are not properly documented.

Interoperability issues are and will be further addressed by SSHOC. It developed recommendations on metadata standards, data formats and formats for controlled vocabularies and will work on interoperability solutions. [28] As the diversity of formats and standards in use in the research communities make achieving the "ideal" solution (one common metadata standard, one common data format) rather unlikely, SSHOC is focussing on "aiming and providing conversion services for what we consider the major recommended

[21] European Language Social Science Thesaurus
[22] CLARIN Vocabulary Service
[23] Gemeinsame Normdatei / Integrated Authority File
[24] Taxonomy of Digital Research Activities in the Humanities
[25] Getty Thesaurus of Geographic Names
[26] Art and Architecture Thesaurus
[27] Virtual Authority File
[28] For details on the recommendations, see ibid., pp. 22-27.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

metadata standards and data formats used in the SSH".[29] These services are planned to be offered and/or brokered via a yet to be established "Interoperability Hub for the SSH" along with guidance. Another goal is to identify and collect more complex interoperability solutions that exceed the capacity of one single service and require complex workflows - and to register them in the "SSHOC marketplace" to be built in the course of the project.

An example of a longstanding community-driven standardisation endeavour from the field of physical science and engineering are the Virtual Observatories (VOs). The aim of the VOs is to enable interoperability of astronomical datasets and other resources across data centres worldwide. To this end, the International Virtual Observatory Alliance (IVOA) is working on agreeing to technical standards for the Virtual Observatories.[30] These comprise specifications for various astronomical data types and related vocabularies.[31] ESCAPE will extend these standards to align them with the FAIR principles, allowing them to be integrated into EOSC.[32]

A series of 13 interviews conducted by the CESAER (Conference of European Schools for Advanced Engineering Education and Research) Taskforce on Open Science has shown that the uptake of research data management practices and the FAIR principles in the area of engineering has been rather low so far. Among the challenges identified are the heterogeneity of data and a lack of documentation guidelines and standards.[33]

In the field of environmental research, there are many metadata schemas in use in the various related infrastructures as well as a large number of standardisation efforts. One example is the NERC Vocabulary server developed and maintained by the British Oceanographic Data Centre (BODC)[34]. The web service supports the machine-readability and interoperability of data from a wide range of disciplines involved in oceanographic research using SKOS.[35]

The Dynamic Ecological Information Management System - Site and dataset registry (DIMS-SDR)[36] provides information about long-term ecosystem research sites all over the world (e.g. location, ecosystems, relevant research themes, facilities and parameters measured) and in some cases also access to associated datasets. The registry is based on

[29] Ibid., p. 27.
[30] http://www.ivoa.net/
[31] http://www.dcc.ac.uk/resources/metadata-standards/international-virtual-observatory-alliance-technical-specifications
[32] https://projectescape.eu/sites/default/files/lamanna-escape-vienne-2019.pdf
[33] CESAER: FAIR Data in Engineering - First Findings.
https://rdm.engineering/fair-data-engineering-first-findings/
[34] https://www.bodc.ac.uk/resources/products/web_services/vocab/#intro
[35] Simple Knowledge Organisation System
[36] https://deims.org/

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

metadata models[37] for several entities (activity, dataset, network, person, sensor, site) which were developed in cooperation with research projects and stakeholders (e.g. LTER).

In the area of agricultural data, there has been a sustained, collaborative effort to build consensus regarding data sharing policies and practices. The RDA Interest Group on Agricultural Data (IGAD)[38], formed in 2013, has created specialist Working Groups to advance data interoperability, namely the Wheat Data Interoperability WG, Rice Data Interoperability WG, AgriSemantics WG, On-Farm Data Sharing WG, and Capacity Development for Agricultural Data WG. Outputs of these working groups include the Wheat Data Interoperability Guidelines[39], which has seen widespread adoption across organisations in Europe, the USA and Australia.[40]

In addition to these groups, the eROSA "Roadmap for a pan-European e-Infrastructure for Open Science in Agricultural and Food Sciences" (Zervas et al. 2018) points to several initiatives to map the standards landscape such as the GODAN map of agri-food data standards.[41,42] The GODAN map currently covers 403 standards, relevant to food and agriculture. The eROSA roadmap offers a maturity assessment of the agri-food community digital assets, in which 'metadata schemes registration' and 'ontology and concept repositories' are both evaluated as advanced in terms of technology readiness.(Zervas et al, op.cit.) Further steps to be taken for these items include bringing operative prototype services to production level, aligning with similar enterprises across domains (e.g. the FAIRsharing registry), and opening services with cross-domain relevance to EOSC partners.

Despite progress already made, Zervas et al. (op.cit.) highlight the proliferation of standards and a risk of maintaining silos of activity within subdomains. In their view, to improve interoperability, there needs to be "high-level agreement on advised standards to use for different types of research applications. The exemplary approach of the RDA Wheat Data Interoperability working group in breaking the boundaries among wheat research community could be generalized and expanded in order to clarify the complex landscape of existing standards and deter communities from duplicating efforts."

[37] https://deims.org/models
[38] https://www.rd-alliance.org/groups/agriculture-data-interest-group-igad.html
[39] http://dx.doi.org/10.15497/RDA00018
[40] https://ist.blogs.inra.fr/wdi/adopters/
[41] https://vest.agrisemantics.org/; https://doi.org/10.7490/f1000research.1115260.1
[42] See also semantic resources such as Global Agricultural Concept Space (GACS) http://agrisemantics.org/node/8; AgroPortal http://agroportal.lirmm.fr; Crop Ontology http://www.cropontology.org

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

3.1.2 Advocacy across domains and of cross-domain standards

Turning FAIR Recommendations

- *r4.2 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.*
- *r4.4: Mechanisms should be facilitated to promote the exchange of good practices and lessons learned in 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.*

To support the interoperability aspect of FAIR, research communities need support in developing interoperability frameworks. Some disciplines have already addressed this and those examples should be highlighted, and promoted to other communities to encourage the broader development of frameworks. Some samples of good practice in collaborative work on interoperability are featured in this section.

Metadata standards are key components of interoperability frameworks, along with standards for persistent identifiers. Existing metadata standards are collated in the RDA Metadata Standards Directory[43] developed through an RDA working group.[44] The directory lists available metadata standards by discipline and provides use cases highlighting repositories that use those metadata standards successfully to describe their collections. Two examples of use cases are the Common Data Index *CDI*[45] of the SeaDataNet data centres using the CDI metadata format to aggregate and provide information about available datasets and the Ecological Metadata Language *EML*[46] which was specifically developed for ecological science and is used by LTER. SeaDataNet and LTER (Europe) are part of the ENVRI community and involved in related activities regarding metadata harmonisation and cataloguing (ENVRIplus (2018).

One discipline with very well developed infrastructures and interoperability frameworks is geosciences. *INSPIRE* (Infrastructure for Spatial Information in the European Community)[47] was set up through a directive of the European Parliament and Council in 2007 to be completed by 2019 to harmonise spatial datasets so cross-border applications can be built on top. The 34 data themes are supported by technical guidelines, data models and metadata to facilitate interoperability between datasets, and the framework can be

[43] https://rd-alliance.github.io/metadata-directory/
[44] https://www.rd-alliance.org/groups/metadata-standards-directory-working-group.html
[45] https://www.seadatanet.org/Metadata/CDI-Common-Data-Index
[46] https://rd-alliance.github.io/metadata-directory/standards/eml-ecological-metadata-language.html
[47] https://inspire.ec.europa.eu/

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

extended to fit national or domain-specific needs.[48] The metadata framework is a European Commission regulation available in 23 languages providing a standard to allow interoperability across most of Europe (CEC, 2008). INSPIRE is and will be the basis for the metadata schemas used by several ENVRI-FAIR infrastructures. The Integrated Carbon Observation System *ICOS* is currently developing its metadata schema building on INSPIRE to ensure their data is findable and interoperable.[49] *IAGOS* (In-service Aircraft for a Global Observing System) data is published through the AERIS data catalogue which also uses a metadata schema that is an extension of INSPIRE to facilitate interoperability (ENVRIplus, 2018).

A metadata schema for all resources on the web is schema.org[50] which is used to provide structured information for web pages which can also cover landing pages providing metadata for datasets. Google Dataset Search[51], for example, relies on schema.org to index datasets in their database[52], enhancing the findability of datasets. schema.org can be extended to cover additional entities or relationships. Bioschemas is an initiative to extend schema.org for the life sciences, by defining profiles of a range of research objects e.g. data repositories, datasets, events, and biosamples to ensure that these are discoverable.[53] Lead by ELIXIR, Bioschemas will be relevant to EOSC-Life and to the current RDA working group activity on applying schema.org for research data discoverability.

A major initiative to facilitate findability, access and interoperability to resources is the work carried out by the FREYA project to extend the infrastructure for persistent identifiers (PIDs). The PID graph developed by the project connects and integrates PID systems, creating relationships across a network of PIDs and thus connecting information and increasing discoverability. Building on existing PID services, FREYA is building a PID graph, a network of interconnected PID systems that new discovery services can be based on (Fenner and Aryani, 2019). For the PID graph to be successful, services need to buy into a wider interoperability framework, e.g. by using standardised services such as DataCite and its metadata schema. The FREYA project is currently working on providing ways to explore the PID graph, which will result in case studies highlighting incentives to contribute to the broader infrastructure (Fenner, 2019).

There is an evident need for researchers to access training resources on both domain-specific and cross-domain standards for interoperability. A range of these resources

[48] https://inspire.ec.europa.eu/data-specifications/2892

[49] https://www.icos-cp.eu/about-icos-data
[50] http://schema.org/
[51] https://toolbox.google.com/datasetsearch
[52] https://support.google.com/webmasters/thread/1960710
[53] https://bioschemas.org/

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

has been developed. However, they are not always findable and accessible for researchers or data stewards. It has been argued that training resources themselves need to be FAIR, and 'interoperable' in the sense that they are described according to a standard set of terms relating to FAIR enabling practice. The initiative 'terms4FAIRskills'[54] is creating a terminology describing competencies, skills and knowledge associated with activities to make data FAIR. SSHOC, ELIXIR and the FAIRsFAIR Competence Centre are looking into applying the terminology to enable the training of FAIR-related skills for researchers, data stewards and data managers and other use cases.

[54] https://terms4fairskills.github.io/

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Example of good practice: Metadata in the endangered languages community

The DOBES initiative (http://dobes.mpi.nl/) started to document critically endangered languages in 2000. From the start, they agreed to principles for data collection that are well aligned with the FAIR principles. From 2008, CLARIN adopted many of the DOBES principles, and the datasets are now accessible through the CLARIN language archive (https://archive.mpi.nl/).

DOBES data is now hosted in a CoreTrustSeal certified repository alongside the MPI for Psycholinguistics Archive. The repository supports searching for as well as browsing data sets in a structured way, where users can browse by collection, contributor, country, format, genre, and language. All datasets are described by structured metadata using the CLARIN developed Component MetaData Infrastructure (CMDI)[55] and assigned a persistent identifier (handles). Metadata can be exported in CMDI format or just standard Dublin Core which makes it available in a specific format to the disciplinary community but also allows findability, interoperability and reuse by the broader scientific community.

The vast majority of the data in the archive is available without or with minimal access restrictions; depositors are required to obtain informed consent for sharing the datasets from study participants. To facilitate the deposit and analysis of datasets, The Language Archive also provides a range of software tools implementing evolving standards in language documentation including the creation of a well-structured metadata catalogue.[56]

3.2 Data Management Plan Support and Usage

Relevant FAIR data practices	Scope and resource, Advise and enable, Plan and design

FAIR Action Plan progress (and relevant action points)	Phys. Science Eng.	Energy	Environ -ment	Health/ Food	Social Cultural	Data Comp. Digital

[55] https://www.clarin.eu/content/component-metadata

[56] http://dobes.mpi.nl/archive_info/tools/

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Responding to funder requirements, enabling support r5.1, r5.2, r5.3						
Put DMPs to work as machine-actionable documents r22.4, r22.5						

Key: FAIR initiatives by stakeholders are: Rare or missing, low take-up Growing, patchy take-up
Common, extensive take-up Information unavailable

Table 3.2. Extent of DMP Support and Usage

3.2.1 Responding to funder requirements and enabling support

Turning FAIR Recommendations:

- *r5.1: 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.)*
- *r5.2: Data Management Plans should be living documents that are implemented throughout the project. A lightweight data management and curation statement should be assessed at the 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.*
- *r5.3: 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 on model approaches, exemplars and rubrics that help to embed FAIR data practices in different settings.*

Funder guidance on DMPs

Funders can influence the data management practices of researchers and institutions they find, by specifying their requirements for a data management plan that identifies how practices will align with FAIR data principles. According to the policy analysis exercise undertaken in D3.1, the majority of funders analysed require the development of a DMP as part of grant applications. For instance, in its policy for funded projects, the Austrian Science Fund states: “All research data and their metadata should be findable, accessible, interoperable and reusable (fulfil the FAIR Principles)...”. Furthermore, they provide

FAIRsFAIR “Fostering FAIR Data Practices in Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

guidance on recommended repositories, licensing procedures and persistent identifiers for citation.[57]

Similarly, the Open Research Data policy for the Swiss National Science Foundation (SNSF) requires the submission of a DMP. Furthermore, it provides many resources for researchers interested in its Open Research Data policy, for example, an extensive DMP guideline document which contains a checklist to identify FAIR-compliant repositories. [58] The FAQ section of the SNSF policy notes that a submitted DMP "does not undergo any scientific evaluation and therefore does not influence the rating of a proposal". However, it does make clear that DMPs can be revised and updated throughout the research project (see section 3.2.2 DMPs as working machine-actionable documents) and that the DMP will be made publicly available through its P3 database of funded projects on project completion.[59]

Supporting researchers' engagement - tools and training

The Skills working group of the Open Science Policy Platform, in their survey of over 1200 researchers found that only a quarter had used a DMP. One third had not used a DMP but would like to, while a quarter did not know what a Data Management Plan is. They also report that "early-career researchers are less likely to use a DMP than senior researchers and are more likely than senior researchers to not know what a DMP is and be interested in using one." (O'Carroll et al, p.12)

Researchers applying for research grants may need to understand and deal with a variety of DMP requirements from different funders. And for collaborative projects they may need to deal with conflicting policies of partner institutions. DMP tools can assist in this context, by consolidating in one place the topics that researchers need to address in their plan, and the guidance available. They also enable institutions to provide guidance that takes into account local support services (Jones et al., 2020). Available tools include DMPonline, DMPtool, Data Stewardship Wizard, RDM Organizer, EasyDMP, Research Data Manager (UQRDM), DataWiz, EzDMP, and OpenDMP. These vary in functionality, approaches to deployment (ibid.)

Research support and training at an institutional level play a key role in the uptake of DMPs by researchers, with these services providing an important complement to data management tools and guidance. These typically also offer generic guidance on issues like ethics and data protection, which can be applied from a national or broad-domain level to specific research community practices. For instance, the University of Manchester's data management planning guide contains resources on the provision of potentially sensitive data, with information on "how you can create, store, share and archive data concerned

[57] www.fwf.ac.at/en/research-funding/open-access-policy/open-access-to-research-data/

[58] http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/data-management-plan-dmp-guidelines-for-researchers.aspx

[59] http://p3.snf.ch/

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

with human subjects."[60] These would be relevant to DMPs for projects in any research domains working with personal data.

The Science Europe (2019) "*Practical Guide to the International Alignment of Research Data Management*" identifies a set of necessary criteria for DMPs. The guide maps out core requirements, which include a model for translating these into a DMP template that is consistent with following the FAIR principles. The guide is intended as a minimum standard, "leaving the flexibility to formulate additional guidelines according to the needs of specific domains or to national or local legislation." (ibid. p.17) Science Europe has also led an initiative to develop disciplinary research data management protocols, aiming to consolidate DMP requirements from each discipline (Science Europe, 2018).

These frameworks would allow for the interoperability for data collected within a discipline based on the domain's protocols, while also setting out a standard set of criteria against which funders and RIs could evaluate and monitor DMPs. The Domain Data Protocols *DDPs* were derived from interviews and desk-based research in the respective domains. Table 3.3 lists the domains and illustrates the importance accorded to Research Infrastructures as representatives of the communities featured in the report.

Table 3.3 Science Europe disciplinary protocols: selected domains and communities

Domain	**Community**
Humanities	DARIAH
Humanities: Archaeology	PARTHENOS-ARIADNE
Linguistics: Language Data	CLARIN
Social Sciences: Survey Research	CESSDA
Social and Behavioural Sciences: Psychology	Psychology departments and associations
Social Sciences: Ageing Studies	SHARE, TILDA
Life Sciences: Bio-informatics	ELIXIR, FORCE11/RDA FAIRSharing
Plant Science	ERA-CAPS (former WG on RDM)
Climate Research	ICOS

Some of the DDPs have been incorporated into DMP templates, for example the PARTHENOS project template.[61] This includes criteria for making data FAIR. PARTHENOS has

[60] https://www.library.manchester.ac.uk/using-the-library/staff/research/research-data-management/planning/

[61] http://www.parthenos-project.eu/portal/dmp

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

also separately produced guidelines to FAIRify data,[62] in 20 steps that refer to the need for discipline-specific protocols for interoperability and machine-actionability. Also within the social sciences and humanities domain CESSDA[63] has produced a data management guide for use by social science researchers, with community-wide input from CESSDA's 11 partner organisations. One of these partners, the Austrian Social Science Data Archive, has produced their own DMP template. *AUSSDA*'s DMP template is intended to be consistent with the FAIR principles, though the template itself does not go into as much detail on FAIR compliance criteria as its PARTHENOS counterpart.

The ESFRI's Strategy Report from 2018 contains direction for RIs on domain level practices, identifying a role for EOSC in identifying gaps in services and procedures within domains that employ less robust data management, or where there exists less consensus on standard protocols. The utility of domain-specific practices within communities is reflected in the extent to which organisations have attempted to formulate guidelines on common DMP criteria.[64]

Example of good practice: RDM training mandatory for graduate students at the University of Glasgow
Data management training is mandatory for research students in the Sciences and strongly recommended in the Arts and Social Sciences. As of 2018/19, it is mandatory for all research students to present a data management plan as part of Annual Progress Review. The University of Glasgow anticipates that by writing a data management plan for their of Annual Progress Review, students will become more aware of good research data management requirements and practice. *University of Glasgow's Annual Statement on Compliance with the Concordat to Support Research Integrity (2018–2019)*[65]

3.2.2 DMPs as working machine-actionable documents

Turning FAIR Recommendations:

- *r22.4. DMPs themselves should conform to FAIR principles and be open where possible.*
- *r22.5. Information gathered from the process of implementing and evaluating DMPs relating to conformity, challenges and good practices should be used to improve practice.*

[62] https://zenodo.org/record/2668479#.XdUHor_gp-V
[63] https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide
[64] http://roadmap2018.esfri.eu/strategy-report/the-evolving-role-of-research-infrastructures/
[65] https://www.gla.ac.uk/media/Media_683476_smxx.pdf

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

The EC Horizon Europe programme will require projects to complete a DMP and make it publicly available (DG Research and Innovation, 2019). This requirement is a step-change from current practice, where DMPs have been published as end-of-project deliverables from projects funded under the Horizon 2020 Open Data Pilot, e.g. in CORDIS. Other funders have not made public the DMPs of projects they fund. Most of those requiring a DMP, demand its submission at the pre-award stage. DMPs are treated with the same level of confidentiality as a grant application, i.e. typically shared with peer reviewers but not more extensively.

There have nevertheless been efforts by researchers and service providers to share DMPs publicly. Some journals are beginning to accept DMPs as article types, for example, RIO journal[66] and BMC Research Notes[67], and the former has published a small number. Some users of DMP authoring platforms have chosen the option to make their plan publicly accessible (see, e.g. DMPonline). Others have done so by depositing in generic repositories, such as Zenodo and Figshare. A DMP Catalogue has been established by LIBER (see box). This aims to address one of the main use cases for sharing DMPs, to help establish good practice in planning and the execution of those plans (Simms et al., 2017).

These early steps towards sharing DMPs suggest tentative acceptance of these plans as research outputs in their own right, or as documentation of steps to make data FAIR. They have been the focus of two RDA working groups, under the umbrella of an interest group 'Active DMPs'. The working groups have addressed two main issues hampering implementation and adoption.

- A lack of standards for expression and interchange of DMPs that would render them FAIR
- Insufficient understanding of user needs, and the benefits and risks in making DMPs accessible to stakeholders to meet those needs

The first of these issues has been addressed by a DMP Common Standards Working Group, by defining a set of terms describing DMP content. These take the form of an application profile, which has been proposed to the RDA as a standard for exchanging machine-actionable DMPs. The second issue is addressed by an Exposing DMP Working Group. The group aims to provide a reference model and use case catalogue of examples. Survey results from this group include respondents' degree of support for a range of reasons to expose DMP content to parties other than a DMP author's direct colleagues and the relevant funder (Myers, 2019). The strongest support was for demonstrating transparency, followed by integration of DMPs into research workflows, and to assist service providers in

[66] RIO Journal: https://riojournal.com/browse_journal_articles.php?form_name=filter_articles&sortby=0&journal_id=17&search_in_=0&section_type%5B%5D=231

[67] BMC Research Notes: https://bmcresnotes.biomedcentral.com

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

supporting implementation of the plans. Respondents frequently wanted some degree of selective exposure (see Figure 3.1).

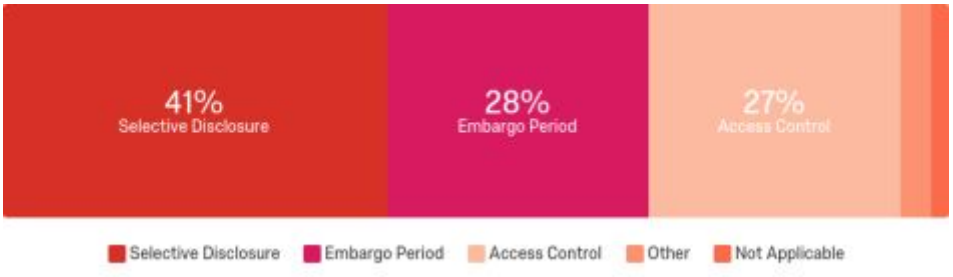

Figure 3.1 What conditions and guarantees should be considered when deciding to expose DMPs? Source: Myers et al (2019).

Efforts are being made to support funders and institutions work with researchers to plan for FAIR data. For instance, the FAIR Funders Implementation Study [68] is a GO-FAIR initiative to enable funders to make it easier to require FAIR-compliant, machine-readable DMPs as part of application requirements, and to develop guidelines allowing applicants to respond to these requirements.

Example of good practice: LIBER Data Management Plans Catalogue
To highlight good practice in providing evaluated examples of DMPs, the LIBER Research Data Management Working Group created a DMP catalogue. It contains DMPs of finished projects from a variety of European states, universities and disciplines. The working group members reviewed those DMPs, providing an assessment according to an established review matrix.[69] The feedback is provided by data librarians or data stewards, and is independent of any funding bodies. Shortly after announcing the DMP catalogue, the LIBER group also ran a survey to gather feedback on how the catalogue can be improved and be more helpful to researchers and data stewards. Respondents asked for the catalogue to be enriched with DMPs in French and examples from more disciplines, especially the Arts and the Humanities. Including DMPs in a machine-actionable form via for example RIOjournal and DMPonline is listed as a potential feature in the future.[70]

[68] https://arxiv.org/abs/1902.11162v2

[69] http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/H2020%20DMP%20compliance%20rubric.pdf

[70] https://libereurope.eu/dmpcatalogue/plans/

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

3.3 Managing Incentives, Costs and Rewards

Table 3.4. FAIR incentivisation, costs and rewards

Relevant FAIR data practices	Scope and resource, Advise and enable, Plan and design

Turning FAIR Into Reality- Action Points (edited)	Phys. Science Eng.	Energy	Environ -ment	Health/ Food	Social Cultural	Data Comp. Digital
Prioritising and incentivising reuse						
Managing costs and rewards for FAIR effort						

Key: FAIR initiatives by stakeholders are: Rare or missing, low take-up Growing, patchy take-up Common, extensive take-up Information unavailable

3.3.1 Prioritising and Incentivising Reuse

Turning FAIR Recommendations

- *r21.1: Researchers – including graduate students – should be required to demonstrate in research proposals and DMPs that existing FAIR data resources have been consulted and used where appropriate, before proposing the creation of new data.*
- *r21.2: Research funders and the academic reward system should ensure that research that reuses data and other outputs is valued as highly as research that creates new content.*

The recommendations above are primarily directed towards top-down action by funding bodies and institutions, but require take-up by research communities to be effective.

Some funding bodies make clear that they expect researchers to have checked that the data they plan to generate does not already exist. For example, the Economic and Social Research Council ERSC policy requires researchers to provide an explanation of the existing data sources that will be used, with references and an analysis of the gaps identified between the currently available and required data.[71] They also guide peer reviewers on what to look for in the DMPs submitted with grant proposals[72]. The Netherlands

[71] https://esrc.ukri.org/files/about-us/policies-and-standards/esrc-research-data-policy/

[72] https://esrc.ukri.org/files/about-us/policies-and-standards/data-management-plan-guidance-for-per-reviewers/

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Organisation for Health Research and Development ZonMw also expects researchers to address the comment "I will be reusing or combining existing data, and I have the owner's permission for using or combining their data." as part of their DMP.[73]

The Open Research Funders Group has published an Incentivisation Blueprint[74] "to more closely align with open access, open data, open science, and open research." This recommendation lays out three stages for funders to embed incentivisation for open research in funding protocols. These stages include:

1. Steps for incorporating incentivisation into funder policies and guidelines;
2. implementing these policies;
3. engaging funded researchers in incentivisation and research assessment procedures.

Funding bodies could do more to reward data reuse through funding calls that focus on secondary data analysis. However, outside of funding organisations and in domains where data reuse is common, there have been some incentive initiatives emerging in recent years in the form of one-off prizes, awards, and open calls. In the Social Sciences, the ESRC has offered its Secondary Data Analysis Initiative SDAI[75] open call since 2015. The initiative was updated in 2018 to increase the maximum funding available from £200k to £300k and to extend the maximum duration of funded projects from 18 to 24 months. While SDAI was originally introduced to 'deliver high-quality high-impact research through utilising existing data resources created by the ESRC' the 2018 update also extended the scope of the data eligible for reuse to include 'a range of UK and international data resources funded by ESRC and by other agencies'.

In the Life Sciences, UK charitable trust Wellcome introduced its Open Research Fund[76] in 2018. The call provides up to £50,000 for projects lasting up to one year in length to support "innovative ways of making health research open, accessible and reusable". In addition, the Wellcome also inaugurated its Data Re-use Prizes in November 2018, with awards of up to £15,00 "to reward new insights and tools that help other researchers to re-use data."[77] Also in the domain of medical science, the QUEST Berlin Institute of Health[78] offers an annual €1,000 prize for a publication which is based on existing data available from public repositories. In the US context, the Research Parasite Awards is an annual prize awarded to applicants who can demonstrate outstanding contributions to the rigorous secondary analysis of data. [79]

[73] https://www.zonmw.nl/fileadmin/zonmw/Instruction_ZonMw_DMP_Eng-__okt_2017_02_def.pdf
[74] www.orfg.org/incentivization-blueprint
[75] https://esrc.ukri.org/funding/funding-opportunities/secondary-data-analysis-initiative-sdai-open-call
[76] https://wellcome.ac.uk/funding/schemes/open-research-fund
[77] https://wellcome.ac.uk/news/new-data-re-use-prizes-help-unlock-value-research
[78] https://www.bihealth.org/en/research/quest-center/initiatives/quest-open-data-reuse-award/
[79] https://researchparasite.com/

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

When it comes to incentivising researchers to manage and share their data, the State of Open Data 2018[80] report found that almost two-thirds of respondents were in favour of national mandates relating to open data. This figure has increased to 79% of respondents in the recently released 2019 State of Open Data report.[81] The survey also provides information on the reuse of existing data. Of the nearly 1,300 researchers who responded, 33% said they were 'Extremely likely' to reuse open data in their research in future, while 38% said they were 'Somewhat likely'. However, nearly one-third of the 'Extremely likely' group and nearly a quarter of the 'Somewhat likely' group identified the 'Reusable' element of the FAIR principles as the one with which their research was least compliant.

Following from this, while mandates are an important factor in driving cultural change the findings of the FAIRsFAIR open consultation highlight the need for an appropriate balance between penalties and rewards. Rewards for making data FAIR were regarded by more respondents as a positive influence on researchers' practice than were penalties. Just under half (43%) of respondents from the open consultation indicated that the introduction of penalties for not making data FAIR would have a very/quite negative influence on practice.

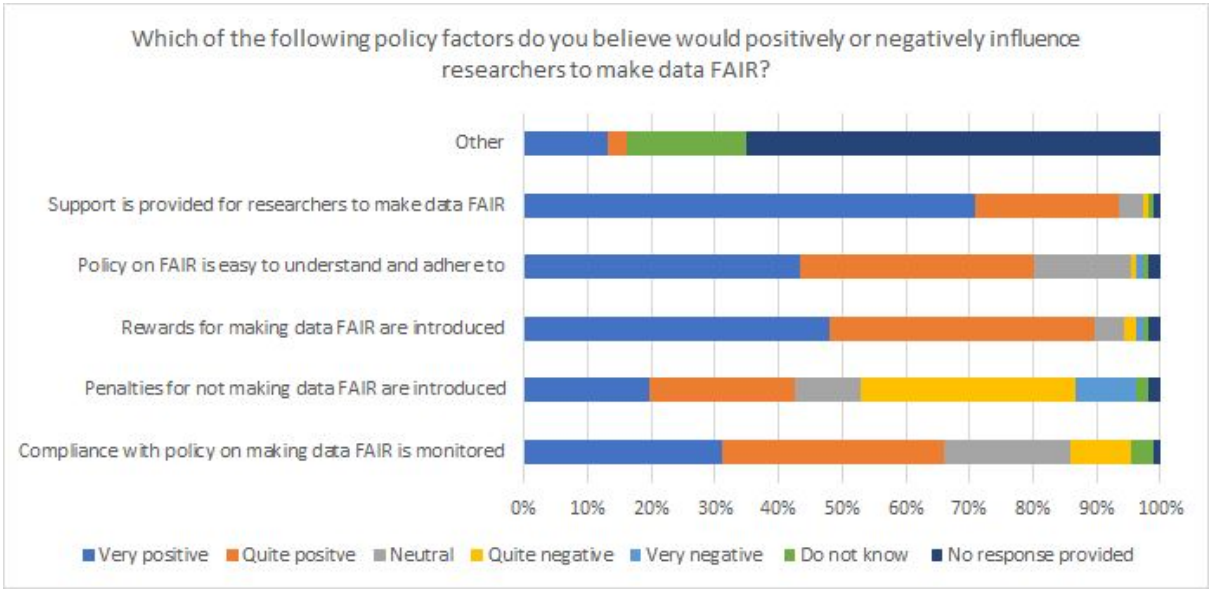

Figure 3.2. Influence of policy factors on researchers' practice, FAIRsFAIR open consultation on Policy and Practice

Wellcome updated its Open Access policy in May 2019 to include an expectation that "Wellcome-funded organisations must sign or publicly commit to the San Francisco Declaration on Research Assessment DORA, or an equivalent. We may ask organisations to show that they're complying with this as part of our organisation audits. This is a new

[80] https://digitalscience.figshare.com/articles/The_State_of_Open_Data_Report_2018/7195058
[81] https://digitalscience.figshare.com/articles/The_State_of_Open_Data_Report_2019/9980783

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

requirement to encourage organisations to consider the intrinsic merit of the work when making promotion and tenure decisions, not just the title of the journal or publisher".[82]

Wellcome introduced this aspect to their OA policy to demonstrate their commitment to ensuring that the funding decisions made are based on the "intrinsic merit of the work and not the title of the journal or publisher"[83]. The policy comes into effect on January 1, 2021, and may well have a significant influence on HEIs publicly adopting a commitment to change the way they assess research outputs. The European University Association EUA and many individual institutions and institutes have signed DORA.

3.3.2 Managing costs and rewards for FAIR effort

Turning FAIR Recommendations:

- *r18.2: 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.*
- *r6.2: 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.*
- *r6.3: 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 the evaluation of research contribution has a legitimate role to play. This evidence should include assessment of graduate students.*
- *r6.4: Contributions to the development and operation of certified and trusted infrastructures that support FAIR data should be recognised, rewarded and appropriately incentivised sustainably.*

Uncertainty over who pays for the research data management effort to make data FAIR, and to resource the infrastructure to keep data FAIR, has been a persistent problem for researchers, institutions and funders. According to a 2016 Knowledge Exchange study "RDM, although recognised as important, is generally not (yet) regarded as a fundable part of the standard research process. The specifics of RDM and the budget scope for funding RDI are usually not clearly defined. The funding is not well connected to specific RDM requirements at different stages in the research process/data lifecycle." (Bijsterbosch et al, 2016).

The Knowledge Exchange-Science Europe survey of research funding organisations, which the 2016 report is based on, found that RDM is generally funded indirectly without any clear budget. It argued that "RDM activities and resulting costs should be considered to be part of the costing breakdown in research funding programmes." There is some evidence that

[82] https://wellcome.ac.uk/news/wellcome-updates-open-access-policy-align-coalition-s

[83] https://wellcome.ac.uk/sites/default/files/wellcome-open-access-policy-2021-faq.pdf

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

funders' positions are changing to allow data management as an eligible cost on grants (as for H2020, for example). The small sample of funder policies reviewed for FAIRsFAIR D3.1 FAIR Policy Analysis found that most accept such costs in project budgets.

Funding processes are however considered poorly adapted to deal with data, both in terms of project-level resourcing for RDM and post-project (or trans-project) funding for infrastructure. Infrastructure funding has been considered an acute problem for research funding organisations and institutions to address, due to the difficulties in sustaining infrastructures from successive short-term projects. This problem has been a focus of major investment at National and European level, as the Turning FAIR report notes. However, it also points out that the "so-called long tail of research remains poorly catered for, and vast amounts of data produced in research are neither FAIR nor stewarded for long-term preservation and access". Project level funding of RDM is not consistently available and, when it is, there is a lack of transparency in how costs may be allocated.

Funding rules prevent institutions from counting the costs of data management support twice by charging them to grants under both direct and indirect cost categories. Researchers are therefore not equipped with either the knowledge or the means to budget for, or control data management costs. Centralised institutional services such as libraries that are normally funded through institutional overheads are therefore limited in their ability to create services that match provision to uncertain demand. For institutions that do not have substantial research income, it may be impossible to justify the risk of building a central RDM support service. Similarly, at research group or institute level, building local RDM support is only likely to be affordable for those that already have a level of income from data-intensive research to justify creating a specialist data support role.

To counter such challenges, disciplinary RDM expertise across institutions can be combined, so that researchers with discipline-specific questions can be contact the staff with the appropriate data management expertise. There are a number of recent initiatives trialling this approach including the Data Curation Network *DCN* [84] in the US, the Dutch National Coordination Point Research Data Management *LCRDM,*[85] and the Swedish National Data Service network.[86]

TU Delft has adopted a similar but institutional-level approach.[87] Senior level support has created a central data steward coordination role and recruited a team of data stewards employed by faculties. Each faculty has a dedicated Data Steward to advise on domain-specific issues. In its Research Data Framework policy, it puts the liability on its data stewards to "help researchers with writing data management plans and with budgeting for

[84] datacurationnetwork.org/

[85] www.surf.nl/en/national-coordination-point-research-data-management

[86] snd.gu.se/en/about-us/snd-network

[87] openworking.files.wordpress.com/2018/06/tu-delft-research-data-framework-policy-version-for-cvb-18-june-2018.pdf

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

research data management costs in their grant applications", while principal investigators in research projects are expected to include the costs of data stewardship into their projects. TU Delft has also developed at Data Management Costing Tool[88] for its researchers which, based on a few questionnaire responses, recommends elements to include in project costings.

Turning to the rewards for effort spent on FAIR enabling activity, the Contributor Roles Taxonomy *CRediT* is an initiative that attempts to secure greater recognition. It leverages traditional citation measures for data management and other tasks that underpin the authorship of research outputs, but may otherwise go unrecognised.[89] CRediT is developed by standards organisation CASRAI to enable authors submitting any scholarly output to a journal to clearly articulate the various contributions made. The taxonomy currently identifies 14 roles from conceptualising to writing up the work and has been adopted by almost 30 publishers. In 2017, the University of Glasgow was the first research institution to refer to CReDIT in its Code of Good Practice in Research[90] to clarify authorship and contributions to publications. Adopting CReDiT at the University of Glasgow is part of a wider effort to change research culture that also includes introducing promotion criteria that reward collegiality and open research.[91]

Considering the reward system in Europe more broadly, the European Commission's Open Science Working Group on Rewards/Recognition examined the extent to which researchers are currently recognised for the openness of their research, and how this may be improved. [92] With a focus on long-term, career-orientated recognition of Open practices, the working group devised a draft Open Science Career Evaluation Matrix. '[93]

Follow-up work has been carried out in the EC's Expert Group on Indicators which has recently produced recommendations on methods to evaluate research quality outside of the traditional 'established journals' measurement.(Wouters et al, 2019). The report emphasises the importance of assessing 'open knowledge practices' rather than the openness of research outputs per se. It calls for the development of indicator toolboxes, each composed of qualitative and quantitative indicators. They fall into the following four categories : -

1. Research Infrastructure creation, growth, nature of their contribution, and their use and uptake, with a particular focus on EOSC.
2. Open knowledge capabilities in research communities and their support personnel: indicators of resource availability in specific communities, success cases and

[88] zingtree.com/host.php?tree_id=511095771

[89] https://casrai.org/CRediT/

[90] https://www.gla.ac.uk/media/Media_490311_smxx.pdf

[91] https://www.gla.ac.uk/myglasgow/ris/researchculture/#d.en.649279

[92] https://ec.europa.eu/research/openscience/index.cfm?pg=rewards_wg

[93] https://ec.europa.eu/research/openscience/pdf/os_rewards_wgreport_final.pdf

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

measures needed to increase the inclusiveness, diversity and equity of the research system.

3. Pioneering open knowledge practices, identified through case-studies maintained and regularly updated on a public platform, highlighting pioneering open knowledge practices, and modelled on the UK Research Excellence Framework.
4. Individual level indicators for career, based on the Open Science Career Evaluation Matrix outlined in the previous report.

Importantly each of these categories takes into account the role of infrastructure and support roles in the career assessment of researchers.

The European Universities Association has recently reviewed current university approaches to research assessment in the context of open science. The results of its 2019 EUA Open Science and Open Access Survey on Research Assessment indicate that publication-based metrics continue to dominate, with only occasional mention of data in indicators (Saenen et al 2019). According to their analysis, Open Science and Access indicators were only "important" or "very important" for research career assessment to 28% of respondents. Moreover, "the open accessibility of research publications and data is often only monitored at institutional level, and is not part of incentive and reward structures for individuals." Nevertheless, the EUA report also says that "...virtually all of the responding institutions are reviewing their approach to research assessment. Most responding institutions indicated that they will incentivise and reward a broader range of academic activities in future".

The association of universities in the Netherlands, VSNU, offers an example of coordinated action with national-level funding organisations, beginning with a recognition of the need for change. Their statement calls for "concrete and meaningful steps to bring about a necessary transition....Towards a system where Open Science is the standard, not the exception." (VSNU, 2019)

The recent State of Open Data report (Figshare, 2019) found that researchers still see citations as the 'holy grail in terms of reward'. Extension of current research assessment frameworks to reward researchers for a broader range of outputs is crucial that data reuse can be tracked As the FAIRsFAIR D3.1 report points out there is a need for stakeholders across the FAIR ecosystem to agree and promote standards for data citation. The EUA report (Saenen, 2019) that only 48% of universities consider 'other types of research output', including research data, as "important" or "very important" for the assessment of researchers. This result compares with 90% of HEIs that consider research publications to be important or very important.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Example of good practice: measuring and rewarding FAIR practices
The University of Bristol will include open research practices for use from the 2020-21 promotion cycle. Open access publication, including in Bristol's institutional repository, has been required for some time. Emphasising that open research is more than simply open access, the new criteria recognise: "Producing open research outputs as appropriate by adopting good practice in, for example, sharing data and code, sharing materials, sharing digital outputs, publishing preprints and pre-registering study protocols." (Munafo, 2019)

3.4 Selecting and Depositing FAIR Outputs

Table 3.5 FAIR data selection, deposition and repositories

Relevant FAIR data practices	Appraise and preserve, publish and release, expose and discover

Turning FAIR Into Reality- Action Points (edited)	Phys. Science Eng.	Energy	Environ -ment	Health/ Food	Social Cultural	Data Comp. Digital
Guide and document selection decisions						
Build repository community capacity and capabilities						

Key: FAIR initiatives by stakeholders are: Rare or missing, low take-up Growing, patchy take-up Common, extensive take-up Information unavailable

3.4.1 Guiding and documenting selection decisions

r19.2: The appraisal and selection of research outputs that are likely to have future research value and significance should reference current and past activities and emergent priorities. Established archival principles and the importance of unrepeatable observations of natural and human phenomena should be taken into account.

r19.3: When data are to be deleted as part of selection and prioritisation efforts, metadata about the data and about the deletion decision should be kept. If data deletion is carried out routinely, the underlying protocols for selection and prioritisation need to be made FAIR.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Decisions on selection of research outputs for FAIR treatment involve two parties, not necessarily working in tandem: the researchers themselves who generate the data, and the archives or repositories that undertake to curate it.

Researchers might find themselves facing what seem difficult decisions on the selection of data. They need to consider, for example, how fundamental it is in underpinning their research results, and whether it is likely to interest others. Practices and judgements will vary by domain and by the nature of the research, but it is clear that there are risks of divergences of individual opinion. From the repository's perspective, there are considerations of scope and sufficiency. Appraisal and selection must take into account the need for data whose absence would reduce the value of the repository's holdings. Sufficiency refers to the need for the selected data to be understandable to the Designated Community (in terms of the OAIS Reference Model). These can both be seen as positive requirements on what must be selected, translating into what the repository expects from the depositors.

The selection decision is based on the relationship between the (potential) depositor and the repository, and on their exchange of knowledge. Researchers need to be aware of the implications of selection decisions, while repositories must take account of researchers' views of data value and the trade-offs in providing it.

The process of identifying what should be kept has grown in significance in recent years as the volume and diversity of research data have grown, and as the available infrastructure for managing research data has become more diverse. Beagrie (2019) provide useful insights in the 'What to Keep' report. This makes ten recommendations aimed at funders, repositories, UK Higher Education Institutions, learned societies and publishers, on data selection issues.

The report identifies research integrity and reproducibility, and the potential for reuse as two major use-cases for keeping research data. Although the use-cases can overlap, it is crucial to recognise that they are distinct. Different types of data may need to be kept to support them. The report notes that different disciplines may also have different reuse cases or derive different value from similar data. Nevertheless, a broad consensus has emerged around high-level generic criteria that are now being applied in multiple domains. This suggests there are examples of effective practice that can be promoted to others. One such set of criteria are described in the NERC Data Value Checklist (below). The report notes that this checklist is specific to the environmental domain, but has been influential in the development of checklists for other repositories and disciplines.

A key recommendation of the report was to bring communities together to evolve disciplinary norms for what research data to keep, where these norms are currently absent

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

or evolving. Though not specified in the report, there would seem to be a key role here for scientific unions, societies and professional associations to help develop these norms for communities that lack them.

Example of Good Practice: NERC Data Value Checklist
The Data Policy of the UK's National Environment Research Council (NERC) requires environmental datasets of long-term value to be submitted to one of the NERC Environmental Data Centres. NERC developed its Data Value Checklist to help NERC funded researchers select this data.[94] Individual Data Centres have collections policies that help decide which Centre is the appropriate place to deposit, depending upon the science area and type of data collected. The Checklist guides but does not determine the decision on long-term value. The checklist is intended for use in developing a project's 'full' data management plan. NERC requires a 'full' data management plan to be produced within three to six months of the start date of a grant. The checklist identifies three sets of criteria for retention – mandatory, important and supporting. If data meets any of the mandatory criteria (e.g. if there is a legislative requirement to keep the data), it will automatically be retained. If data meet at least one of the important criteria (e.g. are the data a unique, unrepeatable measurement of the environment), or if the majority of supporting criteria (e.g. would the data be costly to reproduce) are met then it will probably be selected for retention.

3.4.2 Building repository capabilities

Turning FAIR recommendations

- *r20.3: 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.*
- *r20.4: Outreach is required via scholarly societies, scientific unions and domain conferences so researchers in each field are aware of the relevant disciplinary repositories.*

Repositories manage access to valuable data and metadata and offer services to support access and reuse. Data stewardship and making data FAIR is often beyond the capacity of individual researchers, small teams and most research laboratories. The specialisation and

[94] https://nerc.ukri.org/research/sites/data/policy/data-value-checklist

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

expertise required means that research communities rely on (disciplinary) data repositories to support these activities.

Different repositories offer different levels of stewardship. Generic repositories often rely on user-entered metadata, which may not meet exacting standards of FAIRness. Disciplinary repositories play a key role in the provision and preservation of FAIR data since they pool relevant domain expertise, should implement community standards and may provide quality long-term stewardship and curation. Researchers are recommended to use domain repositories where they exist, and preferably certified repositories (Hrynaszkiewicz et al, 2017). Generic repositories are recommended where they provide a specific service that adds value to the data, which is not available in a relevant domain repository (such as linking the data to a publication), or where there is no relevant domain repository available (Whyte, 2015).

Although the FAIR principles apply to data, their implementation requires several data services and components to be in place in the broader ecosystem that enables FAIR. These services should themselves be FAIR where applicable. Hereinafter, we consider the case of data repositories and services necessary to the FAIR data ecosystem.

The FAIRsFAIR policy and practice consultation included questions to research support staff on repository provision to the communities they work with. The 106 respondents (whose roles are described in Annex) responded on topics including

- which stakeholder groups are providing repositories
- gaps in provision
- community expectations about repository certification, self-assessment, and user feedback ratings

Stakeholders providing repositories

Asked about which stakeholder groups currently provide a service to those communities, most respondents identified institutions and community databases/repositories (both 71%) and research infrastructures (54%). Some (42%) respondents indicated journals provided a repository, and 19% identified funders as direct providers. A number of respondents to the 'other' option mentioned Zenodo or Figshare as preferences, suggesting that some users of these services see them as independent of RIs or publishers (respectively).

There is some contrast between these responses and those of previous surveys where respondents have been asked where researchers data is shared. An Elsevier-backed survey of researchers by Wouters and Haak reported that only 13% of respondents used a data repository to share data publicly, 33% used an appendix to a publication, and 28% a data

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

journals. It is possible that respondents to publisher-backed surveys tend to be more inclined than non-respondents to deposit in publisher-backed repositories.

Gaps in provision

Most respondents (60%) believe there are gaps in repository provision to researchers. Only 6% indicated that they see no gaps. However, there were also a large number of 'don't knows' for this question, suggesting a need for further work to investigate researcher demand for domain-specific support.

Invited to comment on specific unmet needs, respondents listed 58 of these (see Annex 2). Examples included:

- High-quality direct-to-consumer DNA test results which people are sharing and would also to donate for R&D&I
- Many types of chemistry data lacking an appropriate repository where appropriate means expertly curated
- Users are diverse and work across many photon science facilities; federated and common cloud infrastructure would be very helpful - but has to be engineered to match the workflows and fast turnaround demands of experimenters
- There doesn't seem to be a repository dedicated to Engineering
- Lack of repositories for big data and sensitive data outside of the social sciences - especially health sciences and commercially sensitive data, support for software is limited
- Wearable/sensor data repository
- Size limits on deposits problematic for those handling large datasets e.g. medical imaging.
- Dealing with GDPR requirements means we need processing agreements with all repositories. That's not a simple issue to solve.
- Not a suitable repository for clinical research data is available; the ones existing are either disease-specific or faculty/institute-based, with different access modalities

Analysis of the comments identified the following common themes: Lack of domain-specific repository, curation, or quality control (18 comments) Data size limitations (11) GDPR, confidential data handling (5) Other legal (2) Data type limitations (4) Other (39)

Expectations of repository trustworthiness

Respondents were asked, "to what extent do stakeholder groups in the communities you work with expect data repositories to follow 'trustworthy' repository standards, e.g. CoreTrustSeal?" They were asked to identify how far on a five-point scale, they agreed with the options shown in Figure 3.3. They could respond to each option independently.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

The responses suggest greater support for 'community guidelines' than for either self-assessment against a common standard or third-party assessment and certification. The 'other' option elicited the following comments:-

- Repositories need transparency with regard to their own financial and organisational sustainability
- Communities are hardly aware of CoreTrustSeal. However the university policy encourages using certified repositories
- Typically, the stakeholders lack knowledge of what ''trustworthy' repository standards' are and usually care more about technical standards (e.g. https) and user experience (e.g. upload API).
- The role and expectations of citizens. DNA data & databases due to both big national etc initiatives & DTC DNA tests. Citizens expect to be treated as partners (like in All of Us initiative), they want to be informed etc.
- The standards themselves need to be trustworthy
- There is a trend towards certification, but many repositories without certification function very well within data infrastructures.
- Researchers do not want to be constrained by procedures related to infrastructure certification
- Core Trust Seal is not good enough: the requirements of institutions in terms of administration (CRIS requirements, building a base for recognition) are generally not met at all!!
- Quality of repositories should be assessed
- Alignment with community guidelines or best practices should be easily assessable by anyone
- Certification needs to be machine-testable, not documented in prose as currently. Infrastructure-as-Code using Ansible & similar tools.
- Self-assessment can only be the first step to build trust independent of personal knowledge of the repository and their staff, and these certificates need to get known in the research community.
- Core trust seal does not really play a role in trustworthiness in the life sciences
- The Researcher stakeholder is often unaware of certification; the Funder stakeholders find certification relevant but are not very knowledgeable about how many (how few) repos are certified in the various domains.

These responses show that CoreTrustSeal faces some barriers in raising awareness, and needs to win support from stakeholders outside of its parent domains of social science and earth sciences. Some respondents favour alternatives to certification as a means of demonstrating trustworthiness.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

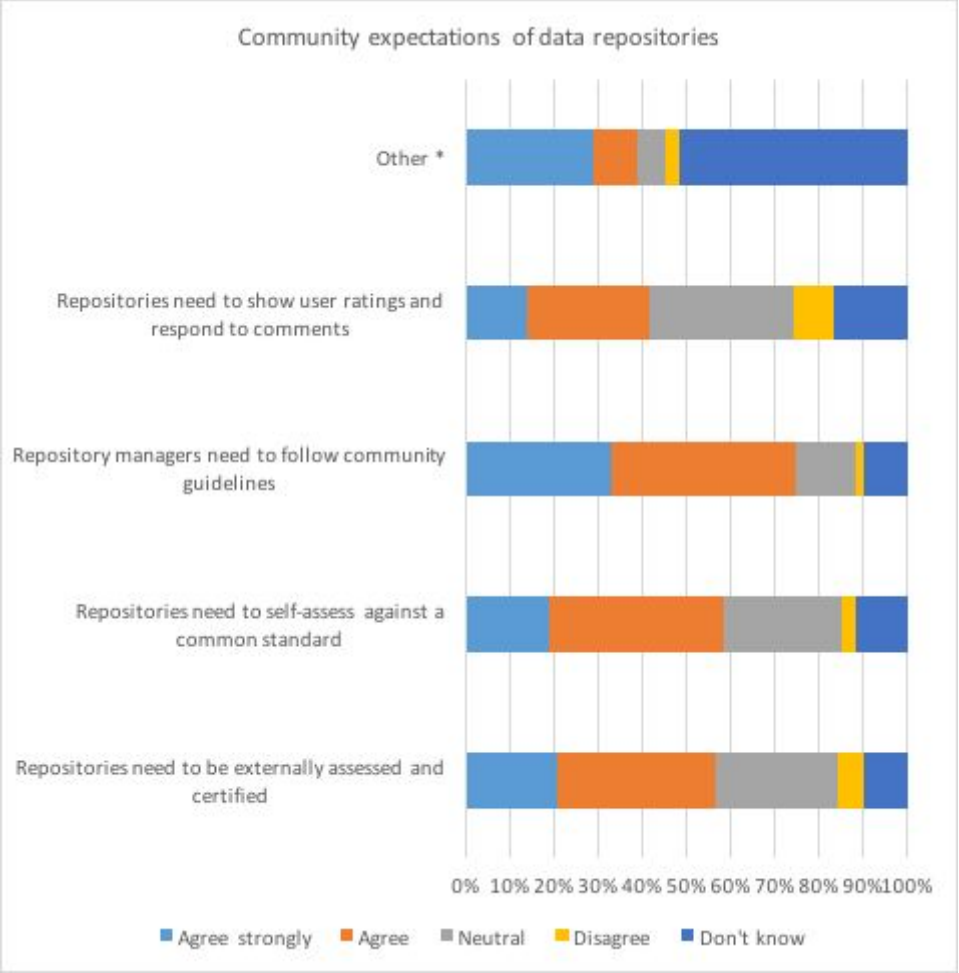

Figure 3.3 Consultation responses on expectations of repository trustworthiness

Journal and publisher stimuli for domain repository deposition

There has been some convergence across the funding, research and publishing communities to promote greater use of domain repositories for FAIR and open data, particularly for deposition of research data underlying published articles. These measures include developing guidance on the following:

- for journal reviewers, on the peer review of data and code
- for authors on choosing repositories to host data underlying articles submitted for publication
- for authors, on providing a 'data access statement' identifying how any data supporting their article may be accessed

On the first point, Springer Nature has since 2016 led an initiative to harmonise journal policies on research data deposition. The resulting research data policy framework (Hrynaszkiewicz et al. 2017) divides data policy for publications into four types, as follows:

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- Type 1 policy encourages data sharing and data citation and provides researchers with a list of data repositories
- Type 2 provides information on preparing data availability statements.
- Type 3 includes mandatory data availability statements
- Type 4 requires open data and requires peer reviewers to access data supporting publications.

The re3data global registry of research data repositories is frequently referenced in RDM guidance as a means of identifying domain repositories for sharing and reusing research data.[95] The registry service is maintained by the Humboldt University, Berlin, the GFZ German Research Centre for Geosciences, the Karlsruhe Institute of Technology (KIT) and Purdue University. The front-end of the online service provides faceted search of repository metadata.

Metadata details are described using the re3data.org schema, a list of metadata properties covering a research repository regarding its general scope, content and infrastructure as well as its compliance with technical, metadata and quality standards. The schema includes required metadata properties and optional properties providing additional information (Vierkant et al. 2014).

Several initiatives have achieved a level of consensus among publishers and stakeholder groups on suitable criteria for recommending repositories to researchers. One of these is the COPDESS and AGU initiative 'Making data FAIR' (see box below). This led to collaboration with the DataCite re3data service to develop a repository finder service, which is being enhanced in FAIRsFAIR WP4. Another is the recent collaboration between the FAIRsharing initiative, DataCite and a group of publishers. The proposed criteria are intended to complement certification standards and to offer the community a basis for assessing repositories that have yet to undergo certification.

The Belmont Forum offers an example of an initiative to harmonise guidance on data accessibility statements *DAS*.[96] The forum is an international partnership of environmental funders and has recently delivered a DAS template to guide grantees when publishing their research results. A DAS requirement encourages researchers to plan for the longevity, reusability, and stability of the data attached to their research publication and results. Additionally, the DAS offers opportunities to credit data collectors and curators by supporting data citation.

[95] http://www.re3data.org/
[96] http://www.belmontforum.org/

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Example of Good Practice: Coalition on Publishing Data in the Earth and Space Sciences

The Enabling FAIR Data initiative effort builds on a 2014 *Statement of Commitment* by the Coalition on Publishing Data in the Earth and Space Sciences (COPDESS). Enabling FAIR Data provides recommendations and guidelines for implementing a research data ecosystem for these domains. [97]

The earlier COPDESS statement identified best practices and goals for journals and repositories.[98] It committed journals to make available the data supporting published conclusions, encouraging data deposition in domain repositories. In turn, domain repositories committed to develop practices that would support data availability and quality. Repositories were also committed to working with publishers on the infrastructure for data curation and integrity in scholarly publishing.

[97] http://www.copdess.org/enabling-fair-data-project/enabling-fair-project-overview/

[98] http://www.copdess.org/statement-of-commitment/

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

4. Conclusions: Opportunities for Culture Change

The FAIR principles have become widely referenced as cross-domain guidance on data stewardship. As the companion report D3.1 on FAIR Data Policy points out, FAIR principles are the cornerstone of data policies for a growing range of national funders, in addition to the EC's Horizon Europe.

Recent surveys from FAIRsFAIR and others reviewed in sections 2 and 3 of the report indicate stronger awareness of FAIR among institutional and infrastructure service providers than the research communities they serve. The report profiles current activity across disciplines, informed by desk research, interviews and open consultation with the research data support community, in institutions and research infrastructures. This will inform further work to identify and amplify support for FAIR, with the ESFRI cluster projects, related EOSC projects, and providers of data stewardship support to research communities.

There are many researchers and others involved in data stewardship who already make research data findable, accessible, interoperable and reusable without publicly aligning their work to the FAIR principles. The *Turning FAIR* into Reality report's recommendations on actions towards a FAIR culture offer a practical scope for analysing the extent of FAIR-enabling practices. The activities defined in the FAIR4S skills and capability framework provide another dimension for the analysis, to be used in Task 3.3 to link learning resources to examples.

This report also offers an understanding of the factors enabling and hindering FAIR data enabling practices, from our literature review and open consultation. The disciplinary dimension of our analysis is the most challenging. The six broad categories represented in the ESFRI Roadmap have been used, in the interest of aligning FAIRsFAIR activity with ESFRI cluster activity.

A FAIR data practice analysis that rigorously examines research data practices across the full scale and diversity of the European research landscape would be an enormous undertaking. Rather than attempt that, our approach is to identify those aspects of practice likely to help us find useful examples to amplify and support.

Using broad disciplinary categories to characterise FAIR practice is potentially misleading, as these categories do not necessarily translate to the actions of specific communities within the disciplines. Previous studies of data management and sharing reviewed in section 2 indicate the following characteristics of communities where data sharing is prevalent:

- Relatively 'hard science' approaches favouring testable predictions, controlled experimentation, quantifiable data and mathematical models, where a high degree of accuracy, objectivity, cumulativeness, and replicability are valued

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- A high degree of consensus on research topics, standards, methods and techniques
- Large teams working collaboratively, within or across disciplines
- A high ratio of researchers to research problem

Communities with these characteristics do not fit neatly into disciplinary containers. FAIR-enabling initiatives may be cross-disciplinary by design or aim to broaden take-up of an approach from one discipline to others. Other FAIR enabling activities, such as establishing incentives and rewards for FAIR, are driven by action at funder or institution level and tend to be cross-disciplinary for that reason.

With the above caveats, Table 4.1 offers a tentative and qualitative estimation of where action is being taken in line with the *Turning FAIR* recommendations.

The general picture illustrated in Table 4.1 is that it is challenging to assess the level of culture change towards FAIR data production practice based on information gathered in FAIRsFAIR to date. Action points relevant to FAIR implementation may be found across most FAIR data stewardship activities in the Environment, Food and Health, and Social and Cultural Innovation disciplines. Further work is needed to fill gaps in knowledge about FAIR data practices in the Energy domain and in Computing and Digital Infrastructures. This work should take account of these domains requirements for support to make code FAIR and ensure its stewardship.

FAIRsFAIR “Fostering FAIR Data Practices in Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Table 4.1. Overview of the extent of initiatives to support culture change towards FAIR data

Turning FAIR Into Reality- Action Points (edited)	Phys. Science Eng.	Energy	Environ -ment	Health/ Food	Social Cultural	Data Comp. Digital
Usage of domain standards						
Advocacy across domains, and of cross-domain standards						
Respond to funder requirements and enable DMP support						
Put DMPs to work as machine-actionable documents						
Prioritise and incentivise reuse						
Manage costs and rewards for FAIR effort						
Guide and document selection decisions						
Build repository community capacity and capabilities						

There is a need to describe FAIR data practices using a framework that is fit for purpose, i.e. one that can encourage and promote collaborations to implement FAIR, and assist wider adoption of the "good examples". In principle, it would be feasible to analyse domains at a higher level of granularity than the six ESFRI categories. In practice, it has been challenging to find appropriate sources of information to analyse the data practice landscape comprehensively at this level.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Conclusion 1- Develop a self-assessment framework for Research Infrastructures, Institutions and other FAIR competence centres

The CESSDA self-assessment of its actions to address the Turning FAIR recommendations is a commendable example that FAIRsFAIR should assist other RIs to emulate. Research communities' self-assessments of their implementation of FAIR are rather few, although recent work in this area by GEDE and GO-FAIR is also very informative.

Further work is needed to develop a self-assessment framework for research infrastructures and institutions. This should be co-designed to help them, and FAIRsFAIR, identify progress to support FAIR enabling practices in the communities they serve. This will underpin further capability building, promote exchange of good practices and lessons learned, and address the highly uneven availability of information on research community implementation.

The RIs and Research Producing Organisations (e.g. Academic Institutions and Institutes) that are partners in ESFRI cluster and EOSC projects are best placed to assess their FAIR enabling activities. Further liaison will be carried out with RIs, Institutions and other competence centres to co-design the framework, and to work with their target communities to identify how best to meet specific needs for support in implementing the Turning FAIR recommendations. Coordination is also needed to ensure alignment with work on FAIR maturity models for repositories.

Conclusion 2- Build an inventory of FAIR practice

Adopting the view that disciplines are a necessary but not very useful unit of analysis for data practices, section 2 of the report considers several conceptual frameworks proposed in recent literature on research data practices. These include *data communities* (Cooper and Springer, 2019), which emphasise collaborative action around data types. Alternatively the concept of *research repertoire* (Leonelli and Ankeny 2015) refers to well-aligned assemblages of the skills, behaviours, and instruments that a group may use to practice data management and train newcomers. For the purposes of the FAIRsFAIR tasks on embedding culture change, the research repertoire concept may help define an inventory of good practice examples. This should focus on the relationships between skills acquisition for FAIR, the FAIR-enabling activities these skills are applied to, the instruments used, and the organisational capabilities to conduct science based on FAIR data.

As the term 'repertoire' term suggests, an inventory of FAIR data practices would identify the instruments and methods that a group has acquired the competences to perform, sustained by an organisation. Accordingly, an inventory of examples of FAIR practice should cover at least one of each of the following

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

1. Infrastructure provider - including institutional and other providers - open to self-assessing its' capacity and capabilities for supporting a community to implement FAIR principles
2. Community - including data stewards, research software engineers, data scientists and domain researchers - open to learning new competencies to make or keep digital objects FAIR
3. Instruments - including plans, protocols, standards, tools and repositories - open to deployment by a community and its Infrastructure provider(s).

Conclusion 3- Lead by example

The inventory of examples will contribute to Task 3.3, which is to embed FAIR data practices, and to definition of *D3.4 Recommendations for practice to support FAIR data principles*, which is due in May 2020. The work done to date suggests a number of themes to prioritise. The areas listed below are ones the partners have the capability to engage with, in collaboration with projects and communities identified in brackets. Forthcoming work will also engage with the EOSC-5b projects, to obtain a consensus on these recommendations and on co-authoring any guidance and learning resources needed to support them.

1. Methods for building consensus on metadata and interoperability frameworks within and across communities (ENVRI-FAIR, PANOSC, FREYA, SSHOC, GO-FAIR, Funders Implementation Network).
2. Machine-actionable DMP templates and guidelines on using them to inform downstream data management activities (PANOSC, ExPaNDS, GO-FAIR Funders Implementation Network).
3. Using instruments to assign PIDs at the point of data creation (FREYA, ENVRI-FAIR)
4. Terminology for competence centres to annotate and retrieve training materials on enabling FAIR (EOSC-Life, SSHOC, GO-FAIR Data Stewardship Competence Centres Implementation Network).
5. Managing FAIR support costs and resources – models for coordinating data stewards and research software engineers (EOSC-Life, GO-FAIR Data Stewardship Competence Centres Implementation Network).
6. Good practice for researchers, repositories and ethics committees on selecting and preparing sensitive data to be FAIR (SSHOC, FAIRplus, EOSC-life).

The D3.4 recommendations will be accompanied by an open call for implementation use cases. By highlighting examples of challenges being met through cross-disciplinary collaboration, FAIRsFAIR task T3.3 will support communities to adopt measures likely to

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

increase the production of FAIR data, or increase their preparedness to do so. Guidance and learning resources produced through WP3 and collated from the above sources will feed into the FAIRsFAIR competence centre.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

5. References

Becher, T., & Trowler, P. (2001). Academic tribes and territories. Buckingham: SRHE

Beagrie, N. (2019) What to Keep: A Jisc research data study. http://repository.jisc.ac.uk/id/eprint/7262

Berghmans, S., Cousijn, H., Deakin, G., Meijer, I., Mulligan, A., Plume, A., de Rijcke, S., Rushforth, A., Tatum, C., van Leeuwen, T., Waltman, L. 'Open Data: The Researcher Perspective' (2017). https://www.universiteitleiden.nl/en/research/research-output/social-and-behavioural-sciences/open-data-the-researcher-perspective

Bijsterbosch, M., Duca, D., Katerbow, M., Kupiainen, I., Dillo, I., Doorn, P. K., ... & de Lucas, J. E. M. (2016). Funding research data management and related infrastructures: knowledge exchange and science europe briefing paper.

CEC (2008) Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata (Text with EEA relevance) ELI: http://data.europa.eu/eli/reg/2008/1205/2008-12-24

Cooper, D., and Springer, R. (2019, May 13). Data Communities: A New Model for Supporting STEM Data Sharing. https://doi.org/10.18665/sr.311396

Cox, A. M., Kennan, M. A., Lyon, E. J., Pinfield, S., & Sbaffi, L. (2019). Progress in Research Data Services: An international survey of university libraries. *International Journal of Digital Curation*, *14*(1), 126-135.

Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *368*(1926), 4023-4038

DG Research and Innovation (May, 2019) Horizon Europe presentation v.25 https://ec.europa.eu/info/sites/info/files/research_and_innovation/ec_rtd_he-presentation_062019_en.pdf

Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L. (2011). Science friction: Data, metadata, and collaboration. *Social Studies of Science*, *41*(5), 667-690.

ENVRIplus (2018). D8.4 Interoperable cataloguing and metadata harmonisation for environmental RIs: prototype. Available at: http://www.envriplus.eu/wp-content/uploads/2015/08/D8.4-Interoperable-cataloguing-and-metadata-harmonisation-for-environmental-RIs-prototype-.pdf

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

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/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283

Fecher, B., Friesike, S., & Hebing, M. (2015). What Drives Academic Data Sharing? PLOS ONE, 10(2), e0118053. https://doi.org/10.1371/journal.pone.0118053

Fenner, M. (2019). Using Jupyter Notebooks with GraphQL and the PID Graph (Version 1.0). https://doi.org/10.5438/HWAW-XE52

Fenner, M., & Aryani, A. (2019). Introducing the PID Graph (Version 1.0). https://doi.org/10.5438/JWVF-8A66

Figshare (2019) State of open data report https://knowledge.figshare.com/articles/item/state-of-open-data-2019

Hrynaszkiewicz, I., Birukou, A., Astell, M., Swaminathan, S., Kenall, A., & Khodiyar, V. (2017). Standardising and harmonising research data policy in scholarly publishing. IJDC 12(1) 65-71. http://dx.doi.org/10.2218/ijdc.v12i1.531

Jones, S. Pergl, R. Hooft, R. Miksa, T., Samors, R. Ungvari, J. Davis, R.I and Lee.T. Data management planning: How requirements and solutions are beginning to converge. Data Intelligence 2(2020), 208–219. doi: 10.1162/dint_a_00043

Lahti, K., Schulman, L., Piirainen, E., Riihikoski, V. M., & Juslén, A. (2019). 'As Open as Possible, as Closed as Necessary'–Managing legal and owner-defined restrictions to openness of biodiversity data. *Biodiversity Information Science and Standards*.

Leonelli, S. and Ankeny, R. (2015) "Repertoires: How to Transform a Project into a Research Community," *BioScience* 65:7: 701-08, DOI: 10.1093/biosci/biv061

Linek, S. B., Fecher, B., Friesike, S., & Hebing, M. (2017). Data sharing as social dilemma: Influence of the researcher's personality. PLOS ONE, 12(8), e0183216. https://doi.org/10.1371/journal.pone.0183216

Lyon, L., Rusbridge, C., Neilson, C., & Whyte, A. (2011). Disciplinary Approaches to Sharing, Curation, Reuse and Preservation: DCC SCARP Final Report. Digital Curation Centre. http://www.dcc.ac.uk/sites/default/files/documents/scarp/SCARP-FinalReport-Final-SENT.pdf

McGeever, M., Whyte, A. & Molloy, L. (2015). 'Five Things You Need to Know About Research Data Management and the Law. DCC Checklist on Legal Aspects of RDM'. Digital Curation Centre. http://www.dcc.ac.uk/resources/how-guides

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Munafò, M. Promoting openness Research professional news 30 October 2019 https://www.researchprofessionalnews.com/rr-news-uk-views-of-the-uk-2019-10-promoting-openness/

Murphy, F, and Samors, M. Belmont Forum Data Accessibility Statement Policy and Template - Endorsed 18 October 2018 https://zenodo.org/record/1476871

Myers, N., Unsworth, K., Jaquemot, M-C, Murphy, F and Whyte, A. (2019) Exposing DMP Update, RDA Plenary 13, Philadelphia(presentation) March 2019 https://www.rd-alliance.org/p13-philadelphia

National Archives, UK (2014) 'Guidance on Cloud Storage and Digital Preservation', www.nationalarchives.gov.uk/documents/archives/cloud-storage-guidance.pdf

O'Carroll, C., Kamerlin, C.L, Brennan, N. Hyllseth, B., Kohl, U. O'Neil, G and v.d. Berg, R. (2017) 'Providing researchers with the skills and competencies they need to practise Open Science' Open Science Skills Working Group Report. European Commission Directorate-General for Research and Innovation http://ec.europa.eu/research/openscience/index.cfm

Pryor, G., Jones, S., & Whyte, A. (Eds.). (2013). Delivering research data management services: Fundamentals of good practice. Facet Publishing.

RDA (2016) RDA & CODATA Legal Interoperability Of Research Data: https://www.rd-alliance.org/rda-codata-legal-interoperability-research-data-principles-and-implementation-guidelines-now

Saenen, B. Morais,R. Gaillard, V. and Borrell-Damián, L. (2019) Research Assessment in the Transition to Open Science 2019 EUA Open Science and Access Survey Results https://eua.eu/resources/publications/888:research-assessment-in-the-transition-to-open-science.html

Science Europe (2018) Framework for Discipline Specific Research Data Management https://www.scienceeurope.org/our-resources/guidance-document-presenting-a-framework-for-discipline-specific-research-data-management/

Science Europe (2019) Practical Guide to the International Alignment of Research Data Management. https://www.scienceeurope.org/our-resources/practical-guide-to-the-international-alignment-of-research-data-management/

Simms, S., Jones, S., Mietchen, D., & Miksa, T. (2017). Machine-actionable data management plans (maDMPs). *Research Ideas and Outcomes*, *3*, e13086.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

SSHOC project (2019) Deliverable 3.1. Report on SSHOC (meta)data interoperability problems. https://sshopencloud.eu/d31-sshoc-report-sshoc-data-interoperability-problems

SPARC Europe (2017) Open data citation advantage briefing. https://sparceurope.org/what-we-do/open-data/sparc-europe-open-data-resources/

Tam, Wan Ting (2019) Discipline and research data in geography. figshare. PhD Thesis, University of Loughborough https://hdl.handle.net/2134/26997

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., ... Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. PloS One, 10(8), e0134826.

Tenopir, C., Talja, S., Horstmann, W., Late,E., Hughes, D., Pollock, D., Schmidt, B., Baird, L., Sandusky, R. & Allard, S. (2017). Research data services in European academic research Libraries. Liber Quarterly, 27 (1).

Vierkant, et al. (2014). Schema for the Description of Research Data Repositories. Version 2.2. doi: 10.2312/re3.006

VSNU (2019) Recognition and Rewards https://www.vsnu.nl/en_GB/news-items/nieuwsbericht/489-vsnu-nwo-nfu-and-zonmw-drive-change-in-recognition-and-reward-of-academics.html

Wenger, Etienne (1998). Communities of Practice: Learning, Meaning, and Identity . Cambridge: Cambridge University Press. ISBN 978-0-521-66363-2

Whyte, A. (2015). 'Where to keep research data: DCC checklist for evaluating data repositories' v.1.1 Edinburgh: Digital Curation Centre. www.dcc.ac.uk/resources/how-guides

Whyte, A. et al. (2019) EOSCpilot D7.5: Strategy for Sustainable Development of Skills and Capabilities.

Wittenburg, P., de Jong, F., van Uytvanck, D.; Cocco, M.: State of FAIRness in ESFRI projects. In: Data Intelligence 2 (2020), 230-237. https://doi.org/10.1162/dint_a_00045

Wouters, P. Ràfols, I. Oancea,A. Caroline, S. Kamerlin,K. Holbrook J.B and Jacob, M. (2019) Indicator Frameworks for Fostering Open Knowledge Practices in Science and Scholarship. European Commission DG Research.

Yakel, Elizabeth, Ixchel M. Faniel, and Zachary J. Maiorana. (2019) Virtuous and vicious circles in the data lifecycle. Information Research, 24(2), paper 821. http://InformationR.net/ir/24-2/paper821.html.

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Zervas, P., Manouselis, N., Karampiperis, P., Hologne, O., Janssen, S., & Keizer, J. (2018). Roadmap for a pan-European e-Infrastructure for Open Science in Agricultural and Food Sciences. Available at http://erosa.aginfra.eu/sites/default/files/Roadmap%20Paper.pdf

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Annex 1 - D3.1 policy sources

Funder policies reviewed as part of D3.1 policy characterisation:

- Austrian Science Fund (FWF) Open Access to Research Data accessed at https://www.fwf.ac.at/en/research-funding/open-access-policy/open-access-to-research-data/
- Netherlands Organisation for Scientific Research (NWO) Open (FAIR) Data accessed at https://www.nwo.nl/en/policies/open+science/data+management
- Research Foundation Flanders Data Management Plan (DMP) accessed at https://www.fwo.be/en/the-fwo/organisation/data-management-plan/
- European Commission DG RTD Unit Open Science Open Access and Data Management accessed at http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm
- Parkinson's UK Data sharing & preservation: Policy & guidelines accessed at https://www.parkinsons.org.uk/sites/default/files/2017-06/Data%20sharing%20policy%20and%20guidelines%20May%202017.pdf
- UK Research and Innovation Common principles on data policy accessed at https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/
- Swiss National Science Foundation Open Research Data accessed at http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx
- Kone Foundation Research funding; Ethical considerations, reuse of research materials and open access accessed at https://koneensaatio.fi/en/grants/forgrantrecipients/
- The Research Council of Norway Open access to research data accessed at https://www.forskningsradet.no/en/about-the-research-council/forskningspolitikk/open-science/open-access-to-research-data/
- Research Council Formas Good to know before you apply: Open access to research results and data accessed at https://formas.se/en/start-page/applying-for-funding/how-it-works/good-to-know-before-you-apply.html
- The French National Research Agency Open Science accessed at https://anr.fr/en/anrs-role-in-research/values-and-commitments/open-science/
- Fundação para a Ciência e a Tecnologia Data Availability Policy and other Results of FCT-funded R&D Projects accessed at https://www.fct.pt/documentos/PoliticaAcessoAberto_Dados.pdf

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- The Royal Society Conditions of Award accessed at https://royalsociety.org/-/media/grants/schemes/Conditions-of-Award.pdf?la=en-GB&hash=73BEAFC806D50AFC020B11165953A3DF
- Arcadia Fund Open Access and Digital Preservation Policy accessed at https://www.arcadiafund.org.uk/about-peter-baldwin-lisbet-rausing/open-access-and-digital-preservation-policy/
- Research Council of Lithuania Resolution on the approval of open scientific publications and data guidelines accessed at https://www.e-tar.lt/portal/en/legalAct/dceeeb10e05711e59cc8b27b54efaf6e
- Wellcome Trust Data, software and materials management and sharing policy accessed at https://wellcome.ac.uk/funding/guidance/policy-data-software-materials-management-and-sharing
- European Research Council (ERC) Open Access Guidelines accessed at https://erc.europa.eu/sites/default/files/document/file/ERC_Open_Access_Guidelines-revised_feb_2016.pdf

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558