



FAIR data principles in practice

To what extent are the FAIRSF AIR “Recommendations on practice to support FAIR data principles” supported by the results of the EOSC-hub project?

*Working paper aimed to share ideas on practices to support FAIR data principles.*¹

Identifier: <http://zenodo.org/record/4638551>

Date: March 2021/ Version 1

Author: René van Horik, DANS

ORCID: <https://orcid.org/0000-0001-6899-760X>

1. Introduction

Since 2016, when the FAIR data principles for scientific data management and stewardship were published², the importance of and attention to the principles has increased. The FAIRSF AIR project formulated ten recommendations on practice to support the FAIR data principles.³ This working paper reports on a consultation (in the form of desk-research) conducted to assess the relevance of these recommendations in relation to services developed by the EOSC-hub project.⁴ This working paper covers the outcomes of the consultation and describes the extent EOSC-hub services can be used to implement the recommendations made by the FAIRSF AIR project. As a result, this working paper can contribute to future activities in support of the FAIR data principles.

Both the FAIRSF AIR project and the EOSC-hub project are aimed at the European research data community, but they differ in terms of target group and focus. The FAIRSF AIR project⁵ aims to supply practical solutions for the use of the FAIR data principles throughout the research data life cycle. The FAIRSF AIR recommendations on practice to support FAIR data principles are aimed at people who support researchers in the production and management of data, code and related research objects. These are data stewards and research software engineers. The EOSC-hub project⁶ brings together multiple service providers to create a single contact point for European researchers and innovators to discover, access, use and reuse a broad spectrum of resources for advanced data-driven research.

¹ This working paper is part of the activities formulated in the MoU between FAIRSF AIR and EOSC-hub projects (July 2020).

² Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

³ The recommendations were published in June 2020 in: “D3.4 FAIRSF AIR Recommendations on practice to support FAIR data principles”. <https://zenodo.org/record/3924131>

⁴ This consultation is based on personal observations of the author and cannot be considered as an official position of the FAIRSF AIR and EOSC-hub project.

⁵ See: <https://fairsfair.eu/>

⁶ See: <https://www.eosc-hub.eu/>

EOSC-hub started in January 2018 and ended in April 2021. FAIRsFAIR started March 2019 and will run until April 2022.

The remainder of this report consists of four sections. First, an overview is given of the ten recommendations for practice in support of FAIR data, as formulated by the FAIRsFAIR project. They are the starting point of the consultation. Next, the main results of the EOSC-hub project are presented. Then, based on the output of the EOSC-hub project, the recommendations are discussed. The point here is whether EOSC-hub provides suggestions for the follow-up and implementation of the recommendations formulated by the FAIRsFAIR project. The fourth part of this report consists of concluding remarks and suggestions for further activities.

2. Overview of Recommendations on practice to support FAIR data

The ten FAIRsFAIR recommendations that aim to take practical actions to support the realisation of a FAIR ecosystem are stated in the table below. The recommendations are divided into four themes.⁷ See table 1.

Theme A: Develop and implement data sharing and interoperability frameworks	
	A1: Describe research outputs using agreed terminologies and metadata standards to make data FAIR
	A2: Build a culture of data citation
Theme B: Ensure data management is supported by data management plans (DMPs)	
	B1: Formalise and support appropriate data management plans (DMPs) for FAIR data
	B2: Develop roadmaps, guidance and workflows for machine-actionable data management plans (DMP) to inform FAIR data stewardship
Theme C: Develop professional support for FAIR data	
	C1: Define and manage FAIR support costs and resources
	C2: Develop and implement models for coordinating and supporting data stewards and research software engineers
	C3: Develop and implement terminology for competence centres to annotate and retrieve training materials on enabling FAIR
	C4: Develop and implement a self-assessment framework for Research Infrastructures, institutions, and other FAIR competence centres
Theme D: Ensure trusted curation of data	
	D1: Develop and implement guidance and support for selection of appropriate trusted digital repositories (TDRs)
	D2: Develop and implement guidance and support for making sensitive data FAIR for reuse

Table 1. Recommendations on practice to support FAIR data principles

⁷ "D3.4 Recommendations on practice to support FAIR data principles" (see footnote 3) describes the recommendations in more detail.

3. Main results of the EOSC-hub project

The most important results of the EOSC-hub project are the Key Exploitable Results (KERs). They are used in this report as a reference for assessing the relevance and applicability of the recommendations supporting the FAIR data principles based on the output of the EOSC-hub project.⁸ Table 2 provides an overview of the nine KERs that have been produced by the EOSC-hub project.

#	Key Exploitable Result of the EOSC-hub project
1	<u>EOSC Portal and Marketplace.</u> (Supports the discovery of and access to services in EOSC.)
2	<u>EOSC Service Management System - SMS.</u> (Defines and implements the EOSC IT service management system (ITSM), i.e. the activities performed by service providers to plan, deliver, operate and control services offered to customers.)
3	<u>EOSC Rules of Participation - ROP</u> (Comprehensive and coherent set of rules and policies for service providers to onboard services and make them discoverable and accessible through the EOSC Portal.)
4	<u>Internal Services in the Hub Portfolio.</u> (Provides basic enabling services for EOSC access and operation, such as access control or accounting, and offer common and standard interfaces to shared tools.)
5	<u>External Services in the EOSC Service Portfolio.</u> (Provides a "one-stop-shop" for a range of services and solutions to speed up the research process of the disciplines and enable cross-disciplinary collaboration and reuse of tools and results.)
6	<u>EOSC Digital Innovation Hub (INI).</u> (Provides a clear interface for commercial innovation that can be supported by EOSC as part of the broader European Digital Innovation Hub landscape.)
7	<u>Business and sustainability models for services and the hub.</u> (In addition to grounding the discussions about finances, they provide foundations for ensuring the trust of users and user communities in the continued delivery of services.)
8	<u>Interoperability and integration guidelines.</u> (Defines the high-level architecture for basic EOSC technical functions and promotes EOSC standards and APIs, that will facilitate access to services.)
9	<u>Training courses and material.</u> (Encompasses training and support for a large variety of project results.)

Table 2. Key Exploitable Results of the EOSC-hub project

In several cases, such as the EOSC Portal and Marketplace (KER1), results of the EOSC-hub project are developed in collaboration with other initiatives. This makes it difficult to distinguish between what is a specific result of the EOSC-hub project and what is delivered by other related initiatives. External Services in the EOSC Service Portfolio (KER 5) are integrated in the EOSC Portal⁹ and it is not

⁸ A more detailed description of the KERs of the EOSC-hub project can be found at: <https://www.eosc-hub.eu/eosc-hub-key-exploitable-results/>

⁹ Link to the EOSC-portal: <https://eosc-portal.eu/services-resources>

always clear which service can be attributed to which project.¹⁰ This will be taken into account when discussing the relevance of the services to the FAIR data principles in the next section.

4. The FAIR data principles and the EOSC-hub services

The thematic classification for recommended actions to enable FAIR data, formulated by the FAIRsFAIR project, is the starting point for the assessment of the KERs of the EOSC-hub project with regard to their support for the FAIR data principles. Table 3 summarises this assessment.

Theme A: Develop and implement data sharing and interoperability frameworks

This theme concerns the collaboration on research community-specific and inter-domain agreements and uptake of standards, shared vocabularies, ontologies and metadata schemes.

A relevant KER of EOSC-hub with respect to this theme are the “Interoperability and Integration Guidelines” (KER 8). These guidelines “defining the high-level architecture for basic EOSC technical functions and promoting EOSC standards and APIs, will facilitate access to services, lower barriers to integrating and composing services and promote the usage of services between adjacent communities”.¹¹ In the EOSC-hub project interoperability is primarily focused on interoperability of services and to a lesser extent on interoperability of data. A “common vocabulary” is used in relation to the development of a technical reference architecture for services and this can be considered as part of an interoperability framework.

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

Training and support for the creation of data management plans was a specific topic of the training program of the EOSC-hub project that is part of KER 9, “Training courses and material”. At EOSC-hub project plenaries, for instance, all participating research infrastructures and service providers were offered to participate in workshops that provided background information on the relevance of data management planning as well as tools and services to create DMPs. The training material is available for reuse.¹² Also support and consultancy was offered to research infrastructures that are part of the EOSC-hub community. EOSC-hub actively participated in cross-project initiatives to align support for data management.¹³

Moreover the EOSC-hub project compiled a data management plan for the project itself. The EOSC-hub DMP briefly covers data handling of 11 research infrastructures and competence centers that are part of the EOSC-hub project.¹⁴ For each infrastructure, the DMP describes the type of data and its origin, the related metadata standards, the approach to sharing and target groups, and the approach to archiving and preservation. The EOSC-hub DMP does not refer to the FAIR data principles.

Theme C: Develop professional support for FAIR data

This theme is related to the emerging data professional roles such as the data steward and research software engineer.

It is difficult to distinguish an EOSC-hub KER that is directly and clearly related to this recommendation. This is mainly due to the fact that the EOSC-hub project is primarily focused on service development rather than FAIR data. Support is mentioned as part of KER1, “EOSC Portal and Marketplace”. With respect to KER5 “External Services in the EOSC Service Portfolio”, the

¹⁰ The “EOSC Portal Catalogue and Marketplace” contains only 2 resources provided by EOSC-hub as “Related Infrastructure and Platform” (situation in April 2021). These resources are “Chipster” and “Elastic Cloud Compute Cluster”. (see: https://marketplace.eosc-portal.eu/services?related_platforms=26)

¹¹ EOSC-hub Key Exploitable Results, page 11.

https://www.eosc-hub.eu/sites/default/files/EOSC-hub_BookletB5_february2021%20rev4_0.pdf

¹² See, for instance, a template for a “Workshop on Domain Specific Data Management Plans” (March 2021) <https://doi.org/10.5281/zenodo.4588456>.

¹³ See for instance the Community of Practice to share training experience. Training on data management is a prominent topic in this community. See: <https://www.openaire.eu/cop-training>

¹⁴ The EOSC-DMP can be found at: <https://documents.eji.eu/document/3497>

services are “onboarded and also often integrated with each other”¹⁵, which means that it is not possible to allocate a service as a dedicated result of the EOSC-hub project. On the other hand the EOSC-hub project does acknowledge the importance of the FAIR data principles by stating that the service portfolio contains “off the shelf FAIR data and compute services with service integration.”¹⁶

Theme D: Ensure trusted curation of data

Research communities are encouraged to explore the use of trusted digital repositories and to find out how they can help research by providing suitable, secure places of deposit for all types of research data including sensitive data.

The “Business and sustainability models for services and the hub” (KER7) seem most applicable to this theme. However, this KER is aimed at the continued delivery of services (rather than data) and provides a definition for the planned “EOSC Federation Core”.¹⁷ This KER contributes to the work of EOSC policy bodies, in particular to the EOSC Sustainability Working Group and EOSC Architecture Working Group.

For each of the 4 themes FAIRsFAIR has formulated concrete recommendations, ten in total see table 1). In table 3 below, each concrete recommendation is linked to an example related to the EOSC-hub project. For this we do not look at the full catalogue of the EOSC portal, co-developed by EOSC-hub, but at services that are developed by means of resources of the EOSC-hub project.¹⁸ For instance the EUDAT service catalogue¹⁹ that forms a prominent service provider in the EOSC-hub project.

<i>FAIRsFAIR Recommendations on practice to support FAIR data principles</i>	<i>Examples of EOSC-hub output that follow the FAIRsFAIR Recommendations</i>
Theme A: Develop and implement data sharing and interoperability frameworks	KER8 “Interoperability and integrating guidelines”
A1: Describe research outputs using agreed terminologies and metadata standards to make data FAIR	B2FIND discovery service supports metadata interoperability between different domain specific metadata catalogues. ²⁰
A2: Build a culture of data citation	B2HANDLE PID service based on the Handle System. ²¹ Handles can be used to cite data.
Theme B: Ensure data management is supported by data management plans (DMPs)	KER9 “Training courses and material”
B1: Formalise and support appropriate data management plans (DMPs) for FAIR data	Workshop Template to introduce and implement Domain Specific Data Management Plans. ²²

¹⁵ See: <https://www.eosc-hub.eu/sites/default/files/EOSC-hub%20KERs.pdf>, slide 23.

¹⁶ Idem

¹⁷ The Federating Core includes the technical, human, policy and resource-related elements which are needed to allow research-targeted services to operate. For an extended description of the Federating Core, see: <https://www.eosc-hub.eu/news/federating-core-governance-and-sustainability>

¹⁸ The EOSC-hub website states: “The project mobilises providers from the EGI Federation, EUDAT CDI, INDIGO-DataCloud and other major European research infrastructures to deliver a common catalogue of research data, services and software for research.” See: <https://www.eosc-hub.eu/about-us>

¹⁹ See: <https://www.eudat.eu/catalogue>

²⁰ <https://doi.org/10.5281/zenodo.4277666>

²¹ <https://eudat.eu/services/userdoc/b2handle>

²² <https://doi.org/10.5281/zenodo.4588456>

	B2: Develop roadmaps, guidance and workflows for machine-actionable data management plans (DMP) to inform FAIR data stewardship	Machine-actionable DMPs are part of some DMP training and workshops provided by the EOSC-hub project.
Theme C: Develop professional support for FAIR data		KER1 “EOSC Portal and Marketplace” KER5 “External Services in the EOSC Service Portfolio”
	C1: Define and manage FAIR support costs and resources	Not applicable.
	C2: Develop and implement models for coordinating and supporting data stewards and research software engineers	To a limited extent the EOSC-hub competence centers support data stewards and software engineers in using services. ²³
	C3: Develop and implement terminology for competence centres to annotate and retrieve training materials on enabling FAIR	Not applicable. ²⁴
	C4: Develop and implement a self-assessment framework for Research Infrastructures, institutions, and other FAIR competence centres	In the EOSC-hub project the focus was on management of services (rather than on management of data) and a (self) assessment framework for the management of IT services was prominent in the EOSC-hub project: The FitSM standard. ²⁵
Theme D: Ensure trusted curation of data		KER7 “Business and sustainability models for services in the hub”
	D1: Develop and implement guidance and support for selection of appropriate trusted digital repositories (TDRs)	The B2SAFE service is the data repository service further developed in the EOSC-hub project. ²⁶
	D2: Develop and implement guidance and support for making sensitive data FAIR for reuse	Sensitive data services are one of the “Core Service Areas” of EOSC. These services are “Off the shelf sensitive data services” (thus not a direct result of the EOSC-hub project). ²⁷

Table 3. Examples of EOSC-hub output that follow the recommendations on practice to support FAIR data principles

5. Conclusion

This report examines the extent to which the recommendations of the FAIRsFAIR project tie in with the activities of the EOSC-hub project. Both projects have different target groups and overlap two years in time (March 2019-March 2021). We should be aware that we are looking back in time from the position formulated by the FAIRsFAIR project to the realised output of the EOSC-hub project.

²³ For an overview of the services developed during the EOSC-hub project, see: <https://www.eosc-hub.eu/services>

²⁴ Training material developed by the EOSC-hub project is annotated and made accessible via the EOSC-hub training catalogue. The material in the catalogue is aimed at the services developed by the EOSC-hub project in which FAIR is not a prominent topic. See: <https://www.eosc-hub.eu/materials>

²⁵ FitSM - Framework for IT Service Management. See: <https://www.egi.eu/services/fit-sm-training/>. The EOSC-hub project has organised a considerable number of trainings on FitSM.

²⁶ In a few instances a B2SAFE repository is certified by means of the CoreTrustSeal, so it can be considered as an TDR. See for instance: <https://www.coretrustseal.org/wp-content/uploads/2018/03/Meertens-Institute.pdf>

²⁷ See: <https://www.eosc-hub.eu/sites/default/files/EOSC-hub%20KERs.pdf>, slide 24.

The EOSC-hub project cooperates with several other initiatives and projects, so the boundary between the EOSC-hub project and the general EOSC activities was blurred, especially towards the end of the project. As such, this is a good sign, as it indicates that the work of the EOSC-hub project has become part of a more sustainable structure, such as the EOSC Portal and the EOSC Marketplace, in which the FAIR data principles have come to play an important role.

Based on the Key Exploitable Results of the EOSC-hub project, examples could be found that follow the recommendations on the FAIR data principles formulated by the FAIRsFAIR project. The FAIR data principles were not a prominent aspect of the original design principles of the services, but they were found to fit the FAIR data model at the end of the EOSC-hub project.

This report takes the FAIR data recommendations as its starting point and it proved to be an appropriate model for evaluating aspects of the work undertaken as part of the EOSC-hub project.

The FAIRsFAIR project also uses the recommendations for future activities, e.g. to formulate "implementation stories" that put the recommendations into practice and inspire research communities to make data FAIR. The EOSC-hub project and its predecessors have laid a foundation for this.