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

D5.6 Report 3 of the Synchronisation Force

Work Package	WP5
Lead Author (Org)	Marjan Grootveld (DANS)
Contributing Author(s) (Org)	Simon Hodson (CODATA), Sara Pittonet Gaiarin (Trust-IT), Joy Davidson (DCC), Ingrid Dillo (DANS)
Due Date	31.08.2021
Date	31.08.2021
Version	0.1 <mark>DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION</mark>
DOI	https://doi.org/10.5281/zenodo.5336657

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	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 is the report of the third FAIRsFAIR Synchronisation Force workshop, organised online as a series of eight sessions from 29th of April until 10th of June 2021. The objective of these sessions was to survey the current state of activities towards implementing the recommendations outlined in the *Turning FAIR into Reality* report (2018), and to provide indicators of the amount of activity being undertaken. To do this, FAIRsFAIR brought together representatives of INFRAEOSC-5 projects, ESFRI clusters and and e-infrastructure projects, the EOSC Association Board of Directors, the FAIRsFAIR European Group of FAIR Champions and High-Level Advisory Committee, as well as project officers of the European Commission, to share information on their FAIR-oriented activities and to discuss commonalities and priorities. This report summarises the workshop outcomes.

Versioning and contribution history

Version	Date	Authors	Notes
0.1	07.06.2021	Marjan Grootveld, Simon Hodson, Sara Pittonet Gaiarin, Joy Davidson	First draft
0.2	14.07.2021	Marjan Grootveld, Simon Hodson, Sara Pittonet Gaiarin	Draft for internal review
0.3	31.07.2021	Hilary Hanahoe (RDA Foundation), Kevin Ashley (Digital Curation Centre)	Internal review
0.4	27.08.2021	Marjan Grootveld, Simon Hodson, Sara Pittonet Gaiarin, Ingrid Dillo	Full draft version with review comments addressed
1.0	30.08.2021	Sara Pittonet Gaiarin	Final version ready to upload

Disclaimer

Abbreviations and Acronyms

API	Application Programming Interface
CMIP6	Coupled Model Intercomparison Project Phase 6
COPDESS	Coalition for Publishing Data in the Earth and Space Sciences
CTS	CoreTrustSeal
DCAT	Data Catalog Vocabulary
DMP	Data Management Plan
EC	European Commission
EGFC	European Group of FAIR Champions
EOSC	European Open Science Cloud
ESGF	Earth System Grid Federation
ESFRI	European Strategy Forum on Research Infrastructures
EUDAT CDI	EUDAT Collaborative Data Infrastructure
FAIR	Findable, Accessible, Interoperable, Reusable
FIP	FAIR Implementation Profile
F-UJI	FAIR Research Data Object Assessment Tool
HLAC	FAIRsFAIR's High-Level Advisory Committee
M4M	Metadata for Machines
maDMP	machine-actionable Data Management Plan
NOAD	National Open Access Desk
RDA	Research Data Alliance
RDM	Research Data Management
RI	Research Infrastructure
RPO	Research-Performing Organisation
RSE	Research Software Engineers
SF	Synchronisation Force (part of FAIRsFAIR)
SSH	Social Sciences and Humanities
TFiR	Turning FAIR into Reality (report)
WG	Working Group
WS1	First workshop of the Synchronisation Force (2019)
WS2	Second workshop of the Synchronisation Force (2020)
WS3	Third workshop of the Synchronisation Force (2021)

Project names, which are often acronyms, can be found in Annex 1.

Executive Summary

This is the report of the third FAIRsFAIR Synchronisation Force workshop (WS3), organised online as a series of six sessions from 29th of April until 21st of May 2021. The objective of these sessions was to survey the current state of activity towards implementing the recommendations outlined in the *Turning FAIR into Reality* report (2018), and to provide indicators of the amount of activity being undertaken. To do this, FAIRsFAIR brought together representatives of INFRAEOSC-5 projects, ESFRI clusters and e-infrastructure projects, the EOSC Association Board of Directors, the FAIRsFAIR European Group of FAIR Champions and High-Level Advisory Committee, as well as project officers of the European Commission, to share information on their FAIR-oriented activities and to discuss commonalities and priorities. A similar exercise was carried out in the second FAIRsFAIR Synchronisation Force workshop in 2020 (WS2).

The two main target audiences for this report are the EOSC Association as well as the European Commission's EOSC programme (together called "the EOSC Governance" in this report) on the one hand, and the EOSC and FAIR related projects and initiatives on the other.

Value for the EOSC and FAIR related projects and initiatives

The report provides an overview of activities being undertaken by projects involved in building a FAIR ecosystem for the European Open Science Cloud. Structured according to the 'pillars' and recommendations presented in the TFiR report, we hope that this will be useful in communicating the current landscape and that stakeholders, actors and the projects themselves will find it valuable as well for the purpose of identifying synergies and opportunities for collaboration.

Value for the EOSC Governance

The report also offers a representation of the amount of activity being undertaken towards each recommendation across the projects and initiatives that participated. These indicators *do not* pretend to show the amount of progress towards achieving each recommendation. The indicators are based on the collective judgement emerging from the workshops: they were discussed and ultimately approved with participants. As such it helps us understand where there are good levels of activity and where more attention might need to be directed.

These findings are summarised in the following infographic.
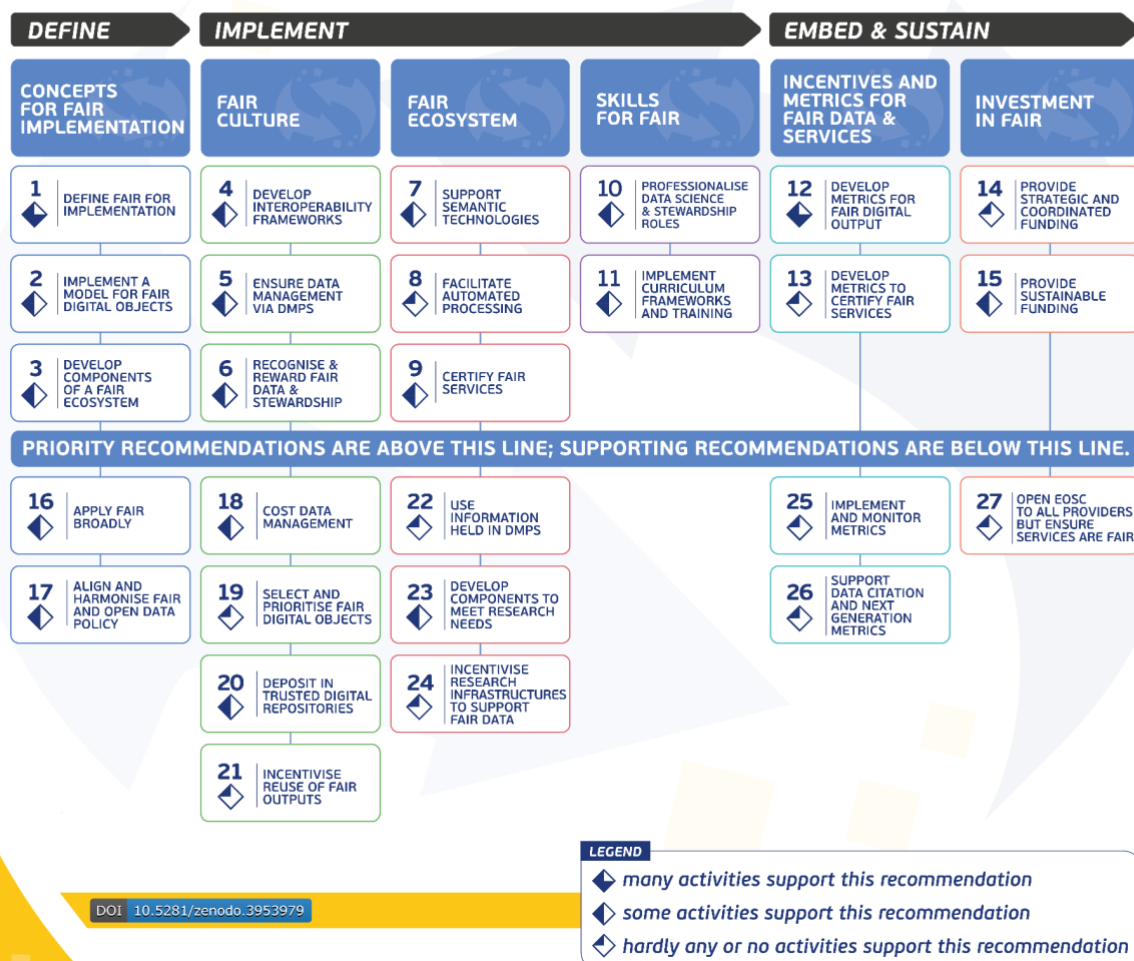
Through the workshop sessions, discussions and the information collected and presented here, we show that **there is broadly a good level of activity towards the TFiR recommendations**, across the participating projects and initiatives. The report also highlights a number of areas where more activity is needed and makes recommendations for additional effort and focus.

Table of contents

Index of figures

1. Introduction to the FAIRsFAIR Synchronisation Force workshops

The FAIRsFAIR project aims to supply practical solutions for the use of the FAIR data principles throughout the research data lifecycle with an emphasis on fostering a FAIR data culture and the uptake of good practices in making data FAIR, in particular in the context of the European Open Science Cloud (EOSC). A key challenge for FAIRsFAIR is to ensure that project activities dovetail with work carried out by the former Working Groups of the EOSC Executive Board and the EOSC Association task forces (as the landscape evolves in the course of 2021), and hence feed into and complement the work done by other projects in the research data and FAIR space. To address this challenge the FAIRsFAIR project set up the **Synchronisation Force**, a team tasked with establishing a dialogue among the various projects and actors in both the EOSC and FAIR ecosystems whose work touches on FAIR in order to:

- Maximise coordination and minimise unnecessary overlap or duplication;
- Encourage the dovetailing of projects' and actors' activities with those of the EOSC Steering Board and the EOSC Advisory Group and Task Forces;
- Promote mechanisms to support collaboration on turning FAIR into reality.

The key activity of the Synchronisation Force (SF) was to run three dedicated workshops as the best platform to create a dialogue between the key stakeholders working in the dynamic landscape of FAIR activities in relation to EOSC.

FAIRsFAIR's primary stakeholders within the EOSC ecosystem are illustrated in Figure 1. They are the INFRAEOSC-5a-c projects, which collaborate in Task Forces under a Cross-Project Collaboration Board, domain-related ESFRI cluster projects, and generic or "horizontal" e-Infrastructure projects. Up to early 2021 the Working Groups of the EOSC Executive Board played an important role as well. In addition, the FAIRsFAIR High-Level Advisory Committee (HLAC[1]) and the European Group of FAIR Champions (EGFC)[2], initiated by the project to work as FAIR ambassadors, were invited to join the workshop. Finally, other FAIR-related initiatives progressively joined the activity, such as representatives from the ARCHIVER[3] and OCRE[4] procurement projects and the FAIRSharing[5] initiative.

---

[1] FAIRsFAIR High-Level Advisory Committee. https://www.fairsfair.eu/advisory-board/hlac
[2] European Group of FAIR Champions. https://www.fairsfair.eu/advisory-board/egfc
[3] https://www.archiver-project.eu/
[4] https://www.ocre-project.eu/
[5] https://fairsharing.org/

Figure 1 FAIRsFAIR primary stakeholders within the EOSC ecosystem

The workshop series

Although the workshops were planned to be held as face-to-face meetings, the second and third workshops were restructured as virtual workshops due to the COVID-19 restrictions at the time. Each of the three workshops resulted in a report.

9

The **first Synchronisation Force workshop or WS1**, held in Budapest, 25 November 2019, explored interaction between FAIRsFAIR and the five EOSC Working Groups[6] that were established at that time[7]. The first workshop highlighted the difficulty around coordination and collaboration of activities as a key challenge. Furthermore, it was concluded that the importance of clustering activities and outputs around recommendations from the *Turning FAIR into Reality* report was underemphasized in the workshop and that more stakeholders should have a seat at the table.

These three issues were addressed in the **second Synchronisation Force workshop or WS2[8]** in 2020. First, the *Turning FAIR into Reality* recommendations were the linch pin of the sessions. Second, the workshop itself was instrumental in bringing stakeholders together to share information about their respective activities. To better enable discussion around TFiR themes, eight sessions were held instead of one workshop where discussions would have to remain at a higher level. Third, for WS2 we invited a wider group of stakeholders. During WS2 diamonds with coloured facets were introduced to indicate the amount of activity towards implementing each recommendation. This led to an updated overview of the original TFiR overview.

many activities (in place and/or planned) support this recommendation

some activities (in place and/or planned) support this recommendation

hardly or no activities (in place and/or planned) support this recommendation

Figure 2 Legend with the coloured facets diamond indicators for the amount of activity concerning the implementation of a TFiR recommendation

Finally, for the **third Synchronisation Force workshop or WS3**, in 2021, we repeated this approach, with invitations extended to few new project representatives as well as members of the EOSC Board of Directors[9]. **Six sessions** in May centered around the six TFiR pillar recommendations and how they are contributing to the EOSC and FAIR vision in Europe as well as the current state of implementation activities. WS3 was concluded with a discussion session on June 10th.

> **Important note for the reader**
>
> This report provides an overview of activities being undertaken by projects involved in building a FAIR ecosystem for the European Open Science Cloud and represented in a FAIRsFAIR workshop. The overview therefore is not exhaustive. The content is structured according to the 'pillars' and

---

[6] https://www.eoscsecretariat.eu/eosc-working-groups

[7] Report on the First workshop: Gerard Coen, Mustapha Mokrane, Sara Pittonet, Simon Hodson, & Renee van Kessel-Hagesteijn. (2020). D5.3 Report on the First Synchronisation Force Workshop (Version 1.0). https://doi.org/10.5281/zenodo.3629159.

[8] Report of the Second workshop: Ingrid Dillo, Marjan Grootveld, Simon Hodson, & Sara Pittonet Gaiarin. (2020). Second Report of the FAIRsFAIR Synchronisation Force (D5.5) (Version 1.0). https://doi.org/10.5281/zenodo.3953978

[9] See Annex 2 List of participants

recommendations presented in the Turning FAIR into Reality report, and we hope therefore communicates the current landscape in a way that various stakeholders, actors and the projects themselves will find useful, including for the purpose of identifying synergies and opportunities for collaboration.

The report also offers a representation of the *amount of activity* being undertaken towards each recommendation across the projects and initiatives that participated. The level of activity is represented by diamonds (many activities, some activities, hardly any activities). These indicators are based on the collective judgement emerging from the workshops: they were discussed and ultimately approved with participants.

The utility of this is to provide an indication of where effort is currently being focussed and to identify areas where more effort may be needed.

What this report (and the indicators used) categorically does not pretend to show is the progress towards achieving each recommendation: i.e. How FAIR are we? or, how far are we? That would require a different level of resource and a detailed methodology beyond the scope of this project and activity.

The stakeholders represented in the third Synchronisation Force workshop and the workshop participants are listed in Annex 1 and 2, respectively. For ease of reading of this report we will use the term "projects" to refer to all types of stakeholders.

2. Turning FAIR into Reality: tracking activities towards FAIR

The European Commission Expert Group Report *Turning FAIR into Reality* (TFiR[10]) lays out an Action Plan for what is needed to implement FAIR. It recognises that in order for data and other research outputs to be Findable, Accessible, Interoperable and Reusable, a broader ecosystem of shared concepts, technologies, services, skills and culture is required. Furthermore, that ecosystem needs to be sustained by appropriate investment and sound governance.

To that end the TFiR report makes a series of structured recommendations, and proposes an Action Plan for delivering a FAIR data ecosystem. In turn, the structure in which those recommendations are presented provides a useful and appropriate framework for discussing and comparing FAIR-oriented activities of a wide range of projects and initiatives across Europe. The recommendations are clustered under six pillars:

1. Core Concepts: FAIR Digital Objects and technical components of the FAIR ecosystem
2. FAIR Culture: agreements on data availability and description, data management plans, recognition and rewards
3. FAIR Ecosystem: key services, semantic technologies, trust and certification of services
4. Skills for FAIR: data science and data stewardship, curriculum frameworks

---

5. Incentives and Metrics for FAIR data and services: citation, metrics and monitoring
6. Investment in FAIR: investment, sustainability and governance
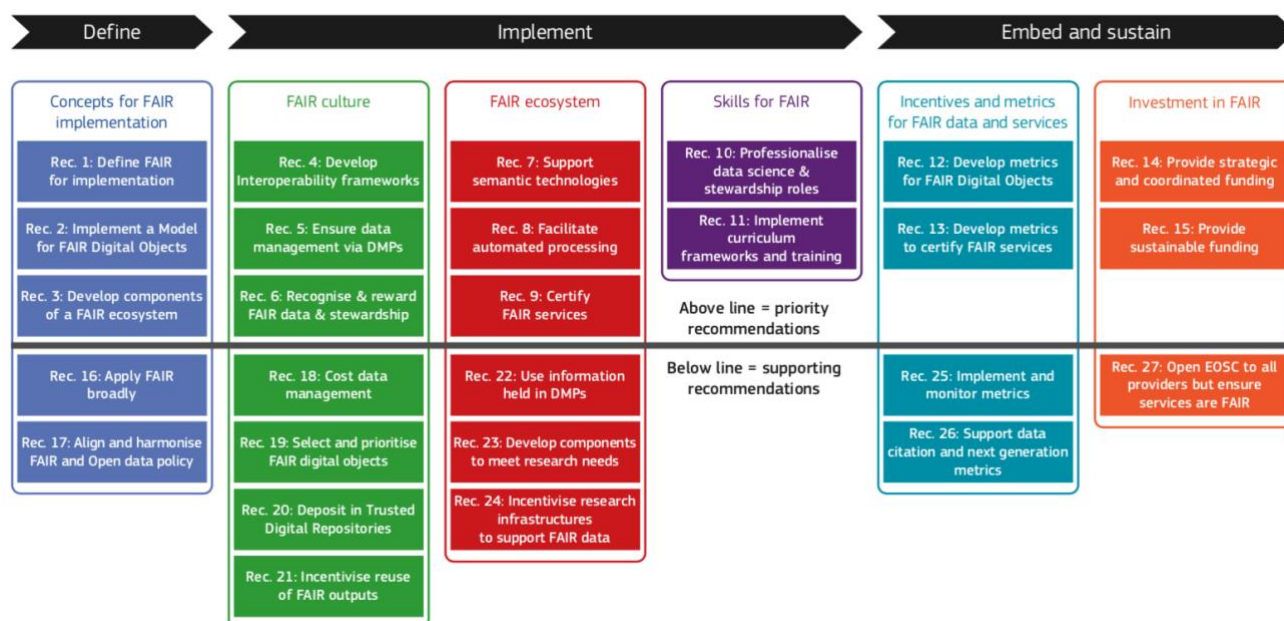
Figure 3 Turning FAIR into Reality priority and supporting recommendations (Source: TFiR, page 17)

The Synchronisation Force workshops aimed to understand how actors within the EOSC and FAIR ecosystems are addressing the TFiR recommendations and how they are responding to the *Turning FAIR into Reality* Action Plan.

In WS3 information from the projects was collected and discussed for each of the six pillars (see Figure 3), hence the term "pillar session". In advance of the six pillar sessions the Synchronisation Force published a spreadsheet listing all TFiR recommendations. Where possible, we had populated the spreadsheet with information from WS2 (2020). For each recommendation—if relevant—the participants provided or updated information about their project's deliverables and ongoing activities that aim to address this recommendation: both "in place" and "planned" activities. During the sessions, participants were invited to flesh out this information. Furthermore, each pillar session discussed whether any significant FAIR-related activities were missing from the TFiR recommendations or deserved more attention or priority.

3. Findings and recommendations

This report provides an overview of activities being undertaken by projects involved in building a FAIR ecosystem for the European Open Science Cloud. This section presents the analysis and the recommendations that FAIRsFAIR bases on the detailed workshop information in Section 4. Structured according to the 'pillars' and the recommendations presented in the *Turning FAIR into Reality* report, FAIRsFAIR hopes that this will be useful in communicating the current landscape and that stakeholders, actors and the projects themselves will find it valuable as well for the purpose of identifying synergies and opportunities for collaboration.

The report also offers a representation of the amount of activity being undertaken towards each recommendation across the projects and initiatives that participated. With the proviso that this is a snapshot in time which is based on input from many but obviously not all EOSC-related initiatives, the following image summarises the amount of activity with regards to the 27 TFiR recommendations. In the final workshop session, the participants were presented with a draft coding made by FAIRsFAIR. They mostly agreed that the draft coding of each pillar was an accurate representation of the activities reported, though verging on being too positive. As a result, the current coding in this overview is a little stricter than the draft coding. **Moreover, the coding is lower than in WS2: the workshop organisers and participants agreed that, one year on, we should raise the bar.**
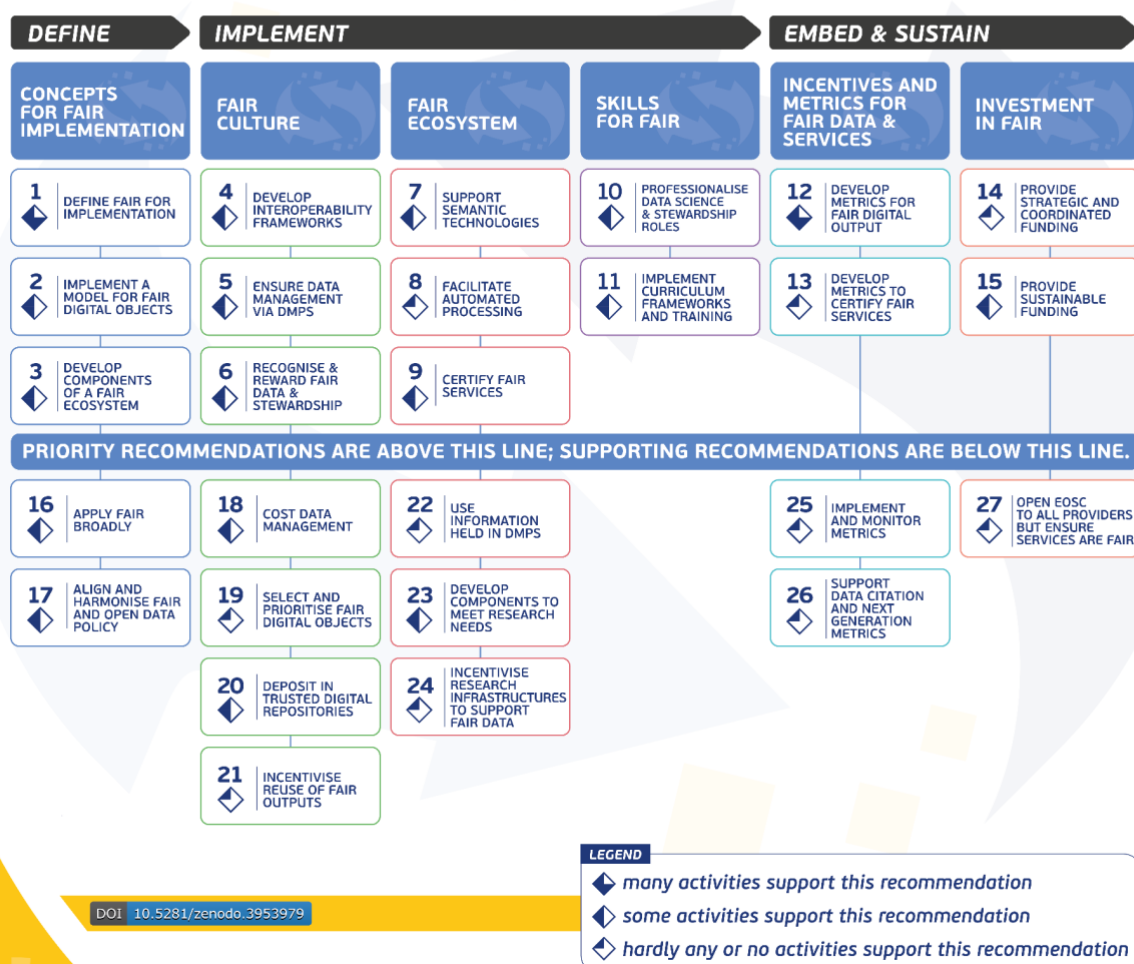
Figure 4 The amount of activity per TFiR recommendation, identified in the workshop

As emphasised clearly above, these indicators *do not* pretend to show the amount of progress towards achieving each recommendation. That would require a different level of resource and a detailed assessment methodology beyond the scope of this project and activity.

The indicators are based on the collective judgement emerging from the workshops: they were discussed and ultimately approved with participants. This methodology is important and useful. As discussed above, it has allowed for a recalibration of the assessment of the amount of activity between the two reports. In the collective view of workshop participants, the level of activity against

some TFiR recommendations was judged to be lower in WS3 (2021) than in WS2 (2020). This communicates, we believe, a rising level of expectations and a realisation, one year further on, after the first blush of excitement for the considerable work being undertaken had faded, that more activity is needed and that there remain significant gaps, and recommendations which require more attention.

The information in Section 4 indicates that **there is broadly a good level of activity towards the TFiR recommendations**, across the participating projects and initiatives. This is particularly strongly the case in the area of defining and communicating the definition of FAIR (Rec.1) and preparing metrics for FAIR Digital Objects and Outputs (Rec.12). Other areas are also receiving a lot of attention and commendable activities: certification for trusted data repositories, adoption of Data Management Plans, interoperability frameworks and semantics and training activities, to list a few examples.

Recommendations needing more attention or an added focus. It emerged from this exercise and workshop that there remain some significant gaps and areas where only part of a given recommendation is being addressed. These areas are listed in the report below with the bulleted subheadings **Missing Element**.

A lot of work is going on around Rec. 4 'Develop Interoperability Frameworks', one of the cornerstone recommendations. Understandably enough, at this stage a lot of effort is directed towards the harmonisation and interoperability of catalogue level metadata, with *inter alia* DCAT playing a central role. This is necessary and significant work. Nevertheless, **it remains important to further extend this activity with research communities and Research Infrastructures on vocabularies, ontologies, and detailed metadata for existing domains and for emerging cross-domain research areas.**

Similarly in relation to Rec.7 'Support Semantic Technologies', it was observed in the workshop that while there is good level of activity among the projects in implementing semantic technologies in some domains, there remains a considerable need to provide support (including training) for domains (and institutions and countries) that have made less progress, so that the current gap in experience and implementation does not increase further. As for Rec.4, a lot of focus remains on the level of catalogue integration. **Stepwise, as progress is made, activity will need to extend to more detailed metadata and semantics, to less advanced domains, and to emerging cross-domain research areas.**

Data Management Plans are essential for good data stewardship and are a necessary step towards ensuring project outputs are well curated and FAIR. As such they are increasingly mandated by research funder policy. Their implementation is touched upon in a number of the recommendations (most notably Rec.5) and the participating projects and initiatives referenced work relating to DMPs for a number of other recommendations. Solid and commendable activity is underway to support the use of DMPs. A number of areas emerged which need further attention. **The value of the Science Europe domain protocol approach to DMP should be explored further. More activity is also needed to refine and implement machine-actionable DMPs**. Understandably, therefore, there is as yet very little activity on actively using the information held in DMPs to better inform various components of the FAIR ecosystem (Rec.22).

A lot of work is being done to assist repositories in the task of attaining the CoreTrustSeal certification (Rec.20, Rec.9). However, **meeting the CoreTrustSeal requirements remains a challenging task for smaller repositories and for institutional repositories.** SSHOC has been engaging with smaller repositories. There is a need for greater clarity on whether and how smaller repositories should engage with certification processes. There is likely also a requirement for a further onramp process, greater support or other steps to support smaller and institutional repositories.

Finally, among the areas that deserve attention in this short discussion, it is evident that a number of projects and initiatives have training activities and that this is extremely valuable work towards Recs. 10 and 11 (Professionalise Data Science and Data Stewardship Roles; Implement Curriculum Frameworks and Training). Nevertheless, we should be reminded (as indeed it emerged in the workshop discussions) that **in addition to providing training, this recommendation also comprises the development of the curriculum framework, of certification and establishing professional bodies, activities to encourage institutions to recruit staff to these posts, to develop a cohort of data stewards and data scientists to support projects.** Further activity is needed in these areas; activity which in some respects will need the participation of funders, policy makers and other stakeholders that set the framework for professionalisation and the recognition of curricula.

Recommendations needing more activity: WS3 identified a number of areas of lower activity. In relation to different recommendations, different factors are at play. **In a number of cases, particularly for the TFiR 'Supporting Recommendations', dependencies or necessary prioritisation means they have received less attention.** For example, this is understandably the case in relation to Recs. 22 'Use Information in DMPs' (already discussed); and 19 'Select and Prioritise FAIR Digital Objects'.

In relation to other recommendations there are **dependencies and the need for involvement of other stakeholders, particularly funders and policy makers**. This is most obviously the case in those Recommendations relating to funding, sustainability and governance (Recs. 14, 15, 27), but also to those concerning metrics (Recs. 21, 13, 25, 26). **It would be very constructive, therefore, to organise an activity (workshops) involving funders and policy makers specifically to survey activity in the areas of funding, sustainability, governance and metrics.**

The most notable area with a lower level of activity is Priority Rec. 8 'Facilitate Automated Processing'. It should be emphasised that one of the most important objectives of the FAIR principles is precisely to better facilitate the machine-processing, at scale, of data and other research outputs: this priority objective is conveyed in the TFiR report. At the same time, we need to recognise that this is a particularly challenging recommendation. A lot of useful work is proceeding to help automate tools used in support of research and to assist RIs, RPOs and repositories (automation of FAIR assessment tools, automation of aspects of DMPs and the use of DMP information). However, the main objective of this recommendation (implementation of FAIR and semantic technologies to enable the automation of research processes) seems to have received less attention beyond certain domains where there are already some automated data processing pipelines and workflows. Although doubtless challenging, further attention to this recommendation, a cornerstone of the FAIR principles, is needed.

Finally, Rec. 18 'Cost Data Management' is an area that evidently needs further attention. Again this is a challenging area, but as better processes and support are put in place and as the data management and repository components of the FAIR ecosystem are put in place, it becomes a priority to provide cost information as evidence for funders and other stakeholders, to help substantiate and quantify the necessary investment.

4. Mapping project activities to the individual TFiR recommendations

4.1 Scoring the activity levels

In section 4 we describe pillar by pillar the main findings for each TFiR recommendation. The project activities reported below are a selection of the information provided in the spreadsheet and during the sessions. As noted above, the overview of activities presents a snapshot in time and is not exhaustive. In addition to the information presented in the spreadsheet, which relates to specific project deliverables for each TFiR recommendation, the extensive session notes are available for more details[11]. Where relevant, possible follow-up actions and/or missing elements are added. Furthermore, the following coding is used:

many activities (in place and/or planned) support this recommendation

some activities (in place and/or planned) support this recommendation

hardly or no activities (in place and/or planned) support this recommendation

In WS2 and WS3 FAIRsFAIR took the same qualitative approach to scoring the activity levels for each TFiR recommendation. The crucial consideration is how broadly the reported activity addresses the recommendation including the actions. When the activities only support a part of the recommendation and actions, the score can't exceed "some activities".

Firstly, the workshop organisers individually scored the activity level for each recommendation and they discussed these scores. Secondly, the scores that resulted from this discussion were shared with the workshop participants in the draft report. Thirdly, based on discussion with the participants, some scores were adapted. Stricter, less subjective approaches were beyond the scope of the SF workshop series, but could be very interesting in future. Please see Section 4.8 for more considerations about a scoring approach.

---

[11] The spreadsheet and the session notes are available from https://doi.org/10.5281/zenodo.5336657

4.2 Pillar 1: Concepts for FAIR implementation

Turning FAIR into Reality Pillar 1: Concepts for FAIR implementation defines how the FAIR principles apply in the context of research communities - based on the data types, the nature of research and the level of existing support for data sharing. It also identifies points where the FAIR principles need to be supported with additional concepts and policies to be further expanded and unpacked.

The recommendations address the concept of "Defining" with a two-fold focus:
- **FAIR Digital Objects** - Defining what is needed for digital objects to be made FAIR.
- **FAIR Ecosystem** - Defining which components are needed in the FAIR ecosystem.

Rec. 1: Define FAIR for implementation

"To make FAIR data a reality it is necessary to incorporate and emphasise concepts that are implicit in the FAIR principles, namely: data selection, long-term stewardship, assessability, legal interoperability and the timeliness of sharing". (TFiR, page 62)

The outcomes of the second FAIRsFAIR Synchronisation Force workshop (WS2 in 2020) had already pointed out that "the definition of how the FAIR principles apply in the context of research communities is widely tackled by all of the EOSC initiatives and collaboration and cross-fertilisation is already in place, with many actors sharing analyses, reports and best practices and contributing to each other's working groups"[12]. During WS3 this was confirmed with precise actions reported by the EOSC cluster projects as well as "horizontal" activities. The feeling is that during this second period (Summer 2020 - Spring 2021) actions moved from a scoping activity, with meetings and events planned to promote and discuss tools and best practices, more to an "adoption" approach with practical implementation stories undertaken.

Looking at the ESFRI clusters, the science cluster **ESCAPE** is mapping community standards in astronomy to the FAIR data principles to ensure these are being implemented complementary to FAIR data standards[13]. **ExPaNDS worked with PaNOSC** to finalise its model for data policy and realised the needs of ExPaNDS facilities for a higher-level data policy framework as well. The final version of the ExPaNDS data policy framework will be presented in August 2021 and workshops on FAIR for facility staff took place last year. ENVRI-FAIR also reported on practical work done to map out what is necessary to increase FAIRness of data in **ENVRI-FAIR subdomains**, introducing awareness where perhaps there was more subconscious activity. The outcomes will convene together in the ENVRI-hub and in the ENVRI Knowledge Base, with a third run of FAIRness assessment scheduled in 2022, following a bottom-up, coordinated approach.

With respect to FAIR assessment tools, the F-UJI tool[14] is being exploited, for instance by **EOSC-Nordic** that used it to map regional repositories, making recommendations on how repositories can improve their scores, and identifying repositories where there has been a significant increase thanks to this support (and the continuous work to improve the F-UJI software). The same was done by the **ARCHIVER** project that is exploring the F-UJI tool to evaluate the Buyers' repositories and guarantee FAIR principles compliance in the procured cloud services. Another tool is **OSSDIP**, the Open source Secure Data Infrastructure and services, released in v1[15] by the Information and Software Engineering Group (IFS) Of TU Wien, offering a reference implementation of a completely isolated data visiting environment (applying the RDA Dynamic Data Citation Working Group (WG-DC) recommendations[16]).

Other initiatives such as **NI4OS** focus more on training and on legal interoperability aspects, such as the level of licenses for derivative works.

Rec. 2: Implement a model for FAIR Digital Objects

"Implementing FAIR requires a model for FAIR Digital Objects. These, by definition, have a PID linked to different types of essential metadata including provenance and licencing. The use of community standards and sharing of rich documentation is fundamental for interoperability and reuse of all objects."

As already confirmed during the second Synchronisation Force series, many of the initiatives represented directly or indirectly at the workshop, are addressing this recommendation. The need to define digital objects is recognised, but work remains at an early stage. For example, in the **ENVRI-FAIR** project, different teams are involved in various RDA groups / EOSC task forces looking to map current practices for PIDs and scoping what is required to support FAIR Digital Objects. Starting in 2021, ENVRI-FAIR is also represented in the **FAIR Digital Object Forum initiative**[17]. In addition, individuals working for ENVRI-FAIR member organisations have been engaging with relevant **Research Data Alliance groups,** including the Data Fabric Interest Group (DFIG)[18]. Since 2020, the **DFIG** is concentrating its work on implementing technologies and strategies towards creating systems that allow machine-actionable management and processing of FAIR Digital Objects. The efforts are, for instance, devoted to the creation of a draft metadata framework for FAIR data management in Photon and Neutron facilities (**ExPaNDS**). The **NI4OS-Europe** project is working with infrastructure

---

[14] F-UJI FAIR data assessment tool is based on a web service to programatically assess FAIRness of research data objects based on metrics developed by Anusuriya Devaraju & Robert Huber (PANGAEA) under the umbrella of the FAIRsFAIR project https://fairsfair.eu/f-uji-automated-fair-data-assessment-tool. Anusuriya Devaraju & Robert Huber (2020) F-UJI - An Automated FAIR Data Assessment Tool (Version v1.0.0). http://doi.org/10.5281/zenodo.4063720

[15] Open Source Secure Data Infrastructure and Processes (OSSDIP)
http://www.ifs.tuwien.ac.at/~andi/secure_data_infrastructure.htm

[16] https://www.rd-alliance.org/groups/data-citation-wg.html

[17] FAIR Digital Objects Forum Initiative https://fairdo.org/

[18] RDA Interest Group. https://www.rd-alliance.org/group/data-fabric-ig.html

and service providers to deliver training on the PID systems available (PURL, Handle, DOI etc). Via **EOSC-Life** and under ISO ISO/TC 276 a provenance model is being developed (ISO 23494 specification).

Rec. 3: Develop components of a FAIR ecosystem

"The realisation of FAIR data relies on, at minimum, the following essential components: policies, Data Management Plans, identifiers, standards and repositories. There need to be registries cataloguing each component of the ecosystem, and automated workflows between them." (TFiR, page 63)

The updated EOSC Interoperability Framework 2.0[19] provides structure for this recommendation. It proposed a reference architecture that is inspired by and extends the European Interoperability Reference Architecture (EIRA) identifying which main building blocks are required. The former Interoperability Task Force of the EOSC FAIR Working Group[20] delivered an analysis of existing metadata models and an initial set of crosswalks among them. This work **may set the initial steps for a future proposal for an EOSC Minimal Metadata Application profile,** which should be widely discussed and agreed upon by disciplinary communities.

Examples of current practices. An example from long-tail ecological research: The **SPI-Birds Network** and Database[21] aims to host raw and standardized data on birds all over the world. Currently it hosts data on over 100 populations. Data are ingested in different formats (as stored by different groups) and code pipelines are written to transfer these data into a standard data format that is FAIR, and then conduct quality checks on data. The current need is to further developments of the database to connect it to other initiatives hosting complementary data, and also to find a viable workflow to connect all of the datasets that are based on the raw data hosted in the SPI-Birds Network database. Another example of a current practice is the **WDCC**[22] (World Data Center for Climate), hosted at DKRZ[23] (German Climate Computing Center, Hamburg, Germany). The WDCC is a CoreTrustSeal certified long-term archive that collects, stores, and disseminates Earth System data with a focus on climate simulation data and climate related data products. The WDCC thus provides central support for the climate modelling research community and ensuring long-term reusability of preserved data is the main goal. The data preservation workflow comprises quality checks and provision of DOIs to make the data reusable and citable. In line with DKRZ's adoption of the COPDESS[24] (Coalition on

---

[19] Oscar Corcho, Magnus Eriksson, Krzysztof Kurowski, ... Frederik Coppens, (2021). EOSC Interoperability Framework. https://www.eoscsecretariat.eu/news-opinion/achieving-interoperability-eosc-interoperability-framework

[20] https://www.eoscsecretariat.eu/working-groups/fair-working-group. The EOSC Working Groups and their Task Forces ended their mandate at the end of 2020 and produced 20 reports with recommendations in 6 priority areas for EOSC. These are key to the EOSC Strategic Research and Innovation Agenda (SRIA) which provides a clear roadmap for the next seven years to achieve the EOSC vision and objectives.

[21] SPI-Birds www.spibirds.org

[22] https://cera-www.dkrz.de/WDCC/ui/cerasearch/

[23] https://www.dkrz.de/en/dkrz-partner-for-climate-research?set_language=en

[24] https://copdess.org/

Publishing Data in the Earth and Space Sciences) Commitment on Data Publications in June 2018[25], the WDCC complies with the FAIR data guiding principles[26]. The WDCC archiving service is available to the entire climate and Earth System research community. The vast majority of WDCC-archived data is associated with an open license (CC-BY 4.0 or similar) and data can be accessed free of charge while authentication via registration is needed, free of charge and open to anybody.

Tools and registries. New work seems to be advancing. In the framework of the **NI4OS** project the RePoI tool[27] (currently running on beta version) attempts to create machine-actionable repositories' policies, while the **OpenAIRE ARGOS** machine-actionable tool[28] is running in a production version. ARGOS DMP outputs enhance the OpenAIRE Research Graph and create links with other outputs and relationships (semantics). **FAIRsharing**[29] has emerged as a registry of policies and standards (linked to associated databases) in many EOSC reports and for all disciplines; FAIRassist is being developed to offer personalised guidance to discover (and understand the interlinks between) FAIR resources, such as data and metadata standards and databases.

The **EOSC-Life** Collection in FAIRsharing represents more than 100 diverse data resources/repositories produced by EOSCLife partners: these FAIRsharing data resources Collection, and others relevant to COVID-19 data, are being designed as entry points for EOSC aggregators, such as OpenAIRE[30], to harvest metadata profiles (via schema.org, Bioschemas or OAI-PMH). WorkflowHub is emerging as the registry of workflows under ELIXIR and EOSC-Life.

ESFRI SSHOC also developed a series of facilities**,** namely the FAIR SSH Citation Service Prototype, the SSH Conversion Hub, the SSHOC Virtual Collection Registry, the SSHOC Switchboard.

Another example is the global distribution of CMIP6[31] data (Coupled Model Intercomparison Project Phase 6). As central part of the global Earth System Grid Federation (ESGF)[32] infrastructure in Europe, DKRZ offers services for making CMIP6 climatology framework data FAIR by providing a citation service allowing for the referencing of high-quality climate model output in the context of the IPCCs 6th Assessment report[33]. CMIP6 data are free and open for use by anybody. The imposed standards required for CMIP6 data yield a very high degree of domain-specific reusability and PIDs are assigned.

For **Data Management Plans**, which are also relevant here, please see Recommendation 5**.**

---

[25] https://www.dkrz.de/communication/news-archive/copdess

[26] See https://cera-www.dkrz.de/WDCC/ui/cerasearch/info?site=fairness and
https://meetingorganizer.copernicus.org/EGU21/EGU21-12560.html

[27] NI4OS RePoI tool. https://repol.ni4os.eu/

[28] Argos DMP tool. https://argos.openaire.eu/splash/

[29] FAIRsharing. https://fairsharing.org

[30] FAIRsharing and OpenAIRE collaboration: https://www.openaire.eu/enriching-science-fairsharing-and-openaire-amke-sign-a-memorandum-of-understanding

[31] CMIP. https://pcmdi.llnl.gov/CMIP6/

[32] http://esgf.llnl.gov/

[33] CMIP6. https://cera-www.dkrz.de/docs/pdf/CMIP6_Citation_Userguide.pdf

Possible follow-up action:

- Although there are quite a lot of activities, more effort could be put towards implementing components, developing registries and implementing automated workflows as key components of the FAIR ecosystem.

Rec. 16: Apply FAIR broadly

FAIR should be applied broadly to all objects (including metadata, identifiers, software and DMPs) that are essential to the practice of research, and should inform metrics relating directly to these objects.

A national dimension is being introduced. FAIRsFAIR started a series of National Roadshows engaging with country specific facilitators and multipliers, and the **NI4OS-Europe** project is trying to foster policy adoption, by setting up National Open Science Cloud initiatives (**NOSCIs**) in the 15 countries of its partners, envisaged to play a prominent role in Member States and Associated Countries and facilitate EOSC governance.

Target Cooperation across initiatives is continuing. ExPaNDS is organising an event for facility librarians and data managers, aimed at linking these two (often separate) areas in the project's facilities.

EOSC-Synergy and **FAIRsFAIR** collaborate on overlap between quality assurance of software and services and what that means for FAIR.

For examples of DMP activities, see recommendation 5.

Rec. 17: Align and harmonise FAIR and Open data policy

Policies should be aligned and consolidated to ensure that publicly-funded research data are made FAIR and Open, except for legitimate restrictions. The maxim 'as Open as possible, as closed as necessary' should be applied proportionately with genuine best efforts to share.

Several analysis and strategy-oriented reports are being produced around this recommendation. EOSC-Nordic is planning interviews and surveys with policymakers by 2022. EOSC-Pillar developed a legal and policy framework and federation blueprint (December 2020) in the region covered by Pillar (IT, DE, FR, BE, AT). **NI4OS-Europe** is in the process of setting up national Open Science cloud initiatives that will bring together key national stakeholders in academia and research in order to form synergies and work on producing outputs on behalf or as proposals to the Ministries, such as national OS plans and policies. **ExPaNDS's** final extended data policy framework for Photon and Neutron RIs is expected for Aug 2021. FAIRsFAIR's open call on policy support led to good response;

FAIRsFAIR will work on this later this year to explore common elements in policies. An **RDA working group**[34] is working to harmonise policies between funders and scholarly publishers, aiming to arrive at a common template, which FAIRsharing will implement to enable comparison between policies. **ENVRI-FAIR** also has delivered a number of documents in this framework. **OpenAIRE** responds to this recommendation on a practical level via its strong infrastructure and on a policy level via the human network of NOADs (National Open Access Desks).

Missing elements:

- Despite valuable efforts, the actual policy harmonisation, with a corresponding decrease in the number of policies across stakeholders, is missing.

4.3 Pillar 2: FAIR culture

Turning FAIR into Reality Pillar 2: Creating a culture of FAIR data defines how FAIR attitudes and workflows can be improved - by means of interoperability, data management planning, and better recognition for FAIR practice. This also relates to selecting which data to deposit in Trustworthy Digital Repositories, costing data management, and encouraging that FAIR data are actually reused.

◆ Rec. 4: Develop interoperability frameworks for FAIR sharing within disciplines and for interdisciplinary research

"*Research communities need to be supported to develop interoperability frameworks that define their practices for data sharing, data formats, metadata standards, tools and infrastructure.*" (TFiR, page 64)**

Several projects continue work to support interoperability and data exchange. Much of the work centres around **metadata standards and integration**. A key aim for **NI4OS** is supporting the long tail of science and the first part of the project focussed on identifying services from different disciplines and identifying gaps. Building on this work, NI4OS opened a call for experts on data standards and have created a team that will help guide service providers to better support interoperability and data exchange across different disciplines. The expert teams will create best practices on scientific topics and have produced best practices for onboarding. Work on this activity will continue over the next six months. **ExPaNDS** has delivered a draft metadata framework for FAIR data management across the whole experimental lifecycle. The final version will be released next year. **ENVRI-FAIR** is focusing on issues relating to vocabularies, use of the DCAT[35] standard, and work regarding interoperability layers in subdomains (e.g. marine science).

DCAT is also being explored by a number of projects such as **FAIRsFAIR** and **EOSC Pillar** to enable **metadata catalogue integration**. Indeed, several projects are working to support metadata catalogue

---

[34] RDA WG. https://www.rd-alliance.org/funder-publisher-research-data-policy-alignment
[35] Data Catalog Vocabulary (DCAT). DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web. https://www.w3.org/TR/vocab-dcat-3/

integration and federation into EOSC, which addresses a follow-up suggestion made in the 2020 workshop. **ExPaNDS** reviewed data catalogues in photon-neutron (PaN) facilities and undertook a gap analysis and **PaNOSC** is also carrying out work around data catalogues.

Ontologies are being progressed by a few projects. At the time of the event, **ExPaNDS** was in the process of publishing their ontology deliverable[36]. It presents several small ontologies that seek to align semantics in PaN science. The ontologies were designed with FAIR vocabularies in mind. Via **EOSC-Life** and under **ISO ISO/TC 276** a collection of formats, schemas and models for data and metadata is being developed (the **ISO/CD 20691** specification).

EOSC Nordic is currently implementing FAIR, following the framework provided by the FAIR assessment tools (FUJI tool and the Wilkinson et al. evaluator[37]). **FAIRsFAIR** developed a reference implementation of an approach to FAIRifying repository metadata, based on the GO FAIR implementation of a FAIR Data Point. The latter is also part of **GO FAIR**'s Three-point FAIRification Framework[38], together with metadata for machines and community-specific FAIR implementation profiles. FAIRsFAIR is also documenting examples of good practice in developing interoperability frameworks, and of the impacts of these on reuse. In **FAIRplus,** frameworks and processes are being developed to make and keep data FAIR in the life and biomedical sciences, using a number of datasets ranging from omics to clinical**.**

In terms of **interoperability of services**, the **OpenAIRE** guidelines[39] cover interoperability for data repositories/archives and **EUDAT CDI**[40]**,** the EUDAT Collaborative Data Infrastructure, provides several interoperable services. EUDAT have updated the core metadata schema for B2Share storage facility[41] and B2Find discovery service[42], which are now compatible with DataCite metadata schema definition v4.3. **SSHOC** is also working to implement an interoperability hub with a first milestone report outlining the interoperability issues in the Social Sciences and Humanities (SSH) domain. SSHOC has mapped the DDI Codebook and CMDI to its reference ontology (SSHOCro[43]). SSHOC is currently discussing interoperability standards in the context of their marketplace. **EOSC-Pillar** is working towards releasing an MVP (Minimum Viable Product) for the Federated FAIR Data Space[44] (F2DS) in early summer. The F2DS will enable interoperability between the repositories it will harvest.

---

[36] See here https://expands.eu/deliverables/ for further updates

[37] FAIR Evaluator tool. https://w3id.org/AmIFAIR

[38] Three-point FAIRification Framework https://www.go-fair.org/how-to-go-fair/

[39] OpenAIRE Guidelines for Data Archives https://guidelines.openaire.eu/en/latest/data/index.html

[40] https://eudat.eu/eudat-cdi

[41] https://www.eudat.eu/services/b2share

[42] https://www.eudat.eu/services/b2find

[43] SSHOC D4.19 Mapping of two indicative selected standards to the SSHOCro. https://www.sshopencloud.eu/d419-mapping-two-indicative-selected-standards-sshocro-0

[44] EOSC Pillar Federated FAIR Data Space. https://www.eosc-pillar.eu/news/federated-fair-data-space-space-federate-them-all

Possible follow-up actions:

- Research communities, research infrastructures and service providers could go beyond integration of metadata and metadata catalogues, towards other components of disciplinary and eventually cross-disciplinary frameworks.
- Exchange formats for documenting use cases and implementation stories, and coordinate sharing of these (carried over from last year). Broad statements of agreement and good practice are very welcome.

Rec. 5: Ensure Data Management via DMPs

"Any research project producing or collecting research data must include data management as a core element necessary for the delivery of its scientific objectives, and should address this in a Data Management Plan. The DMP should include all the relevant project outputs and be regularly updated to provide a hub of information on FAIR Digital Objects." (TFiR, p. 65)

Data Management Plans (DMPs) are being addressed in some of the projects but not all. Several projects are working to increase capacity for developing and updating DMPs through training and guidance and through policy harmonisation. **SSHOC** has been developing training and advocacy around DMPs to support the implementation of good practice in the community. **SSHOC** is also exploring DMP issues in relation to certain data types. **NI4OS** have also delivered training on developing DMPs and events have been delivered via **EOSC-Hub**. Work to enhance DMP training is planned over the next couple of months. **NI4OS** been promoting FAIR DMPs via the ARGOS tool. **OpenAIRE** provides the ARGOS tool[45] and emphasises the importance of training activities and providing links to guides. OpenAIRE tries to normalise the different types of data and the criteria for reuse in ARGOS. ARGOS also creates links with different research entities and outputs included in the OpenAIRE Research Graph, enabled by the DataCite schema and PIDgraph dashboard. **ENVRI-FAIR** is working to support the harmonisation of policies relating to DMPs in their policy working group. **FAIRsFAIR**'s policy recommendations also address DMPs and place an emphasis on requiring end-stage DMPs to support the provision of provenance and context to inform reuse. FAIRsFAIR recommends that research communities, with support from data stewards, agree on templates for DMPs that identify useful steps towards making data FAIR in a domain-relevant way.

A few projects are working to progress machine-actionable DMPs (maDMPs). **ExPaNDS/ PaNOSC** have undertaken work to create a template of questions which are relevant for photon and neutron researchers. They aim to support the automatic population of DMPs as researchers move through the experimental lifecycle, as reported in the second Synchronisation Force workshop. A deliverable in the later stage of ExPaNDS will focus on how active DMPs will work in the facility and how they will be completed throughout the experimental lifecycle. While this is clearly a domain-specific approach, it isn't clear to what extent Research Infrastructures are currently developing and implementing the

---

[45] ARGOS DMP tool. https://argos.openaire.eu/

"domain protocol" approach for DMPs that was initiated by Science Europe[46]; this was a follow-up suggestion made after the 2020 workshop. OpenAIRE have applied the RDA DMP Common Standard[47] to ARGOS so DMPs are machine actionable. OpenAIRE also created a collection of 841 Horizon 2020 DMPs, as a public resource. A recent exercise on H2020 DMPs highlighted that issues around the licencing of the DMPs themselves can make them difficult to share, which could be an issue for progressing maDMPs. OpenAIRE is currently working on disciplinary aspects with a couple of research communities to implement Domain Data Protocols. Work was also undertaken with ARIADNE+ to develop a machine-actionable DMP instance for archaeological data. OpenAIRE has been approached by publishers to develop a bridge between data availability statements and DMPs. The aim has been to help close the DMP publication lifecycle through integration with Zenodo and in the future with other repositories. FAIRsFAIR is also working with the RDA Interest Group on Active DMPs[48] to progress recommendations on machine-actionable DMPs.

Possible follow-up actions:

- More research communities and Research Infrastructures could explore machine-actionable DMPs for their domains.
- Likewise, more research communities should provide input to the disciplinary aspects of DMPs, in the form of domain protocols (TFiR Action 5.3).[49]

Rec. 6: Recognise and reward FAIR data and data stewardship

"*FAIR data should be recognised as a core research output and included in the assessment of research contributions and career progression. The provision of infrastructure and services that enable FAIR data must also be recognised and rewarded accordingly.*" (TFiR, page 65)

A couple of projects are working on rewards or incentives or are including it in institutional or domain policies. In the German climate computing data centre (**DRKZ**) a DataCite DOI is only given for quality-assured data that complies with the submission guidelines such as domain-specific standards and the provision of rich metadata, in order to ensure long-term reusability. The same demand on highly reusable data holds for publishing data to the ESGF (Earth System Grid Federation) via the DKRZ node of the federation. This way, FAIR data stewardship or data generation is rewarded by obtaining a DOI.

---

[46] Science Europe: Science Europe Guidance Document Presenting a Framework for Discipline-specific Research Data Management. https://www.scienceeurope.org/media/nsxdyvqn/se_guidance_document_rdmps.pdf

[47] RDA DMP Common Standard for Machine-actionable Data Management Plans is the outcome of the RDA DMP Common Standards WG
https://www.rd-alliance.org/group/dmp-common-standards-wg/outcomes/rda-dmp-common-standard-machine-actionable-data-management

[48] https://www.rd-alliance.org/groups/active-data-management-plans.html

[49] A new RDA working group called "Discipline-specific Guidance for Data Management Plans WG" has been set-up to provide support to disciplinary aspects of DMPs https://www.rd-alliance.org/groups/discipline-specific-guidance-data-management-plans-wg

DRKZ is working on approaches to make the quality of archived data visible for data reusers through the EASYDAB[50] concept developed in the framework of the AtMoDat[51,52] project. **ExPaNDS** observes that in the photon and neutron domain data stewardship often falls on the instrument scientists, not on data stewards. In practice the scientists are just expected to take this on as yet more additional work. There isn't really a widely promoted concept of formal recognition for that. What also affects recognition is the cultural question in the PaN research infrastructures about the issue of authorship of the data, that is, how to balance the observation made by the facility and the way the data are put to use by the scientist.

With respect to **policies, OpenAIRE'**s Open Science policy templates include rewards and incentives for FAIR data and stewardship; their NOADs (National Open Access Desks) communicate it to national academic and research stakeholders. Similarly, in its blueprint for national Open Science policies **NI4OS** advises to incorporate reward and incentives. **GO FAIR** addresses and informs research funders about how they can realistically mandate and value FAIR data. Earlier this year **EOSC Nordic** started a study of existing incentives for FAIR uptake, aiming for more consistency. Their project coordinator NeIC provided data stewardship training to more than 160 individuals in the nordic region in the hope that institutions will support culture changes on data stewardship and FAIR in general.

Repositories also play a role in crediting and being credited. **EOSC Nordic** considers the FAIR assessments (See also Pillar 5) they carry out to be an incentive to FAIRify data for some of the repositories involved. **ENVRI-FAIR** notes that partner repositories have an important responsibility for ensuring that correct credit is given to all involved in the data production and curation processes. **SSHOC** wants to reward contributions to community curation in the SSHOC Open Marketplace and is considering badging[53] as an instrument for this.

For the topic of career progression, see Pillar 4.

Possible follow-up action:

- The TFiR actions associated with this recommendation address research funders. Research communities, research institutions and research infrastructures should include rewards & incentives metrics and support tools in their design so that they can actually implement the funder policies.

---

[50] Earth System Data Branding, https://www.easydab.de

[51] Atmospheric Model Data: Data Quality, Curation Criteria and DOI-Branding, https://www.atmodat.de

[52] Ganske, A., et al., 2020. A short guide to increase FAIRness of atmospheric model data. Meteorologische Zeitschrift, https://www.doi.org/10.1127/metz/2020/1042

[53] Badging was a discussion topic in the second Synchronisation Force workshop, see p. 18 in Ingrid Dillo, Marjan Grootveld, Simon Hodson, & Sara Pittonet Gaiarin. (2020). Second Report of the FAIRsFAIR Synchronisation Force (D5.5) (Version 1.0). https://doi.org/10.5281/zenodo.3953978

Rec. 18: Cost data management

"Research funders should require data management costs and other relevant costs to be considered and included in grant applications where relevant. To support this, detailed guidelines and worked examples of eligible costs for FAIR data should be provided." (TFiR, page 71)

Research funders were hardly represented in the workshop. Nevertheless, the topic of costs is being addressed in a few of the projects represented. **ARCHIVER** focusses on commercial long-term archiving and preservation services for petabyte scale datasets. The ARCHIVER contractors have an initial deliverable about cost effectiveness as part of the decision making for the solution architecture, describing all aspects of the architecture that influence the future cost effectiveness of the resulting services. ARCHIVER works on use cases that make the costs of data stewardship more predictable. **DRKZ** observe that the awareness that DM requires dedicated funding, especially in terms of FTEs, has risen in the community. Members of the **OpenAIRE** AMKE (the OpenAIRE legal entity) have developed guidance on the costs of data management, also in relation to funder requirements[54]. The guidance can be used in trainings, which is also the approach taken by **SSHOC** and **NI4OS**. NI4OS supports both researchers in calculating costs through the data lifecycle and institutions in calculating the cost of building curation services. High-level responsibility for costs is a theme of the **ExPaNDS** data management framework. Their DMP template contains several questions in relation to costs, but this presents a challenge to facilities, as it is difficult to calculate cost information. Understanding costs and budgeting of data stewards and research software engineers is part of **FAIRsFAIR** interviews and planned guidance.

Missing element:

- This recommendation addresses research funders. Given the low level of participation to the workshop by research funder representatives, this report does not cover the funder-side activity at all.

Rec. 19: Select and prioritise FAIR Digital Objects

"*Research communities and data stewards should develop and implement processes to assist the appraisal and selection of outputs that will be retained for a significant period of time and made FAIR." (TFiR, page 71)*

The FAIR Digital Object (FDO) concept is being adopted, but work to support selecting and prioritising them for long-term reusability is underrepresented among the participating projects. **EOSC-Nordic** strongly recommend the FAIR Digital Object concept to the 100 repositories in their sample and observe that it is indirectly expected and rewarded in FAIR assessment tools such as F-UJI[55] and the

---

[54] OpenAIRE cost guidance. https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs
[55] F-UJI tool. https://doi.org/10.5281/zenodo.4063720

FAIR Evaluator tool by Wilkinson et al[56]. **EOSC-Pillar** is evaluating the impact that the FDO concept could have on services and how it could be implemented in their FAIR Data Space. **GO FAIR** provides support for nanopublication-based FDOs in the form of an online FAIR Implementation Profiles (FIP) Wizard[57].

The Persistent identifier (PID) is an essential element of a FAIR Digital Object. **ExPaNDS** raises the question to which data and in what stage of the process the PID should be assigned: to the 'raw' data or later in the process. The PaN community has much data, which goes through many processes throughout the data lifecycle. ExPaNDS runs a series of workshops for librarians and data managers, in which they discuss what data needs to be curated to what level, and what data needs to be retained in the long term, i.e., how they can be selective in the data made FAIR. Selection is also addressed in the Draft extended data policy framework for Photon and Neutron RIs[58].

In the context of selecting research outputs **ENVRI-FAIR** is well aware that observational data collected by environmental or earth science research initiatives are unique, as they represent the state of a given point in space and time that cannot be reproduced. From the perspective of long-term preservation services, **ARCHIVER** offers support for handling unstructured and missing metadata to support making data FAIR.

Missing elements (carried over from 2020)**:**

- In particular research communities and Research Infrastructures should play a larger role in defining and refining policies about what to keep and make FAIR and what not to keep. Ideally this relates to data management planning, addressing FAIRness and the level of data curation at each stage of the data lifecycle.

Rec. 20: Deposit in Trusted Digital Repositories

"R*esearch data should be made available by means of Trusted Digital Repositories, and where possible in those with a mission and expertise to support a specific discipline or interdisciplinary research community.*" (TFiR, page 72)

FAIRsFAIR, **EOSC-Nordic**, and **SSHOC** continue working with certified repositories or supporting the certification of repositories via the CoreTrustSeal (CTS) Requirements[59]. In addition to its support programme for ten repositories and working towards linking certification guidelines and FAIR principles, **FAIRsFAIR** reaches out to researchers and data stewards with a lightweight educational

---

[56] FAIR Evaluator tool. https://w3id.org/AmIFAIR

[57] FIP Wizard. https://fip-wizard.readthedocs.io/en/latest/about/about.html

[58] Brian Matthews, Abigail McBirnie, Andrei Vukolov, … Mirjam Van Daalen. (2020). Draft extended data policy framework for Photon and Neutron RIs. https://doi.org/10.5281/zenodo.4014811

[59] https://www.coretrustseal.org/why-certification/requirements/

tool to assess their knowledge of FAIR[60]. To help researchers to identify FAIR-enabling repositories, the Repository Finder tool - a filter on the re3data.org repository registry - has been implemented in the DataCite Commons discovery tool[61]. Users can search the connected PID graph for connections between data sets and other research output (via DOIs), people (via ORCID), and organisations (via ROR). FAIR-enabling repositories are also discoverable via FAIRsharing faceted search, and the results will show which (meta)data standards and identifier schemas these resources implement[62]. **EOSC Nordic** is supporting repositories in considering CTS certification or self-assessment against CTS requirements. **SSHOC** also runs a certification support programme[63]. Several cross-links among participating projects exist, for instance between SSHOC and NI4OS-related repositories, and between FAIRsFAIR and repositories related to **ENVRI-FAIR** and **EOSC Synergy.**

The **OpenAIRE** Validator[64] allows for repositories that are part of the OpenAIRE network to be certified on their metadata quality and completeness by checking metadata they expose against the OpenAIRE guidelines. Work is also being done on APIs to search across all facility catalogues in EOSC (**ExPaNDS/PaNOSC**) and on certifying certain RDM processes (**NI4OS**). The recently started **DICE** project includes a task for long-term preservation that will draft a strategy for long-term preservation of EUDAT services.

Possible follow-up actions (carried over from 2020)**:**

- There is a need to provide outreach to and concrete support for researchers, e.g. on how to select Trustworthy Digital Repositories for data deposit.

◆ Rec. 21: Encourage and incentivise reuse of FAIR outputs

"Funders should incentivise the reuse of FAIR outputs when appropriate by promoting this in funding calls and requiring research communities to seek and build on existing content wherever possible." (TFIR, page 72)

This recommendation addresses research funders, which were considerably underrepresented at the workshop. Contrary to last year however, several participants now felt their projects can play a role in encouraging the reuse of FAIR outputs. **Making reuse easier** is the way to do this. **ExPaNDS** is federating into EOSC to encourage the findability of data generated at PaN Research Infrastructures. Likewise, **ENVRI-FAIR** will provide the ENVRI-hub to advertise and make the ENVRI Community's FAIR data products and services easy to discover and use, both for research and innovation. They also facilitate comprehensive statistics of usage. Through training and outreach activities via e.g. conferences ENVRI-FAIR informs and educates end user communities on how to leverage FAIR data

---

as inputs to their research, and how to make their outputs FAIR in their turn. Sharing of notebook or workflow methods between users can be a means as well. The **ARCHIVER** architecture supports compute capabilities that allow scientific analyses to be carried out off-premise and provides interfaces that allow access to data no matter where they are stored.

Possible follow-up action:

- There is a need to bring research funders together in order to take stock of and encourage relevant activities on their side. Research communities, research institutions and research infrastructures should prepare for such incentives their design so that they can actually implement the funder policies.

4.4 Pillar 3: FAIR ecosystem

Turning FAIR into Reality Pillar 3: Creating a technical ecosystem for FAIR data centers around FAIR Data Objects. It addresses semantic technologies and automated processing, as well certification of FAIR services and incentives for research infrastructures. Ecosystem components should meet research needs and information in DMPs should be used.

Rec. 7: Support semantic technologies

"Semantic technologies are essential for interoperability and need to be developed, expanded and applied both within and across disciplines." (TFiR, page 66)

The report of the second Synchronisation Force workshop suggested that about half of the INFRAEOSC-5b projects and ESFRI cluster projects focus their work on the support, promotion and use of standards, vocabularies and ontologies, and/or standard metadata schemes[65]. Updates provided during the third workshop showed that several projects are actively working to provide guidance and develop recommendations on data exchange standards as well as promoting existing guidelines and resources. Some projects are focusing on developing services and tools to support data exchange and FAIR data assessments. Several are carrying out assessments on the FAIRness of data holdings within repositories.

Building on earlier work, **FAIRsFAIR** has developed a set of data object assessment metrics and based on these, has provided free access to F-UJI which is a tool to support the programmatic assessment of the FAIRness of data holdings. FAIRsFAIR has been working with a number of trusted digital repositories to scope and carry out assessments using the tool and refining it based on feedback. **EOSC-Nordic** is working with repositories in the Nordic region to assess the FAIRness of their data holdings. Using FAIRsFAIR's F-UJI tool, which after testing was found to better meet the project's

---

needs, EOSC Nordic is carrying out quarterly re-assessments of the repositories in the sample to follow their FAIR maturity journey. The assessment includes 7 metadata metrics and there is an emphasis on improving these areas by providing advice and practical recommendations on data standards. The project has found that metadata-related scores are not improving as quickly as hoped, and it now aims to explore some of the challenges in more detail to better understand how to support progress.

NI4OS's primary focus with regard to repositories is to provide general support and share good practices such as those as outlined in OpenAIRE's guidelines[66]. While not a key focus for the project, NI4OS is also carrying out some assessments on the FAIRness of data holdings and recently put together a team to start implementing F-UJI assessments. **EOSC Synergy**, meanwhile, is working to make services aware of the standards and good practices outlined in the EOSC Interoperability Framework with an ultimate objective of supporting those services participating in the project to a common level. A mid-point deliverable is expected soon which will outline this work and a handbook will be shared towards the end to help those services not involved in the project.

Among the discipline-oriented cluster projects, **ENVRI-FAIR** is developing training for environmental researchers on how to use vocabularies as part of good research practice. A key objective for ENVRI-FAIR is raising awareness of semantic technologies and issues such as data provenance among the ENVRI-FAIR community which has been quite successful. A series of training events will begin in June 2021. ENVRI-FAIR will publish a report on their work in early June followed by a set of recommendations due in September. **SSHOC**, meanwhile, has recently begun work with data communities in the Social Sciences and Humanities (SSH) to design a discipline specific knowledge graph. The **Dataverse** web application[67] is being tested as a semantic gateway in SSHOC and this work will ultimately result in recommendations for implementing Dataverse at the institutional level. SSHOC is drawing on the **PARTHENOS** project[68] to develop a semantic annotation framework. An important issue raised during the discussion on support was the need to ensure that additional help will be needed by countries and institutions that are newcomers to such activities, such that they can get up to speed with the quickly evolving and complex landscape otherwise there is a risk that the current gap could widen.

A suggested follow-up action from the second Synchronisation Force was that FAIRsFAIR, SSHOC, ExPaNDS and PaNOSC should explore potential synergies in relation to their respective metadata catalogue activities. This was undertaken and, as a result, **FAIRsFAIR** is working closely with **PaNOSC**, **ExPaNDS** and **SSHOC** to carry out a metadata integration pilot to map domain specific standards to DCAT2 in summer 2021[69]. **EOSC Pillar** is also using DCAT as they are developing and deploying a

---

[66] OpenAIRE Guidelines for Data Archives. https://guidelines.openaire.eu/en/latest/data/index.html

[67] https://dataverse.org/about

[68] https://www.parthenos-project.eu/

[69] Eva Mendez, Tony Hernandez, Angus Whyte, & Joy Davidson. (2020). D3.6 Proposal on integration of metadata catalogues to support cross-disciplinary FAIR uptake.https://doi.org/10.5281/zenodo.4134787

Federated FAIR Data Space[70] which is based on the FAIR Data Point (FDP) concept. The FAIR Data Space will harvest metadata from repositories and map these to the DCAT format. EOSC Pillar is aiming to improve the Federated FAIR Data Space with semantic technologies but this is still in the development phase. Through their work on developing the **ENVRI Hub**[71], **ENVRI-FAIR** have been drawing on the RDA InteroperAble Descriptions of Observable Property Terminology WG (**I-ADOPT**)[72] work which may prove useful for other research infrastructures for improving search facet in portals. I-ADOPT offers a mapping framework for machine-readable and interoperable variable descriptions across semantic artefacts.

Missing element:

- While there appears to be a good level of activity among the projects in implementing semantic technologies in some domains, there remains a need to provide support (including training) for domains (and institutions and countries) that have made less progress, so that the current gap in experience and implementation does not increase further.

Rec. 8: Facilitate automated processing

"Automated processing should be supported and facilitated by FAIR components. This means that machines should be able to interact with each other through the system, as well as with other components of the system, at multiple levels and across disciplines." (TFiR, page 66)

Several projects are looking into the development of machine actionable pipelines and testing aspects of the EOSC Interoperability Framework. In most cases, however, work is being carried out to support the automation of research *related or supporting* processes (such as use of FAIR assessment) rather than research processes themselves. The exceptions are those communities of practice where the processing of large amounts of data is so fundamental to the research that automation of research processes have already largely been implemented (e.g., astronomy).

Some of the current activities in this area focus on defining workflows in relation to FAIR assessment. For example, **EOSC-Nordic** are using the F-UJI tool to support the automation of FAIR data assessment and are currently working to define how and when such assessments should be carried out within research workflows. **EOSC Synergy** is currently working to develop criteria for assessing software quality and maturity. They are developing a pipeline orchestrator and looking at tools available for carrying out the assessments against the criteria (e.g. F-UJI). Following this, the project will aim to develop pipelines for implementing automated assessments. The ultimate objective is that

---

[70] EOSC Pillar Federated FAIR Data Space. https://www.eosc-pillar.eu/news/federated-fair-data-space-space-federate-them-all

[71] ENVRI Hub https://envri.eu/envri-hub/

[72] InteroperAble Descriptions of Observable Property Terminology WG (I-ADOPT WG) https://www.rd-alliance.org/groups/interoperable-descriptions-observable-property-terminology-wg-i-adopt-wg

automated and efficient FAIR assessment will help increase the amount of FAIR data and software that is accessible and reusable via the EOSC Portal.

In the domain-focussed cluster projects, the main focus is on automating interoperability rather than FAIR data assessment. For instance, **EOSC Pillar** is working to support automated processing through mappings between their FAIR Data Space and repositories. **ENVRI-FAIR** is encouraging the reuse of FAIR processing workflows to ensure that researchers and other stakeholders in ENVRI FAIR can find data across the various Research Infrastructures (RIs). The project has developed a Knowledge Base which stores information about FAIR assessments and, if some aspect is lacking, suggestions are provided on how to reuse existing processing workflows. ENVRI-FAIR is also participating in activities to progress FAIR digital object discussions with efforts to engage at both the RI and subdomain levels. A key area for discussion revolves around the metadata and semantics for data 'types', data structure, provenance, and consistent variable definition, in order to support machine actionability, and which are crucial for enabling services to work with automated workflows. **SSHOC** has been working to progress both manual and automated curation workflows in relation to data curation workflows for the SSHOC marketplace. The main aim is to work towards realising semi-automated processes and a deliverable about this activity is due to be published late 2021.

With regard to data management plans (DMPs), there is work underway in **OpenAIRE** to automate the processes relating to the creation and sharing information in the Argos tool (e.g. to notify repositories of the future deposit of a very large dataset). **FAIRsFAIR** is documenting similar examples using other DMP platforms. One of the follow-on actions suggested during the second Synchronisation Force workshop was that **FAIRsFAIR**, **EOSC-Nordic**, **ExPaNDs** and **PANOSC** should explore potential synergies on machine-actionable metadata and DMPs. This is being progressed collaboratively through implementation stories on maDMPs being drafted in FAIRsFAIR[73].

Follow-up action:

- A lot of useful work is proceeding to help automate tools used in support of research and to assist RIs, RPOs and repositories (automation of FAIR assessment tools, automation of aspects of DMPs and the use of DMP information). However, the main objective of this recommendation (implementation of FAIR and semantic technologies to assist with the automation of research processes) seems to have received less attention beyond certain domains where there are already some automated data processing pipelines and workflows. Although challenging, further attention to this key aspect of the recommendation is needed.

Rec. 9: Certify FAIR services

"Data services must be encouraged and supported to obtain certification, as frameworks to assess FAIR services emerge. Existing community-endorsed methods to assess data services, in particular CoreTrustSeal (CTS) for trusted digital repositories, should be used as a starting point to develop

---

[73] https://www.fairsfair.eu/adoption-stories

assessment frameworks for FAIR services. Repositories that steward data for a substantial period of time should be encouraged and supported to achieve CTS certification." (TFiR, page 67)

The large majority of projects are involved, albeit to varying degrees, in activities that will contribute to the implementation of this recommendation. In general, there is an emphasis on certifying repositories as key service providers in the FAIR ecosystem. With regard to the certification of FAIRness and FAIR-enabling services, suggestions were made during the WS2 (2020) that metrics should be both quantitative and qualitative and that as metrics are still being defined and tested, they should be used at this stage to provide guidance on constructively improving services. Responding to a related follow-up action, work is underway to explore for what types of data services (in addition to repositories) FAIR-enabling assessment and certification is valuable and feasible. At the time when the recommendation was written no-one agreed whether it was even possible to call services 'FAIR', much less what properties they needed to have to call them that. Pprojects all are working to establish that.

FAIRsFAIR continues to support a cohort of 10 repositories to reach CoreTrustSeal (CTS) certification. In addition, FAIRsFAIR is developing its concept of *CTS+FAIR* which is an extension of CTS that better aligns with the FAIR principles. In addition to supporting repository certification, FAIRsFAIR has developed a draft assessment framework for assessing a broader range of FAIR-enabling services[74]. A workshop was held to introduce the draft framework on May 20, 2021[75] and to allow for community feedback to be collected. **EOSC-Nordic** is similarly working with 10 repositories to either achieve CTS certification or to carry out self-assessments against the CTS criteria.

ENVRI-FAIR is working on defining requirements and common elements among the individual RIs who are working on repository certification. ENVRI-FAIR is contributing to FAIRsFAIR work on developing a FAIR enabling service assessment framework. **SSHOC** is currently working with 14 repositories to progress toward certification. A SSHOC deliverable related to this work has been published and aims to help repository managers consider whether they should seek certification[76]. The report aimed to focus on inclusiveness and accordingly includes information targeted to smaller repositories. Further outputs related to the certification support in SSHOC are due in early 2022. Some discussion has focused on whether different repositories can be certified with the same set of requirements: for example, should the same set of certification criteria be applied to those repositories that are operated at the national level or by ESFRIs, as to those that are at run at the institutional level.

Missing Element:

- CoreTrustSeal remains challenging for institutional repositories dealing with a much broader range of content that domain repositories. By engaging with smaller repositories, SSHOC's

---

[74] Hylke Koers, Patricia Herterich, Rob Hooft, Morane Gruenpeter & Tero Aalto. (2020). M2.10 Report on basic framework on FAIRness of services (Version 1.0). https://doi.org/10.5281/zenodo.4292598

[75] FAIR-enabling Services: Validating The Framework. https://fairsfair.eu/events/fair-enabling-services-validating-framework

[76] Mari Kleemola, Tuomas J. Alaterä, Niko Koski, … René Van Horik. (2020). SSHOC D8.2 Certification plan for SSHOC repositories. https://doi.org/10.5281/zenodo.4558303

work addresses this to some extent, there remains a need for greater clarity on whether and how smaller repositories should engage with certification processes. Is there a need for a further onramp process, greater support or other steps to support smaller and institutional repositories? FAIRsFAIR is addressing this issue and will provide recommendations as a result of its TDR Support Programme[77].

A few projects are working to develop digital badging schemes. **EOSC Synergy** has carried out a state-of-the-art review regarding digital badge issuing in relation to FAIR data assessments[78]. The EOSC Synergy module for FAIR assessment could potentially provide a digital badge based on a F-UJI assessment. **FAIRsFAIR** is similarly investigating the badging of datasets that have undergone F-UJI assessments with work beginning in spring 2021. **ELIXIR** has a formal selection process for Core Data Resources[79] and Recommended Interoperability Resources[80], and associated badge schemes.

Several projects are working to provide general support in the area of certification of FAIR services. For example, **OpenAIRE** is working to align their existing guidelines with FAIR to support services to assess their FAIR-enabling capacity while **GO FAIR** is working to support the certification of vocabularies and metadata schemas.

For some projects, repository certification is not a core activity, however, there is interest to see an increase in the availability of certified trusted digital repositories (TDRs) to ensure that the EOSC has a larger pool of FAIR data to make accessible for reuse. For instance, **EOSC Pillar** considers certified repositories as essential components of the FAIR ecosystem for feeding into their FAIR Data Space and they continue to monitor FAIRsFAIR activity around certification.

Rec. 22: Use information held in DMPs

"DMPs hold valuable information on the data and related outputs, which should be structured in a machine- actionable way to enhance reuse. Investment should be made in DMP standards and tools that adopt common standards and support 'active' DMPs to enable information exchange across the FAIR data ecosystem." (TFiR, page 72)

As noted in the report from the second Synchronisation Force workshop, only some projects are actively working in this area. **ExPaNDs** and **PaNOSC** continue work to develop and trial an approach for 'active DMPs', which integrates with the experimental lifecycle and assists automated metadata collection throughout the lifecycle. **EOSC Synergy** provides some process guidance for developing DMPs for services joining the EOSC.

---

[77] https://www.fairsfair.eu/application-results-open-call-data-repositories

[78] German Molto et al. (2020). State of the art regarding digital badge issuing technologies. http://hdl.handle.net/10261/206348

[79] https://elixir-europe.org/platforms/data/core-data-resources

[80] https://elixir-europe.org/platforms/interoperability/rirs

SSHOC does not actively focus on DMPs in their project but are trying to promote and support the publication of DMPs among the SSH domain. The participants stressed that as many communities are still in the process of implementing good practice around developing and updating DMPs there is a limit to what can be done to reuse the information held in DMPs. Until good practice is fully embedded, we will not be able to extract content from DMPs at scale.

Missing element:

- As community good practices are relatively new with regard to DMPs, it is recommended that this theme should be explored more fully in the next generation of thematic projects.

Rec. 23: Develop components to meet research needs

"While there is much existing infrastructure to build on, the further development and extension of FAIR components is required. These tools and services should fulfil the needs of data producers and users, and be easy to adopt." (TFiR, page 73)

During the second Synchronisation Force Workshop, there was strong consensus that this supporting action should be made a higher-priority in TFiR in recognition of the crucial importance of ensuring that services in the European Open Science Cloud are researcher-centric and meet research needs. This was echoed during discussions during the third workshop which focused on some of the challenges of engaging with end users communities.

ENVRI-FAIR is working to ensure they meet researchers' needs through dedicated activity to collect researchers' requirements. The work involves collaborations between the RIs of four subdomains and aims to come up with concrete use cases to test the FAIRness of services that are currently available as well as informing the development of those services still in the planning stage. **SSHOC** also agrees on the need to align with researchers' needs and the need to secure a better understanding of *potential* users' needs. SSHOC are in the early stages of user testing of the SSHOC Marketplace which will involve about 150 testers from the research communities. SSHOC also has a dedicated Data Communities work package which is doing valuable work in reaching out to current and potential users.

In **FAIRplus**, the FAIRification process, its indicators and the necessary tools and services, are defined and selected with the researchers from the pharmaceutical industry (including GSK, AstraZeneca, Pfizer, Jenssen, Bayer) involved in the projects. This comes with its own challenges, due to the confidential nature of some datasets and the formal process required to obtain access, but it allows anchoring of the work to real use cases.

During the **ENVRIplus** project it became clear that in many cases the people involved in managing and operating the research infrastructure have one foot in the research community and one in the RI. In such cases, it is often the case that the needs of specific subdomains represented among the RI partners are very well understood but the needs of the broader actual and potential user base of the RI are less well understood. There is also often a difference between well-versed users (esp. in large

organisations) and new users when it comes to being vocal about usability needs. We must ensure that there is an equal focus on feeding in broad-based usability needs from the longer tail (smaller research groups, etc). Having one foot in research and one in the RI is something that is also encountered in **SSHOC**. There are also ongoing challenges associated with how we can obtain input on needs from the longer tail of the research community. It is also often difficult to secure participation in surveys and/or testing which can be time consuming for participants. SSHOC's forthcoming deliverable on challenges in engaging with the community is due in April 2022 and could have valuable insights for others.

Rec. 24: Incentivise research infrastructures to support FAIR data

"Research facilities, in particular those of the ESFRI and national Roadmaps, should be incentivised to provide FAIR data by including it as a criterion in the initial and continuous evaluation process. Investments should be made strategically and consider data service sustainability." (TFiR, page 73)

As noted in the report of the second Synchronisation Force Workshop, there were few activities relating to incentivising research infrastructures (RIs). **EOSC Nordic** is continuing to progress its mapping of existing policy incentives across Nordic countries and work to engage with stakeholders with the aim to develop and promote new and/or efficient incentives. A number of interviews will be carried out with stakeholders over summer/autumn 2021. EOSC Nordic is also engaging with policy makers to try to harmonise the policies in this region. **GO FAIR** is working to promote the "Metadata for Machines" workshop series (M4M)[81] with a rapid pipeline and handbook for research infrastructures.

Several projects are working to support RIs indirectly to be more able to support FAIR data provision, such as **EOSC-Pillar's** approach to support data exchange through their FAIR Data Space which will connect data islands. Also noted in the WS2 report, incentivising research infrastructures to support the production and use of FAIR data requires active involvement and commitment from a range of national and international stakeholders. Discussions during WS3 highlighted national roadmapping exercises as a potentially valuable tool for engaging with politicians.

The participants of the third workshop also observed that while there is an emphasis placed on RIs to engage with politicians and policymakers, there may be less of an emphasis for them to reach out to their user communities to encourage them to make their research data outputs FAIR.

Missing element:

- There is a need to ensure that research infrastructures are working to encourage and support their user communities not only to reuse others' FAIR data but also to commit to creating and

---

[81] The M4M workshops are agile, hackathon-style events to assess the state of metadata practices in data-related communities and stimulate the creation and re-use of FAIR metadata standards and machine-ready metadata templates (definitions of metadata categories). They have been launched by GO FAIR and RDA members https://www.go-fair.org/resources/go-fair-workshop-series/metadata-for-machines-workshops/

contributing to the body of FAIR data available. Training will be crucial for realising this necessary culture change.

Possible follow-up action:

- Understanding the funder perspective and actions on this recommendation (and others) is recommended in order to have them engaged as an active and driving partner. A workshop would be a good start.

4.5 Pillar 4: Skills for FAIR

Turning FAIR into Reality Pillar 4: Skills and capacity building identifies the need for professionalising roles and curricula in data science and data stewardship.

Rec. 10: Professionalise data science and data stewardship roles and train researchers

"Steps need to be taken to develop two cohorts of professionals to support FAIR data: data scientists embedded in research projects, and data stewards who will ensure the management and curation of FAIR data. All researchers also need a foundational level of data skills." (TFiR, page 67)

Most projects are continuing to organise and offer training: for data stewards (e.g. **EOSC Nordic** and **FAIRsFAIR**), service providers (**NI4OS Europe**), data providers (**ESCAPE**), and researchers (many). Training topics include reproducibility, awareness of data skills, data FAIRification, data science, Open Science, EOSC, science with interoperable data, and sharing COVID-19-related data. Depending on the type of stakeholder, training is local, disciplinary, national or cross-cutting. **EOSC Nordic** and **EOSC Synergy** have a multi-country remit, and the latter is working on a data science programme for universities in its partner countries. The EOSC Synergy learning platform[82] also supports course development.

In addition, train-the-trainers' workshops are organised by ExPaNDS/PaNOSC about creating courses in the PaN training portal. SSHOC set up three data communities in support of the train-the-trainer bootcamps they run for a wide variety of SSH disciplines. In the framework of the **CODATA-RDA Schools of Research Data Science**[83], **FAIRsFAIR** has developed data steward instructor training workshops to be run on a regional basis, identifying that most data stewards have only recently been co-opted or employed in this role; hence building regional communities of data stewards is essential to enable sharing of training resources, instructors and best practice[84]. These workshops incorporate advanced training, pedagogy and community building. To bring such initiatives and expertise together, OpenAIRE provides a Community of Practice for training coordinators[85].

---

[82] EOSC-Synergy Learn Platform. https://learn.eosc-synergy.eu/

[83] https://www.datascienceschools.org/

[84] https://www.fairsfair.eu/events/training

[85] Community of Practice of training coordinators. https://www.openaire.eu/cop-training

The **FAIRsFAIR** FAIR data competence framework aims to support the uptake of FAIR data principles and practice by higher education institutes (HEIs). Work will continue in order to offer comprehensive guidelines for HEIs to implement FAIR data competences in their curricula[86]. (See also Recommendations 11).

FAIRsFAIR is also working with the RDA Interest Group on Professionalising Data Stewardship[87] on a survey of the community to identify models for providing data stewardship, and for developing the role in conjunction with other professional groups such as Research Software Engineers. This will discern, for example, national variations in job titles used, and the extent to which stewardship is also done by 'research data managers' and 'data librarians'. Examples of collaboration between these groups on developing and accrediting training for researchers are being documented.

Little other activity around establishing professional bodies or accreditation of training and trainers is reported, with the exception of the **GO FAIR** Foundation Certification Pioneer Programme[88].

FAIRplus is developing a Fellowship Programme[89] in FAIR Data Management, with expert members of the FAIRplus consortium; ultimately the fellow will be able to advise and initiate FAIR data processes in their respective companies and organisations.

A new development can be seen in the **EOSC Association** Task Forces "Data stewardship curricula and career paths" and "Upskilling countries to engage in EOSC", which started in May 2021. Multiple stakeholder groups will be addressed, specifically research communities, data stewards and national bodies. The Task Forces will build on for instance the recommendations from the EOSC Working Group on Skills and Training[90] to define data stewardship curricula and ensure these are recognised and aligned across Europe. Attention will also be paid to career paths to ensure appropriate recognition and rewards for data management activities. For more information on curriculum development, see recommendation 11.

Missing element:

- Significant work is underway to provide training and to further develop curriculum frameworks. More activity is needed to develop certification, to establish professional bodies, to encourage institutions to recruit staff to these posts, and to develop a cohort of data stewards and data scientists to support projects.

---

[86] Three additional reports are expected to be published by FAIRsFAIR by February 2022. See
https://www.fairsfair.eu/fair-frameworks-training-programmes

[87] https://www.rd-alliance.org/groups/professionalising-data-stewardship-ig

[88] https://www.gofairfoundation.org/certification/pioneer-program/

[89] FAIRplus fellowship programme. https://fairplus-project.eu/get-involved/fellowship

[90] Michelle Barker, Natalia Manola, Vinciane Gaillard, Iryna Kuchma, Emma Lazzeri and Lennart Stoy (2021). Digital skills for FAIR and Open Science. https://op.europa.eu/en/publication-detail/-/publication/af7f7807-6ce1-11eb-aeb5-01aa75ed71a1/language-en

Rec. 11: Implement curriculum frameworks and training

"A concerted effort should be made to coordinate and accelerate the pedagogy for professional data roles. To support uptake, skills transfer schemes, fellowships, staff exchanges and informal training opportunities are needed, as well as formal curricula." (TFiR, page 68)

Building on earlier projects such as EDISON and EOSC-Pilot, **FAIRsFAIR** has developed a "FAIR Competence Framework for Higher Education"[91], which will form the basis for a practical FAIR adoption handbook for universities - currently under development. The "FAIR Adoption Handbook for Universities" will be a practical tool for universities to implement research data management (RDM) and FAIR-related content in university programmes at the different levels (Bachelor, Master and Doctoral). It will provide universities with model curricula, courses and supporting materials to enhance provisions for FAIR data training and support the uptake of FAIR skills and competences. A good practices report is also in development, which will analyse examples coming from different European higher education institutions and make recommendations for universities that aim to start or improve their own training activities related to RDM and FAIR. The report will complement the adoption handbook when both are published in December 2021. **FAIRsFAIR** and **OpenAIRE** co-organised a workshop on national policy and support actions for research data skills. These are domain-independent activities. In the context of recommendation 11 it was remarked that the balance between generic and domain-specific data stewardship is not fixed: organisations position their data stewardship capacity differently, e.g. embedded in a research team or at central research support. The Dutch government recognises this in the local Digital Competence Centers, which fund and coordinate data stewardship in research institutions. In on-the-job learning, training and (emerging) curricula for data stewards there may be a need for balancing domain-specific and domain-agnostic competences and content.

For the life sciences, in **FAIRplus** and **ELIXIR** the **FAIR Cookbook**[92] is being created as a hands-on online training resource. The FAIR Cookbook is already on the EC Innovation Radar[93], and will feature in the upcoming version of the Guidance to Researchers by the Innovative Medicines Initiative (IMI)[94].

For harmonising curriculum development and training provision itself, the **terms4FAIRskills**[95] terminology supports describing the competencies, skills and knowledge associated with making and keeping data FAIR; among others this is scheduled for implementation (to facilitate annotation and discovery of the content according to skills) in ELIXIR TeSS, FAIRsharing Resource Selectors and the FAIR Cookbook.

EOSC Synergy, **ExPaNDS/PaNOSC**, **ENVRI-FAIR**, and **SSHOC** set up training platforms, portals or catalogues. **ENVRI-FAIR** is also involved in the training work package of the recently started **EOSC**

---

[91] Yuri Demchenko, Lennart Stoy, Claudia Engelhardt, Vinciane Gaillard (2021). D7.3 FAIR Competence Framework for Higher Education (Data Stewardship Professional Competence Framework), https://doi.org/10.5281/zenodo.4562089

[92] FAIR Cookbook. https://fairplus.github.io/the-fair-cookbook

[93] EC Innovation Radar. https://www.innoradar.eu/innovation/39384

[94] https://www.imi.europa.eu/about-imi/imi-funding-model

[95] terms4FAIRskills. https://terms4fairskills.github.io

Future project, which will set up a "Knowledge Hub" for EOSC, encompassing catalogue, training platform and associated repository services.

Missing elements (carried over from 2020)**:**
- Fellowships and staff exchanges (TFiR Action 11.3)
- Certification and endorsement (TFiR Action 11.4). These elements are probably missing because they target activities that are on a longer timescale than the projects and stakeholders that are not well represented by the projects. Even so, for project partners who are affiliated with universities, like FAIRsFAIR, the university would be a route for progress.

4.6 Pillar 5: Incentives and metrics for FAIR data and services

Pillar 5: Incentives and metrics for FAIR data and services identifies the need to develop metrics to certify FAIR objects and services and their implementation.

Rec. 12: Develop metrics for FAIR Digital Objects

"A set of metrics for FAIR Digital Objects should be developed and implemented, starting from the basic common core of descriptive metadata, PIDs and access. The design of these metrics needs to be guided by research community practices and they should be regularly reviewed and updated." (TFiR, page 68)

Several projects are using, testing and assessing FAIR metrics. **FAIRsFAIR** published version 04 of its Data Object assessment metrics[96], which among other things are inspired by the RDA FAIR Data Maturity Model[97]. In **ExPaNDS/PaNOSC** the draft Recommendations for FAIR Photon and Neutron Data Management use the prioritisation indicators from the RDA FAIR Maturity Model to prioritise metadata to be collected for FAIR across the PaN experimental lifecycle. Recently the **EOSC Association** started a Task Force on 'FAIR metrics and data quality', to oversee the implementation of FAIR data and to make sure these are fit for purpose. Also relatively new is the RDA CURE-FAIR Working Group[98], which aims at devising recommendations for **computational reproducibility of analyses and workflows** in line with the FAIR principles. The RDA/ReSA/FORCE11 FAIR for Research Software (FAIR4RS) Working Group[99] is starting scoping work for a task force to develop metrics for **FAIR software**. Another related initiative is the FAIR Digital Object Forum[100], which is developing a technical specification for FAIR Digital Objects. Here, metrics are still in early phases but will likely be developed to evaluate the readiness of Digital Objects to be machine processed.

---

[96] Anusuriya Devaraju, Robert Huber, Mustapha Mokrane, Patricia Herterich, Linas Cepinskas, Jerry de Vries, … Angus Whyte. (2020). FAIRsFAIR Data Object Assessment Metrics (Version 0.4). http://doi.org/10.5281/zenodo.4081213

[97] RDA FAIR Data Maturity Model Working Group (2020). FAIR Data Maturity Model: specification and guidelines. Research Data Alliance. https://www.doi.org/10.15497/RDA00050

[98] RDA CURE-FAIR WG. https://www.rd-alliance.org/groups/cure-fair-wg

[99] RDA FAIR for Research Software WG. https://www.rd-alliance.org/groups/fair-research-software-fair4rs-wg

[100] FAIR Digital Object Forum. https://fairdo.org

Moving from the metrics to their implementation, many activities were reported. After a pilot with five certified repositories **FAIRsFAIR** launched the F-UJI tool, which automatically assesses the FAIRness of research data in repositories; in addition to automation FAIRsFAIR offers consultancy to F-UJI users about the outcome and to learn about possible improvements. **EOSC-Nordic** continuously evaluates the FAIRness of data in a sample of more than 100 repositories. First they used the Wilkinson tool[101] but then they moved to F-UJI, because that addresses more aspects. **DKRZ** has made FAIR maturity assessments of their data holdings using the F-UJI tool. A set of metrics to evaluate climate science related data has been used to assess FAIRness in recent work. **ARCHIVER** is integrating F-UJI in the testing pipeline in order to monitor the compliance of repositories and the uptake of the FAIR principles. They help to adapt the tool to specifics of research communities and also to improve the tool.

ENVRI-FAIR has performed a FAIR assessment exercise with the help of methodologies generated in GO FAIR for the participating Research Infrastructures. The result of the assessment is currently being implemented in a common system. ENVRI-FAIR looked at what metadata is available at each stage of the (idealised) experimental lifecycle; for example, the PI name and identifier become available at the first stage of the life cycle, i.e. the proposal stage. For each metadata type, they then used the RDA FAIR Data Maturity Model prioritisation indicators to prioritise that metadata type (i.e. as essential, important, useful, not relevant) for each element of FAIR (i.e. F, A, I, R). **FAIRplus**[102] has FAIRified a variety of life science examples, using the RDA/FAIRsFAIR indicators and testing the various evaluation tools. FAIRplus noticed that the tools and the indicators are too generic for their datasets and use cases, and has adapted them.

Missing element (carried over from the 2020 workshop):
- Developing a governance process for the maintenance and revision of metrics and associated assessment processes is important.

Rec. 13: Develop metrics to certify FAIR services

"Certification schemes are needed to assess all components of the ecosystem as FAIR services. Existing frameworks like CoreTrustSeal (CTS) for repository certification should be used and adapted rather than initiating new schemes based solely on FAIR, which is articulated for data rather than services." (TFiR, page 68)

Whereas last year's workshop found that the majority of the work was done around repositories, we now see a little more activity on services in a broader sense. **ENVRI-FAIR**'s work package "Common requirements and testbed for (meta)data services, community standards and cataloguing" is working on this together with the subdomains in ENVRI-FAIR. **FAIRsFAIR** published a milestone report on the

---

[101] FAIR Evaluator tool. https://w3id.org/AmIFAIR
[102] FAIRplus project. https://fairplus-project.eu

FAIRness of services[103]. **ELIXIR** has drawn up their formal criteria for selecting Recommended Interoperability Resources, which are resources and services that facilitate FAIR-supporting activities and are themselves FAIR; one of them is FAIRsharing. **EOSC Nordic** developed a maturity framework for services, but intentionally did not address FAIR in this because of work by other initiatives. FAIR aspects can be added later to the framework.

NI4OS and **SSHOC** apply metrics for certifying repositories. SSHOC work will include feedback to the CoreTrustSeal Board on the CTS requirements. **ExPaNDS** and **PaNOSC** are working on quality assurance for their data repositories and an open self-certification assessment. They will be exploring synergies with the F-UJI tool and CTS and other initiatives (e.g. related to work done in PaNOSC). **EOSC Synergy** is developing an automated FAIR Evaluator tool for multidisciplinary repositories. The tool will be integrated into EOSC Synergy Software and Services Quality as a Service process.

Missing elements:
- An unambiguous definition of "service" and possibly a typology of services. There is the risk of misusing or misinterpreting metrics when the object is unclear, in addition to the trend that "FAIR" is being used for resources other than Digital Objects.
- Developing a governance process for the maintenance and revision of metrics and associated assessment processes is important. (Carried over from the 2020 workshop)

Rec. 25: Implement and monitor metrics

"*Agreed sets of metrics should be implemented and monitored to track changes in the FAIRness of data sets or data-related resources over time. Funders should report annually on the outcomes of their investments in FAIR and track how the landscape matures.*" *(TFiR, page 74)*

Last year's workshop[104] suggested to "organise assessment and comparisons of metrics tools: are the outcomes of competing tools comparable, when applied to the same data set or repository? Do users - potential FAIRness auditors included - apply them in the same way?". While there are no definite answers yet, let alone consensus on metrics, tools and badges, some projects explore and apply more than one approach, see Recommendation 12. In the 'InfraEOSC-5 FAIR data and infrastructure Task Force' **FAIRsFAIR**, **EOSC Secretariat**, **EOSC-Nordic**, **EOSC Synergy**, **ExPaNDS**, **EOSC Pillar**, and **NI4OS Europe** organised a joint workshop for the European Commission about their FAIR assessment and certification activities**.**

EOSC-Nordic is **monitoring the evolution of the FAIRsFAIR metrics** among 100+ repositories in their sample and will monitor the progress of these over a time period of 2 years. They noticed that the repositories recently have not increased their FAIRness level that much. Therefore EOSC-Nordic introduces a number of new activities, for instance informing communities about the recent score

---

for their repositories, summarising recommended actions and issues, and offering 1-on-1 meetings to communities[105].

ARCHIVER helps to **improve the F-UJI tool** by adapting it to use cases such as the CERN Open Data Portal, taking in consideration for example custom metadata from physics datasets at a certain scale (8k records per dataset) and by apply F-UJI scores to TDR assessments. It also investigates the use of alternative input sources[106] for F-UJI to start the analysis from, instead of starting from repository landing pages.

For qualitatively assessing the FAIR maturity level of research infrastructures **ENVRI-FAIR** has used the FAIR Implementation Profile (FIP) approach[107]. The FIP concept, initiated by **GO FAIR** and co-developed with ENVRI-FAIR, makes explicit the FAIR-enabling resources deployed by communities and can be used to monitor over time the continuing evolution of the FAIR implementation landscape. **ExPaNDS** will publish the final extended data policy framework for Photon and Neutron RIs in August 2021, part of which concerns the monitoring of FAIR, including regular audits on the degree of compliance to FAIR data. Currently, monitoring of data compliance is rarely in place. Around the same time **FAIRsFAIR** will publish its data badging schema. **DRKZ** periodically assesses the FAIRness of its certified long-term archive. FAIR is not explicit in their submission guidelines, which however ensure a high degree of long-term reusability. Comparing the assessment scores over time is not planned yet; DRKZ reasons that this might be misleading because assessment tools are still under development.

Possible follow-up action:

- The new Task Force 'FAIR metrics and data quality' under the EOSC Association oversees the implementation of FAIR metrics for the European Open Science Cloud (EOSC), testing them with research communities to ensure they are fit for purpose. This should foster the"agreed set of metrics".

Missing element:

- Concrete reporting activity by funders on the outcomes of their investments in FAIR cannot be identified in the workshop input. This probably stems from their underrepresentation in the workshop, but on a more principled level also from the lack of the "agreed set of metrics".

Rec. 26: Support data citation and next generation metrics

"Systems providing citation, reuse and impact metrics for FAIR Digital Objects and other research outputs should be provided. In parallel, next generation metrics that reinforce and enrich

---

citation centric metrics for evaluation should be developed." (TFiR, page 74)

SSHOC raises the question how to credit intermediaries such as the Europeana[108] platform, where reusers have found the data for reuse. A SSHOC deliverable about 'making data findable by being citeable' is in the making. **ExPaNDS**' workshop series to promote FAIR principles looked at the question of how data citation can be encouraged and supported and will discuss with librarians and data managers (September 2021) how to better bring together publications and data, including through citation. **EOSC Nordic** will organise a FAIRification event after the summer on rich community-specific metadata using F-UJI, as an example of 'next gen' metrics.

It is mentioned that citation ties in with the **licensing of data**, when licences are used to try and force data citation. However, citation is an 'academic norm' issue while licensing is a copyright/legal issue. In terms of data, **ExPaNDS** has noticed a tendency for organisations to assume that what works for publications (e.g. a CC-BY licence) also works for data. However, certainly in the photon and neutron community, much of the data is measurement data so it is not even clear that copyright applies to those data. Therefore, a CC-BY licence is not appropriate, as this assumes copyright applies.

JATS for Reuse (JATS4R) is a working group devoted to optimising the reusability of scholarly content. The NISO JATS4R WG[109] has set up a **software citation** subgroup to ensure that metadata to support software citation is accurately passed through the scholarly publishing infrastructure ecosystem. The WG, with representatives from the FAIR4RS WG[110] and FORCE11 Software Citation Implemention WG[111] builds on the guidance published earlier this year on software citation[112].

In the certified WDCC repository, hosted by **DRKZ,** citation information is given for every dataset curated in WDCC. DKRZ is a signatory of the COPDESS[113] (Coalition for Publishing Data in the Earth and Space Sciences) declaration; COPDESS provides guidance on sharing and citing of research outputs.

Missing element:

- Activities regarding systems that provide reuse and impact metrics, for the full range of valuable research outputs and FAIR Digital Objects, are hard to identify (Action 26.1).

4.7 Pillar 6: Investment in FAIR

The recommendations under Pillar 6: Investment in FAIR, primarily concern funders as the key actors in relation to investment. Nevertheless, other organisations have a role in contributing to the

---

[108] Europeana. https://www.europeana.eu/
[109] https://jats4r.org/current-subgroups/ NISO = National Information Standards Organization, https://www.niso.org/ JATS = Journal Article Tag Suite, http://jats.niso.org/.
[110] FAIR4RS WG. https://rd-alliance.org/groups/fair-research-software-fair4rs-wg
[111] FORCE11 Software Citation. https://www.force11.org/group/software-citation-implementation-working-group
[112] https://f1000research.com/articles/9-1257
[113] COPDESS: Coalition for Publishing Data in the Earth and Space Sciences. https://copdess.org/

coordinated and strategic approach to sustain a FAIR ecosystem. In relation to this pillar, the questions for the projects consulted in the FAIRsFAIR Synchronisation Force workshops were:

● How are they contributing to a view of coordination and strategic funding?
● What are they saying that contributes to a view of the strategically important parts of the FAIR ecosystem and how funding can be coordinated to meet these objectives?

Rec. 14: Provide strategic and coordinated funding

"Funders should adopt a coordinated approach to supporting core infrastructure and services, building on existing investments where appropriate. Funding should be tied to certification schemes, sustainable business models and other community-vetted indicators that demonstrate viability." *(TFiR, page 69)*

This recommendation is mainly directed at funders but there are implications for other stakeholders. The **FAIRsFAIR** project contribution towards this recommendation mainly focuses on providing insights, information and evidence which will inform strategic and coordinated funding. For example, a coordination plan for a sustainable network of FAIR Trusted Digital Repositories (deliverable due Sept 2021) will provide evidence in relation to the funding, income streams and business models of repositories that contribute to their sustainability. Similarly, FAIRsFAIR is providing tools and instruments that will inform funders in developing policies through policy enhancement recommendations[114] that will assist funders in investment decisions to improve FAIR data through 'Changes in Data Policy and Practice - an updated analysis' (D3.8 due March 2022). Similarly, **NI4OS-Europe** will contribute to this recommendation through reports addressing business models and other issues relating to sustainability. A Business Model for Recommendations will be delivered by the project in October 2021. For the project itself and related services, a Sustainability Report will be delivered in August 2022. **EOSC-Nordic** is supporting communities with certification and FAIR support and guidance, which will result in better practices and enable the research community and repositories to fulfil and align with funders' requirements. Among the objectives is to help repositories and other data services to become more sustainable by improving their practices.

ExPaNDs' closeness to the facilities provide an indirect opportunity to influence issues of sustainability. In particular work in **ExPanDs** and **PANOSC** to chart the costs for making data FAIR at their facilities will be significant. For the ExPanDs project itself, a sustainability plan[115] exists for each work package focussing on which existing standards (e.g. OAI-PMH, Nexus format), initiatives and services should be built upon. The project's data management plan[116] primarily concerns the projects' own timeframe but references the allocation of resources and longer term access to outputs. To contribute to issues around strategic and sustainable funding, **ENVRI-FAIR** aims to

---

influence the services being developed through the clusters. This will form part of the project sustainability plan. The ENVRI strategy is being updated and investigations are underway about working towards a more permanent structure (e.g. through mutual MoUs or an association).

GO FAIR Foundation runs a Programme on Certification[117], in which they advocate to dedicate five percent of project funding to FAIR data stewardship: sustainability, vocabularies, resources etc. The initiative also explored a FAIR Funder pilot programme to make it easy for funders to require and for grantees to produce FAIR Data[118].

Strategic and coordinated funding can be provided at the national level. For example, since 2018 **Sweden** has been addressing sustainable research data management and making research data outputs FAIR at the national level with universities cooperating as a consortium through the Swedish National Data Service (SND).[119] The SND works with a common data repository and catalogue[120] with competence centre activities for researchers and research support staff. **France** is also a good practice with their national Plan on Open Science[121].

◆ Rec. 15: Provide sustainable funding

"Funders who issue requirements on FAIR must provide support to ensure the components of the FAIR ecosystem are maintained at a professional service level with sustainable funding. Service providers should explore multiple business models and diverse income streams." (TFiR, page 69)

The repository and service certifications in **FAIRsFAIR** (WP4, WP2) are relevant to this recommendation as they address business models and long-term sustainability of activities and contain outputs which can inform funders seeking to sustain a FAIR data ecosystem. The FAIRsFAIR sustainability plan (D1.6 Feb 2022) includes an examination of post project continuation and a possible network of Trustworthy Digital Repositories which aims to ensure sustainability. Relevant future deliverables include a 'Framework for Assessing FAIR Services' (D2.7 due Aug 2021), a report on the testbed of FAIR metrics and data assessment badging scheme (D4.5 due Aug 2021), and a report on a maturity model towards FAIR data in FAIR repositories (D4.6 due Feb 2022).

EOSC Nordic provides support for FAIRness and certification of repositories that can enhance the FAIR data ecosystem and its sustainability. Reports that will inform funders include 'An Assessment of FAIR-uptake among Regional Digital Repositories'[122] and a Report on Nordic repositories and their uptake of FAIR (February 2022). **EOSC-Pillar** looked at sustainable business models and recommendations, i) through a survey run in 2020 including funding issues and covering institutions, universities, research infrastructures, funding bodies and others and ii) through interviews with a

---

[117] GO FAIR Foundation pioneer programme. https://www.gofairfoundation.org/certification/
[118] GO FAIR funder pilot programme. https://www.go-fair.org/today/fair-funder/
[119] Swedish National Data Service https://snd.gu.se/
[120] Swedish National Data Service https://snd.gu.se/en/catalogue
[121] https://www.ouvrirlascience.fr/open-science/
[122] Andreas Ortmann Jaunsen, Mari Kleemola, Tuomas J. Alaterä, Heikki Lehvaslaiho, … Pauli Assinen (2020). D4.1 An assessment of FAIR-uptake among regional digital repositories https://doi.org/10.5281/zenodo.4045402

dozen service providers including project partners and other scientific use cases (WP6/WP7), run in December 2020. This analysis will inform a business model and sustainability study (D4.5 expected in June 2022) which will report findings and make recommendations with a particular focus on sustainability. **NI4OS-Europe** will provide information on possible business models (October 2021). A Cost-Benefit Analysis is being undertaken at the national rather than the project level with some partners working with their ministries in the context of national open science clouds. Work on 'Innovation Management and Sustainability' produced a 'Marketing, Dissemination and Sustainability Plan'[123] and will produce a Sustainability Report in September 2022.

ExPaNDs is working on a framework for facilities' data policies including the development and application of Data Management Plans (DMP) and locally based data stewards. Once FAIR is integrated into the policies, these issues become a focus for funding. Completed outcomes of the project include a 'Sustainability Policy Report'[124] detailing the strategy of making ExPaNDS outputs indispensable to the facilities (and their sustainable funding) and a draft extended data policy framework for Photon and Neutron RIs[125] that enables indirect strategic funding for FAIR to be included in the facilities data policy. In the subdomains covered by the project (atmosphere, marine, solid earth, ecosystems) **ENVRI-FAIR** is working on sustainable data management for the contribution to the EOSC through i) an atmospheric subdomain development strategy, (D8.12 August 2022); ii) a marine subdomain white paper for sustainable data management, (D9.10 December 2022); iii) a Final report on EPOS policies and governance for FAIR and EOSC, (D10.7 December 2022); and iv) a biodiversity and ecosystem subdomain long term development and management plan, (D11.4 October 2022). These outputs will inform sustainability for the RI and Funder policy. In **PaNOSC**, a future piece of work will aggregate the costs of data management at 6 ESFRI facilities to be included in business models and see what is feasible (T7.2 refined cost collection, completed in mid-June 2021). This work will highlight costs due to FAIR DM and the additional costs of linking with the EOSC e.g. costs associated with providing access to data and services to anyone, outside the embargo period. **PaNOSC** (Photon and Neutron ESFRI and ERICS) is also working with **ExPaNDS** (national facilities) to examine business models to sustain developments beyond the project. Business models will be validated by project work package leaders and facilities' directors. Possible governance structures are being studied (T7.4) where the sustainability plan will cover governance, funding streams and the long-term operation of the PaN (Photon and Neutron) EOSC. In this work, the assessment of the maintenance needs for the current federated services has been completed; work to study the feasibility of different funding streams, including in-kind contributions is ongoing; then a long-term strategy for the PaN facilities will be developed. For **SSHOC**, work on the governance and sustainability of the SSH part of the EOSC (T8.1) will result in a Roadmap in December 2021 (D81).

---

[123] Aneta Karaivanova (2019). NI4OS-Europe Marketing, dissemination and sustainability plan https://zenodo.org/record/3736165#.XrR7bKgzY2x

[124] Mirjam van Daalen, Valentina Piffer, Patrick Fuhrmann, Sophie Servan (2020). D1.7 ExPaNDS Sustainability policy report. https://doi.org/10.5281/zenodo.3715279

[125] Brian Matthews, Abigail McBirnie, Andrei Vukolov, … Mirjam Van Daalen. (2020). Draft extended data policy framework for Photon and Neutron RIs. https://doi.org/10.5281/zenodo.4014811

Rec. 27: Open EOSC to all providers but ensure services are FAIR

"The Rules of Participation for EOSC must be based on the diverse mix of infrastructure and tools currently in use to enable service providers from all sectors to be part of EOSC. The Rules should ensure that services are FAIR-compliant and use open APIs and interchange standards." (TFiR, page 75)

The projects surveyed do not set the Rules of Participation, but they can influence policies through studies, reports and recommendations and can offer new FAIR-compliant services for EOSC.

EOSC Pillar made a two-month open call in January 2021 (WP6/T6.10) with the objective of assisting new services to initially join the EOSC Pillar portfolio and ultimately the EOSC catalogue. The project will invest in eight selected thematic service providers[126] in compliance with the EOSC Rules of Participation. An internal hackathon which included the level of FAIRness required for onboarding, was held in 2021. Outcomes should be published in Autumn 2021.

ExPaNDs' future plans include a report on 'Deployment of Analysis Services for EOSC' (December 2022). At the moment the **ENVRI-FAIR** coordination is organising a workshop with other cluster projects to look into this issue. The implications of this recommendation will also be included in discussions of cluster roles in the EOSC at a workshop foreseen in 2022 with the Commission, EOSC Association and cluster projects.

A sustainable model to procure services has been proposed by **ARCHIVER** to the EOSC Future project to address concerns raised by the EC. These will deliver sustainable, production quality long-term data preservation services for user communities that fill a gap in the existing EOSC portfolio. The scope will be extended, in the EOSC Future project, to accommodate diverse use-cases, including small-scale deployments requiring terabyte-range preservation capacity.

4.8 Additional scoring considerations

This section reflects part of the concluding discussion with the workshop participants and provides points to consider when one wants to score or even to measure "activities".

Informative as the six pillar sessions are, a caveat is in order with respect to scoring the findings and interpreting the scores. First, we collected and scored *activity levels* in the area of the TFiR recommendations, not aspects like progress, implementation, or quality of any kind of result. In other words, we don't try to answer the questions "How FAIR are we? or, how far are we?". An answer to this goes beyond the scope and the means of the Synchronisation Force, not least because it requires agreement on how to define and measure such aspects. Weighing diverse activities ranging from reports to workflows or certification support to derive at a simple score comes before that and is challenging enough.

---

126 https://www.eosc-pillar.eu/news/six-applicants-awarded-eosc-pillar-open-call-thematic-service-providers

Second, we asked workshop participants about *their* activities, and although many were generous with information beyond their own project, our overview clearly doesn't cover all EOSC-related initiatives. Third, "gaps" in the spreadsheet don't imply an omission on the side of the projects, as it is not to be expected that each project or initiative, let alone a FAIR Champion, would be addressing all TFiR recommendations. In this context it was also remarked, for example, that in particular activities regarding incentives and metrics for FAIR data and services (pillar 5) would be hard and only possible once other aspects are in place. Fourth, a couple of TFiR recommendations target other stakeholders than those invited to attend the workshop, for example research funders. We can't give such recommendations a high activity score.

Finally, it should be recognised that this report can only provide a snapshot in time, although we try to relate the new information to the findings from last year's workshop. This in itself presented us with an interesting challenge. While analysing the information the question arose whether we could score the activity level *lower* than we did in the second workshop. On the one hand, adding up the activities – reported in both 2020 and 2021 – suggests at least the same as and probably a higher level and score than in 2020. On the other hand, an extra year has passed since the TFiR publication, in which one would hope for an overall increase of FAIR-related activities, which so to speak raises the bar. We decided on the latter, more conservative approach. Participants in the concluding workshop session agreed with this approach, with 19 out of 20 votes, without yet knowing the differences between 2020 and the draft 2021 scores. (Actually, the draft 2021 scores were lower.) Next, we asked the participants for each pillar whether the draft scoring was accurate, too negative, or too positive. Although for each pillar the majority or even large majority found it accurate, there was also a tendency for a significant minority to find the scores "too positive". This made us review the draft scores, leading to somewhat lower scores in the current report. To summarise: the activity level scores in this report are lower than the draft scores, which were already more conservative than the 2020 scores.

In light of these caveats, the observations and analysis presented in the preceding sections should be considered impressionistic: they express an *interpretation* of the situation and do not pretend to offer a comprehensive or quantified list of all work being undertaken.

Annex 1 - Stakeholders invited to the second workshop

The stakeholders marked in Figure 5 were invited to participate in the second Synchronisation Force workshop.

Figure 5 FAIRsFAIR primary stakeholders within the EOSC ecosystem; invitees highlighted

INFRAEOSC-5 projects: alongside the EOSCSecretariat[127] and FAIRsFAIR, five regional nodes or thematic projects receive funding in the Horizon INFRAEOSC-5 call. These are EOSC Nordic[128] (Nordic and Baltic countries), EOSC Pillar[129] (Open Science across Austria, Belgium, France, Germany and

---

[127] https://www.eoscsecretariat.eu/

[128] https://www.eosc-nordic.eu/

[129] https://www.eosc-pillar.eu/

Italy), EOSC Synergy[130] (national publicly-funded digital infrastructures), ExPands[131] (EOSC Photon and Neutron Data Service), and NI4OS-Europe[132] (National Initiatives 4 Open Science in Europe). All INFRAEOSC-5 projects collaborate in topical Task Forces. For this workshop the members of the Task Force FAIR data and infrastructure have been invited.

ESFRI Cluster projects: five cluster projects were launched in early 2019, namely ESCAPE[133], ENVRI-FAIR[134], SSHOC[135], PaNOSC[136] and EOSC-Life[137]. These five initiatives gather projects from the European Strategy Forum on Research Infrastructures (ESFRI)[138] and European Landmarks, to cover the scientific domains of Astronomy and particle physics, Environmental research, Social Sciences and Humanities, Photon and Neutron research, and Life sciences, respectively.

"Horizontal" projects: several generic or "horizontal" e-infrastructural projects contribute to building the EOSC. Represented in the workshop were OpenAIRE[139], which provides interoperability services and engages all stakeholders for effective Open Science implementation; ARCHIVER[140], to procure R&D services for archiving and digital preservation; the EUDAT Collaborative Data Infrastructure (EUDAT CDI[141]); and the recently started DICE[142] project (Data Infrastructure Capacity for EOSC).

FAIR Champions: the European Group of FAIR Champions (EGFC)[143] was initiated by FAIRsFAIR and is composed of scientific experts and "doers" in the field of FAIR data, who work as ambassadors.

FAIRsFAIR HLAC: the High Level Advisory Committee (HLAC)[144] is a group of international experts appointed by the Executive Committee of FAIRsFAIR.

FAIR Champions and HLAC members also contributed information about other FAIR initiatives.

EOSC Governance: members of the Board of Directors of the EOSC Association have been invited to the workshop.

In addition to these stakeholders, Project Officers from the European Commission have been invited to the workshop.

---

[130] https://www.eosc-synergy.eu/
[131] https://expands.eu/
[132] https://ni4os.eu/
[133] https://projectescape.eu/
[134] https://envri.eu/home-envri-fair/
[135] https://www.sshopencloud.eu/
[136] https://www.panosc.eu/
[137] https://www.eosc-life.eu/
[138] https://www.esfri.eu/
[139] https://www.openaire.eu/
[140] https://www.archiver-project.eu/
[141] https://www.eudat.eu/
[142] https://www.dice-eosc.eu/
[143] https://www.fairsfair.eu/advisory-board/egfc
[144] https://www.fairsfair.eu/advisory-board/hlac

Annex 2 - List of participants

N.	Name	Surname	Organization Type	Affiliation	Country	Stakeholder group
1	Fernando	Aguilar	Universities and Research Performing Organisations	CSIC	ES	other EOSC horizontal projects
2	Mark	Allen	Research Infrastructures & e-Infrastructures	Strasbourg Astronomical Data Centre (CDS), Observatoire de Strasbourg, UMR 7550	FR	FAIR Champions
3	Gingold	Arnaud	Research Infrastructures & e-Infrastructures	Centre National Recherche Scientifique	FR	EC project officers
4	Ari	Asmi	Universities and Research Performing Organisations	University of Helsinki	FI	ESFRI clusters
5	Andreas	Athenodorou	Universities and Research Performing Organisations	The Cyprus Institute & University of Pisa	CY	other EOSC horizontal projects
6	Karla	Avanço	Specialised Service Providers	OpenEdition	FR	EC project officers
7	Jaana Bäck	Bäck	Universities and Research Performing Organisations	University of Helsinki	FI	FAIRsFAIR High-level Advisory Committee
8	Isabel	Bernal	Universities and Research Performing Organisations	CSIC	ES	FAIR Champions
9	Timea	Biro	Universities and Research Performing Organisations	Digital Repository of Ireland / Royal Irish Academy	IE	other EOSC horizontal projects
10	David	Carr	Research Funding Organisations & National Agencies	Wellcome Trust	UK	FAIRsFAIR High-level Advisory Committee

11	Carlos	Casorran	Research Funding Organisations & National Agencies	European Commission	BE	EC project officers
12	Neil	Chue Hong	Universities and Research Performing Organisations	Software Sustainability Institute / University of Edinburgh	UK	INFRAEOSC-5 FAIR Task Force representatives
13	Patricia	Clarke	Research Funding Organisations & National Agencies	Health Research Board	IE	FAIRsFAIR High-level Advisory Committee
14	Gerard	Coen	Research Infrastructures & e-Infrastructures	Data Archiving & Networked Services	NL	FAIRsFAIR project
15	Aoife	Coffey	Universities and Research Performing Organisations	University College Cork	IE	EOSC Board of Directors
16	Christian	Cuciniello	Research Funding Organisations & National Agencies	European Commission	BE	EC project officers
17	Antica	Culina	Universities and Research Performing Organisations	Netherlands Institute of Ecology	NL	FAIR Champions
18	Joy	Davidson	Universities and Research Performing Organisations	Digital Curation Centre, University of Glasgow	UK	FAIRsFAIR project
19	Ornela	De Giacomo	Research Infrastructures & e-Infrastructures	CERIC-ERIC	IT	ESFRI clusters
20	Claudio	Dema	Research Infrastructures & e-Infrastructures	CNR - IMAA	IT	other EOSC horizontal projects
21	Richard	Dennis	Universities and Research Performing Organisations	Royal Danish Library Copenhagen University Library	DE	INFRAEOSC-5 FAIR Task Force representatives

22	Richard	Dennis	Universities and Research Performing Organisations	Royal Danish Library Copenhagen University Library	DE	FAIRsFAIR project
23	João	Fernandes	Universities and Research Performing Organisations	CERN	CH	ESFRI clusters
24	Sandro	Fiore	Universities and Research Performing Organisations	University of Trento	IT	FAIR Champions
25	Vinciane	Gaillard	Other	European University Association	BE	FAIRsFAIR project
26	Federica	Garbuglia	Other	European University Association	BE	FAIRsFAIR project
27	Francoise	Genova	Universities and Research Performing Organisations	CDS/Observatoire Astronomique de Strasbourg	FR	FAIRsFAIR High-level Advisory Committee
28	Stan	Gielen	Research Funding Organisations & National Agencies	NWO	NL	FAIRsFAIR High-level Advisory Committee
29	Marjan	Grootveld	Specialised Service Providers	DANS	NL	EOSC Board of Directors
30	Natalie	Harrower	Research Infrastructures & e-Infrastructures	Digital Repository of Ireland	IE	other EOSC horizontal projects
31	Maggie	Hellström	Research Infrastructures & e-Infrastructures	ICOS ERIC and Lund University	SE	ESFRI clusters
32	Odile	Hologne	Universities and Research Performing Organisations	INRAE	FR	FAIR Champions
33	Andreas	Jaunsen	Research Infrastructures & e-	Nordforsk	NO	INFRAEOSC-5 FAIR Task Force

			Infrastructures			representatives
34	Maria	Johnsson	Universities and Research Performing Organisations	Lund University	SE	FAIR Champions
35	Bob	Jones	Research Infrastructures & e-Infrastructures	CERN	SW	EOSC Board of Directors
36	Sarah	Jones	Research Infrastructures & e-Infrastructures	GÉANT	NL	other EOSC horizontal projects
37	Nick	Juty	Universities and Research Performing Organisations	The University of Manchester	UK	ESFRI clusters
38	Mari	Kleemola	Research Infrastructures & e-Infrastructures	Finnish Social Science Data Archive	FI	INFRAEOSC-5 FAIR Task Force representatives
39	Oliver	Knodel	Research Infrastructures & e-Infrastructures	Helmholtz-Zentrum Dresden-Rossendorf	DE	other EOSC horizontal projects
40	Iryna Kuchma	Kuchma	Other	EIFL	LT	other EOSC horizontal projects
41	Hervé	L'Hours	Research Infrastructures & e-Infrastructures	UK Data Archive, UK Data Service, University of Essex	UK	FAIRsFAIR project
42	Marialuisa	Lavitrano	Universities and Research Performing Organisations	University Milano Bicocca - EOSC Association	IT	EOSC Board of Directors
43	Barbara	Magagna	Research Infrastructures & e-Infrastructures	Environment Agency Austria	AT	ESFRI clusters
44	Natalia	Manola	Research Infrastructures & e-Infrastructures	OpenAIRE	GR	other EOSC horizontal projects

45	Abigail	McBirnie	Research Funding Organisations & National Agencies	UKRI/STFC	UK	INFRAEOSC-5 FAIR Task Force representatives
46	Elizabeth	Newbold	Research Infrastructures & e-Infrastructures	STFC	UK	FAIRsFAIR project
47	Josefine	Nordling	Research Infrastructures & e-Infrastructures	CSC	FI	FAIRsFAIR project
48	Eoghan	Ó Carragáin	Universities and Research Performing Organisations	University College Cork	IE	EOSC Board of Directors
49	Milan	Ojsteršek	Universities and Research Performing Organisations	University of Maribor	SI	other EOSC horizontal projects
50	Zafer	Ozturk	Universities and Research Performing Organisations	University of Twente	NL	other EOSC horizontal projects
51	Elli	Papadopoulou	Universities and Research Performing Organisations	Athena Research Center	GR	INFRAEOSC-5 FAIR Task Force representatives
52	Jessica	Parland-von Essen	Research Infrastructures & e-Infrastructures	CSC	FI	FAIRsFAIR project
53	Karsten	Peters-von Gehlen	Research Infrastructures & e-Infrastructures	German Climate Computing Center (DKRZ)	DE	FAIR Champions
54	Rūta	Petrauskaitė	Universities and Research Performing Organisations	Vytautas Magnus University	LT	FAIRsFAIR High-level Advisory Committee
55	Sara	Pittonet Gaiarin	Small-Medium Enterprises (SMEs)	Trust-IT Services	IT	FAIRsFAIR project
56	Vanessa	Proudman	Other	SPARC Europe	NL	FAIRsFAIR project

57	Antti	Pursula	Research Infrastructures & e-Infrastructures	EUDAT CDI	FI	other EOSC horizontal projects
58	Andreas	Rauber	Universities and Research Performing Organisations	TU Wien	AT	FAIR Champions
59	kostas	repanas	Policy Making Organisations	European Commission	BE	EC project officers
60	Olivier	Rouchon	Research Infrastructures & e-Infrastructures	CINES	FR	INFRAEOSC-5 FAIR Task Force representatives
61	Bregt	Saenen	Policy Making Organisations	European University Association	BE	FAIRsFAIR project
62	Susanna-Assunta	Sansone	Universities and Research Performing Organisations	University of Oxford	UK	FAIR Champions
63	Katrin	Seemeyer	Universities and Research Performing Organisations	Forschungszentrum Juelich GmbH	DE	ESFRI clusters
64	Sophie	SERVAN	Research Infrastructures & e-Infrastructures	DESY	DE	INFRAEOSC-5 FAIR Task Force representatives
65	Hugh	Shanahan	Universities and Research Performing Organisations	Royal Holloway, University of London	UK	FAIRsFAIR project
66	John	Shepherdson	Research Infrastructures & e-Infrastructures	CESSDA ERIC	UK	ESFRI clusters
67	Carthage	Smith	Policy Making Organisations	OECD	FR	FAIRsFAIR High-level Advisory Committee
68	Miriam	Tormin	Universities and Research Performing Organisations	Research Assistant	DE	FAIRsFAIR project

| 69 | Erzsébet | Tóth-Czifra | Research Infrastructures & e-Infrastructures | DARIAH-EU | DE | FAIR Champions |
| 70 | Angus | Whyte | Universities and Research Performing Organisations | Digital Curation Centre | UK | FAIRsFAIR project |

Annex 3 - Notes from the pillar sessions

In advance of the six pillar sessions the Synchronisation Force published a spreadsheet that lists all TFiR recommendations. Per recommendation - if relevant - the participants provided information about their project's deliverables and activities that address this recommendation: both "in place" and "planned". In the sessions they were invited to flesh out this information. Furthermore, the question was discussed whether any FAIR-related activities were missing from the TFiR recommendations or deserved more attention or priority.

This spreadsheet and the session notes are available from Zenodo[145] as data underlying the current report. As mentioned before, it should be recognised that the data and the report can only be a snapshot in time. It should also be noted that "gaps" in the spreadsheet don't imply an omission on the side of the projects: it was not to be expected that each project or working group has planned deliverables to address each of the TFiR recommendations.

---

[145] The spreadsheet and the session notes are available from https://doi.org/10.5281/zenodo.5336657