



FAIR4S

EOSCpilot framework of FAIR data stewardship skills for science and scholarship, and draft recommendations on FAIR training

Key points for consultation

About FAIR4S

The EOSCpilot FAIR4S Framework aims to help you and your organisation to identify the capabilities and skills required to ensure research objects are 'FAIR' (findable, accessible, interoperable and reusable), and that they stay FAIR. In other words the skills to provide stewardship of research outputs. Stewardship skills development is likely to involve a range of 'data experts' - data stewards and other professional groups involved in producing research objects (data managers, data service engineers, and data scientists/analysts).

The European Open Science Cloud is taking shape through EOSCpilot, and its successor projects starting with EOSC-hub, OpenAIRE+, and FREYA. The EOSCpilot Framework offers a template for these projects, their service providers, and the many universities and other Research Producing Organisations (RPOs) that will comprise the EOSC environment. The Framework can be used to identify and describe the competences and learning materials that match the capabilities you need. It offers core competences for data stewardship, relating topics to expertise levels for researchers and the professional groups that support them. The Framework also offers examples of capability and competence statements, focusing these on identified skills gaps.

The Framework has now been published as version 1.0 on the EOSCpilot portal (->) alongside our proposals for Training-as-a-Service in the EOSC. We invite feedback from all those with an interest in skills development for the EOSC, especially training coordinators in Research Infrastructures, e-Infrastructures, and Research Institutions.

We are committed to finalizing the FAIR4S Framework and recommendations about FAIR training in the final months of EOSCpilot to take account of all input received.

Below you will find 5 key points of FAIR4S described further. The 6th point includes our draft recommendations: -

1. Key definitions for the FAIR data stewardship skills context
2. Competences described across relevant professional groups
3. Key dimensions for benchmarking skills development
4. Lifecycle-based approach to describe skills groups
5. Individual-level competences and organisation-level capabilities
6. A path towards FAIR training on FAIR data stewardship

1. Key definitions for the FAIR data stewardship context

FAIR 4S offers the following definitions that apply key terms in skills development to the FAIR data stewardship context.

Stewardship	The formalisation of roles and responsibilities and their application to ensure that research objects are managed for long-term reuse, and in accordance with FAIR data principles.
Capability	<i>Competence</i> applied at a research team or organisational level, with a defined level of expertise and responsibility, to perform a service role or work in the EOSC environment.
Competence	An element (topic) of theory or practice e.g. ‘workflow set-up and management’, combined with an <i>expertise level</i> to indicate whether someone has an awareness of the area, or an ability to do it, or expert knowledge of it.
Skill	<p>A <i>competence</i> or <i>capability</i> acquired or applied in a specific context, e.g. producing a research output or deploying a service. A skill may be specified in a ‘skills user story’.</p> <p>A badge or certificate may provide evidence that a skill has been acquired, and a publication, personal profile, portfolio or CV may provide evidence that a skill has been applied.</p>

EOSCpilot will make recommendations on the wider adoption in EOSC of relevant standards to enable tracking of open science skills from the acquisition of competences through to their application.

2. Competences for relevant professional groups

Institutions and other research performing organisations employ a wide range of professional staff to support researchers. A number of high-level statements on the EOSC have emphasised data stewardship as a key role in enabling the effective use of the EOSC to conduct research ‘as openly as possible, and as closed as necessary’. Data stewards are becoming established as a professional group, and other groups including researchers themselves will need some level of expertise on data stewardship. The EOSCpilot FAIR4S framework views stewardship as a responsibility that’s shared between data stewards and the professional groups shown in Figure 1:

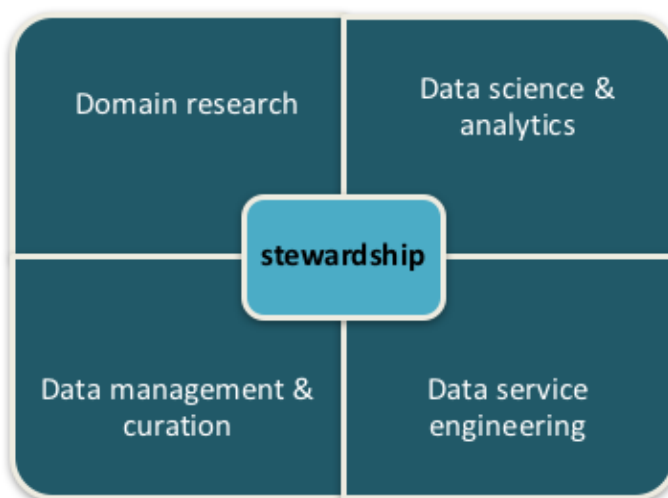


Figure 1. Professional groups sharing stewardship responsibility

FAIR4S draws on existing competence profiles of the groups shown in Figure 1 e.g. those developed in the EDISON project. From jobs currently being advertised in the academic sector it is clear that professional research support roles are evolving quickly. For that reason FAIR4S has avoided describing them at a more granular level, and groups these roles together:

- Data manager, data curator, data librarian
- Data service engineer, data science engineer, research service developer, research software engineer

We recognise that research leads and service managers also have key roles, and associated competences. Rather than include these as a separate professional group, FAIR4S differentiates between levels of *responsibility*, one of the key dimensions to use when applying the framework (see 4 below).

FAIR4S sees the *data steward* role as combining a data management/curation function with at least one other role, i.e. research, data science, or engineering as shown in Figure 1.

3. Key dimensions for benchmarking skills development

When considering how to get the most out of the EOSC environment, individuals and organisations will want to take stock of the skills they already have, and those they need to acquire. FAIR4S bases its 'key dimensions' on existing frameworks for data science, data handling, 'information age skills', and data information literacy. These dimensions are shown below in Figure 3.

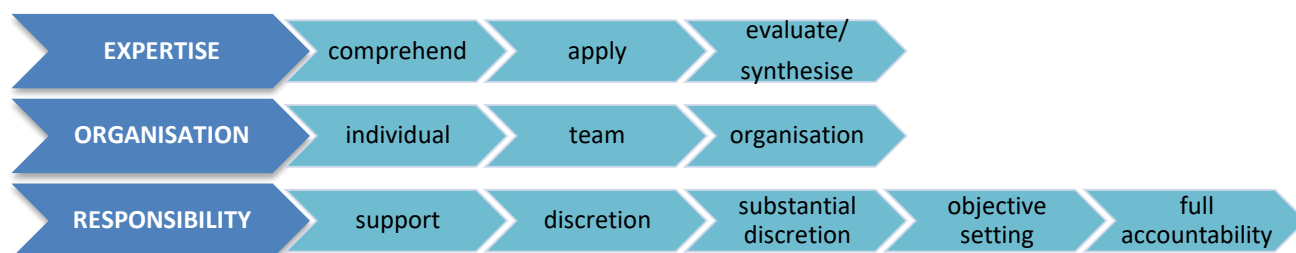


Figure 3. Key dimensions of FAIR4S

Expertise, the first dimension, involves assessing the levels of competence of the individuals involved. The scale begins with 'basic knowledge' i.e. a level of *comprehension* of the topics involved. The next level is to *apply* that understanding in specific contexts, and the third is to *evaluate* applications of that competence, or *synthesise* new knowledge contributing to the topic.

FAIR4S describes topics covering stewardship, and offers examples of competence statements at each level. By identifying gaps between current and desired levels, FAIR4S can be used to identify relevant skills resources, eg. training events and learning materials that can be used to acquire competence.

Organisation, the second dimension, involves assessing the level of capability needed. The Framework uses the term *capability* to refer to the collective competences a group needs. The group can be at research team, or organizational level, and capabilities may be in-house or third party.

FAIR4S offers capability statements that contextualise its subject matter for the research group and organisational level. The capabilities are generic and can be used to guide definition of roles required, whether for a generalist or to support more specific services (more on this under point 5 below).

Responsibility is the third dimension. FAIR4S defines 5 levels of responsibility, and assumes organisations and teams will match these according to the needs of specific professional groups, roles and job functions. As these will depend on the organisational context, FAIR4S does not give specific statements but offers examples to illustrate the approach.

4. A lifecycle-based approach to describing skills groups

The FAIR4S framework describes ‘skills groups’ - high-level categories representing the competences and capabilities included in the framework. These comprise two broad categories shown in Figure 2.

- 1) *Thematic* competences and capabilities. These can be described according to a typical research lifecycle, as shown in the inner circle in Figure 2. Thematic competences and capabilities are applied in research domain- or project-themed ways, and served by training services and learning materials that are similarly themed according to the domain or project.
- 2) *Generic* competences and capabilities. These are applied in project-independent and cross-domain ways. For example they may be applied in governing and assessing the stewardship of outputs; or in scoping and resourcing the support needed. Or they may involve advising others how to do FAIR research, and managing the enabling infrastructure.



Figure 2. FAIR4S skills groups

Each of the skills groups is broken down further into the competences and capabilities described in part 5 below. The overall model is based on the Research 360 research lifecycle, used in Essentials 4 Data Support,¹ and the DCC RISE model of research data management capabilities.²

¹ Essentials 4 Data Support: <https://datasupport.researchdata.nl/en/start-the-course/i-definitions/research-lifecycle/>

² DCC RISE: <http://www.dcc.ac.uk/resources/how-guides/RISE>

5. Describing individual-level competences and organisation-level capabilities

In FAIR4S ‘competences’ and ‘capabilities’ each describe an ability to achieve desired outcomes by applying skills. The FAIR4S skills groups are high-level descriptions of both kinds of ability. These are broken down further into competences and capabilities. A competence describes an ability that an *individual* can acquire and apply as a skill. A capability describes what happens when individuals apply their skills together in a *team or organisation*, using a service, or enabling others to do so.

FAIR4S describes the top level of the model - the 9 skills groups. At the next level there are 59 topics relevant to professional groups involved in open science data stewardship, according to their role in service application as a user or operator.

Competence and capability statement examples are given for selected topics, using the dimensions mentioned earlier. Competence statements are described at 3 levels: comprehend (basic level), apply (intermediate level), and synthesise/evaluate (expert level). The capability statement for the same topic has 2 levels; research team or organisation. The table below illustrates this for the ‘plan and design’ skills group.

Key to expertise levels: comprehend (basic level) ○ apply (intermediate level) ◐ synthesise/ evaluate (expert level) ●

FAIR4S competences and capabilities		Recommended expertise by professional group and service role								
		MIN.	Service users				Service operators			
			Researcher	Data Scientist/analyst	Data service engineer	Data manager/curator	Researcher	Data Scientist/analyst	Data service engineer	Data manager/curator
Plan and design	Planning data management and sharing (DMP)	○	◐	◐	○	●	●	●	○	●
	Open data model and database design	○	○	●	●	○	◐	◐	◐	●
	Metadata, persistent id. specification	○	○	◐	○	◐	○	○	◐	◐
	Open source software / service requirements	○	◐	●	●	◐	◐	●	◐	●
	Repository and data management platform appraisal	○	○	◐	○	●	◐	○	◐	●

Table 1. Extract from FAIR4S competence and capabilities

The table gives a minimum level representing a basic undergraduate-level knowledge of data stewardship. This is based on the ‘data information literacy’ skills framework (Sapp Nelson, 2017), and the Open Science Career Assessment Matrix proposed by the working group on Rewards under the Open Science Policy Platform. The other main source for the competences listed in the table is the EDISON Data Science framework. The complete table is included as an annex to the EOSCpilot Skills and Capability Framework (D7.3), available from the EOSCpilot website here:

<https://www.eoscipilot.eu/content/d73-skills-and-capability-framework>

The range of topics in FAIR4S is intended to cover the skills to both use and operate the kinds of EOSC service foreseen in the EOSC implementation

roadmap.³ EOSCpilot FAIR4S recognises that the nature of EOSC services will evolve as the EOSC matures. It also recognises that research organisations vary greatly, and will want to tailor their descriptions of local services and FAIR stewardship capabilities to support the kinds of research they do. As the relevant skills will change rapidly, FAIR4S offers organisations the dimensions and examples they can adapt for their own purposes, rather than a comprehensive set of off-the-shelf competence and capability statements for every topic conceivable.

Use Cases

The main use cases for the FAIR4S framework are expected to be:

Specification- helping EOSC Service providers or operators specify the competences and capabilities involved in using their service. The main steps to do this:

- 1) Use the FAIR4S skills list to find topics on which the service offers users added value i.e. high-level requirements the service fulfils
- 2) For each topic describe the capability offered using the existing examples for guidance
- 3) Identify the professional group targeted, and expertise level
- 4) Identify relevant material from the EOSC training and learning resources catalogue *
- 5) Link the entry in the EOSC service catalogue to further information on the learning resource

Learning- helping EOSC users to identify learning resources for their own professional development, relating to specific services and more generally to FAIR/open research

- 1) Use the EOSC Service Catalogue to find services relevant to the research goals (or research support goals) and follow the link to training information, if present
- 2) Find the closest match in the FAIR4S skills groups to the Service Catalogue entries for 'user value' and 'target groups' and (preferably) from any training information provided
- 3) From the topics selected from the FAIR4S skills list, compare the suggested competence level for your professional group and career level with those you have already.
- 4) Where there are any gaps, use an EOSC training and learning resources catalogue * to identify material relevant to your professional group and career level

³ The 5 main EOSC service classes include: 1) persistent identification, authentication, 2) workspace, 3) catalogues of datasets and data services, 4) services to make data FAIR, and 5) service information and guidance; according to the Implementation Roadmap. For further details see: http://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf

Planning- helping research leads and human resource professionals to plan skills development.

- 1) Use the EOSC Service Catalogue to find services relevant to the research goals (or research support goals) and follow the link to training information, if present
- 2) Find the closest match in the FAIR4S skills groups to the Service Catalogue entries for 'user value' and 'target groups' and (preferably) from any training information provided
- 3) Describe any additional roles needed, adapting the FAIR4S capability statements for skills groups relevant to the service, referring to EDISON role profiles where relevant
- 4) Select topics selected from the FAIR4S skills list, within the relevant professional group and identify the expected level of expertise and responsibility appropriate for the career level
- 5) Identify a) the tasks the role is expected to deliver for the team/organisation , b) verbs expressing expected behaviours.

6. Towards FAIR training on FAIR data stewardship

EOSCpilot aims to recommend effective strategies for the EOSC to support education and training. These will be consistent with existing recommendations of the OSPP working groups on Skills and Rewards, and intended to help the relevant stakeholders implement those recommendations.

FAIR4S focuses on stewardship as this is an area of 'data expertise' highlighted by successive reports. EOSCpilot WP7 also aims to identify the means to deliver training and learning resources, whether for stewardship or other skills. To that end, in the D7.2 report we consider the applicability of FAIR principles to training, and set out how 'training as a service' could offer a coordinated approach.

EOSCpilot stakeholder consultation has so far led us to draft conclusions below. We now seek wider input to check which of these have broad agreement, and on the strategies that will promote their implementation. Please share your views on the actions that need to be taken here:

<https://ec.europa.eu/eusurvey/runner/EOSCpilotSkillsSurvey>

DRAFT CONCLUSIONS

- 1) A common EOSC framework of skills and competences for data stewardship will help researchers and support professionals to implement FAIR and secure data principles, by identifying relevant skills they should receive recognition for acquiring .
- 2) The FAIR4S framework will help promote education for data stewardship experts, and their reward and recognition.
- 3) The EOSC should monitor training provided, and determine where there are gaps that need filled to ensure data skills keep pace with changes in science and society.
- 4) EOSC should broker skills exchange between research infrastructures and institutions to help ensure training and development is effectively targeted.
- 5) Data skills may be considered either 'generic' (cross-project or cross-domain) or 'thematic' (project-specific or domain-themed). EOSC should develop and/or commission training and learning resources to fill gaps in generic skills and capabilities, leaving domain-themed services and infrastructures to address any thematic skills gaps.
- 6) Training-as-a-service in EOSC should provide a *training catalogue and registry*, aggregating metadata about training and learning resources, from training portals offered by) organisations in the EOSC ecosystem.
- 7) EOSC should offer certification of EOSC 'compliant' and 'compatible' training and learning resources, which must meet a minimal set of criteria to ensure these resources are findable, accessible, interoperable and reusable. These criteria should be consistent with FAIR principles and good practices recognised by the Open Educational Resource community.
- 8) An EOSC skills stakeholder group should determine the set of minimal criteria and standards required for 'EOSC compliant' training resources, and a further set of recommended criteria and standards for resources to be 'EOSC compatible', starting with those drafted by EOSCpilot.
- 9) EOSC should coordinate and support further work to relate standards on learning resource metadata (e.g. LRMI), learning records (e.g. experience API), badges (e.g. Open Badges) and contribution to research outputs (e.g. CRedit) to improve information flow on FAIR/open science skills delivery, acquisition, and help measure resulting contributions to open science.