# D7.5: Strategy for Sustainable Development of Skills and Capabilities

| Author(s) | Angus Whyte, Ellen Leenarts, Jerry de Vries, Frans Huigen, Eileen Kuehn, Gergely Sipos, Vasso Kalaitzi, Elly Dijk, Sarah Jones, Kevin Ashley |
|---|---|
| Status | **Final** |
| Version | V1.1 |
| Date | 29/04/2019 |

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 report summarises the outcomes, recommendations and conclusions of the EOSCpilot WP7 on Skills and Capability. The emerging EOSC ecosystem provides the context for a focus on developing the data stewardship skills of researchers and professional groups supporting research, including data stewards themselves. The report sets these against the broader landscape, with a brief recap and update of related skills initiatives. It then offers a strategy comprising two main aspects.

The first is a skills and capability framework (FAIR4S), which revises that previously published in D7.3: Skills and Capability Framework, following consultation and test applications of the framework. The second aspect is a model for 'training as a service', which is accompanied by an assessment of models for implementing a registry of learning resources, which is a key part of the TaaS model.

Finally, the report offers recommendations on actions for stakeholders and draws conclusions from the work undertaken. Annexes to the report provide details of the FAIR4S framework and supporting information for the comparison of learning resource registry models.

| Document identifier: EOSCpilot –WP7-D7.5 | |
|---|---|
| Deliverable lead | UEDIN-DCC |
| Related work package | WP7 |
| Author(s) | Angus Whyte, Ellen Leenarts, Jerry de Vries, Frans Huigen, Eileen Kuehn, Gergely Sipos, Vasso Kalaitzi, Elly Dijk, Sarah Jones, Kevin Ashley |
| Contributor(s) | René van Horik (DANS), Marta Teperek (TU Delft), Natasha Simons (ARDC), Nancy Hoebelheinrich (ESIP), Niall Beard (U. Manchester, ELIXIR) |
| Due date | 30/03/2019 |
| Actual submission date | 29/04/2019 |
| Reviewed by | Cees Hof; Celia van Gelder |
| Approved by | Mark Thorley (UKRI) |
| Start date of Project | 01/01/2017 |
| Duration | 28 months |

## Versioning and contribution history

| Version | Date | Authors | Notes |
|---|---|---|---|
| 0.1 | 08/01/2019 | Angus Whyte | ToC Description |
| 0.2 | | Angus Whyte, Jerry de Vries, Frans Huigen, Vasso Kalaitzi | Landscape, skills framework |
| 0.3 | 11/03/2019 | Ellen Leenarts, Elly Dijk | Landscape, rules of participation |
| 0.4 | 20/03/2019 | Rene van Horick, Nancy Hoebelheinrich, Niall Beard, Natasha Simons | Cost-benefit information, data stewardship workshop skills responses |
| 0.5 | 25/03/2019 | Angus Whyte, Kevin Ashley, Ellen Leenarts, Eileen Kuehn, Jerry de Vries, Frans Huigen | All sections |
| 0.6 | 18/04/2019 | Angus Whyte, Kevin Ashley, Sarah Jones, Gergely Sipos | Finalising deliverable for review |
| 1.0 | 26/04/2019 | Angus Whyte; Kevin Ashley | Corrections in response to review comments; Final proof checking |
| 1.1 | 29/04/2019 | Mark Thorley (UKRI) | Final typographic proof-read and edit. |

# TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## EXECUTIVE SUMMARY

Developing and sustaining the skills of researchers, research support staff and EOSC service providers is essential for the success of the EOSC vision. This report provides evidence from the EOSCpilot project and elsewhere on this issue, summarises the (sometimes fragmented) work taking place in Europe and elsewhere to address it, and makes a number of recommendations on actions to be taken by EOSC stakeholders following the end of the EOSCpilot project.

The project developed a skills and competence framework (entitled FAIR4S) which has undergone extensive revision after wide consultation. The framework is accompanied by worked examples of its application in a variety of contexts. The project has also developed a view of how training - one important method for skills acquisition - can be presented as a service within EOSC. Most of our recommendations relate to one or other of these outputs.

Key recommendations are:

1. EOSC Executive should ensure the creation of a registry of learning resources and training opportunities, accessible to all players through the rules of engagement.

2. The EOSC Governance should engender a coordinated approach among stakeholders in skills development in order to share learning and training resources in a FAIR way, promote broader application of useful resources and avoid duplication of effort.

3. Employers, learning providers and others should utilise the FAIR4S framework to describe learning resources and training opportunities, and to produce job descriptions in areas relevant to EOSC.

4. Further EOSC projects, including those funded under INFRAEOSC calls, and other stakeholder organisations among the 'coalition of doers' engaged in skills development should be supported to adapt and apply FAIR4S, e.g. to offer disciplinary adaptations of the skill and role profiles.

5. The registry should utilise metadata associated with the FAIR4S framework to assist in locating and describing the materials within it.

6. EOSC governance should assign responsibility for monitoring the provision of training and learning against the FAIR4S framework, allowing any gaps in provision to be identified.

Recommendation 2 includes an extensive set of specific actions for various stakeholders – see Section 2.2 for full details of these actions.

# 1. INTRODUCTION

## 1.1. Overall objectives in the project context

The scope of the EOSCpilot project includes raising awareness of the importance of data skills that will help change incentive structures for researchers to share data, and realise the social and economic benefits of treating shared data as a public good. Recognising the importance of an integrated approach to data skills development, the EOSCpilot project has a dedicated work package (WP7) for 'skills and capability'. The overall objectives of the work package are as follows:

1. Design an open data science skills framework described in terms of the individual competencies and organisational capabilities required to provide EOSC services of the required levels of quality.

2. Catalogue the current provision of education and training against this framework and identify gaps in delivery.

3. Develop an EOSC education and training strategy to address the gaps and set up a sustainable technical training infrastructure to ensure shared resources are openly accessible and reusable.

4. Coordinate delivery of relevant education and training materials and events to address those gaps that stakeholders identify as of the highest priority.

5. Connect with established national and international training schools and institutions, and collaborate with "champion" universities in their implementation of open-data science curricula.

## 1.2. Purpose of this document

This final report of EOSCpilot WP7 offers a strategy to address these objectives comprising two main aspects. The first is a skills and capability framework ('FAIR4S'), which revises that previously published in deliverable D7.3: Skills and Capability Framework[3], following consultation and test applications of the framework. The second aspect is a model for 'training as a service' (TaaS), which is accompanied by an assessment of models for implementing a registry of learning resources - a key part of the TaaS model. The main purposes of the report are therefore as follows:

- Make recommendations for future action by EOSC, its supporting projects, and other stakeholders.

- Review the rationale for a sustainable skills and capability strategy, placing this in the context of the evolving EOSC landscape and related initiatives on data skills, particularly those for stewardship of research outputs

- Describe the skills and capability framework, named 'FAIR4S'. This refers to stewardship skills to make the outputs of science and scholarship FAIR, and keep them FAIR. The community consultation on the draft version of the framework is described, with a summary of the changes made as a result.

- Recap the evidence of skills gaps collected from three sources; reviews of skills resources already provided, consultative workshops with the training community, and analysis of the reports from 'science demonstrators' conducted in the pilot project.

- Describe training the work package partners provided to address the skills gaps, through a Summer School (GridKa School) and webinars.

- Describe a model for 'training as a service' and options for implementing this, providing exemplars of these from current initiatives, and comparing their costs and benefits.

The report also describes activities undertaken by project partners in this work package which have not been described in other deliverables.

## 1.3. Relationship to other project outputs

As the final deliverable from WP7 the report summarises the key outputs seen in earlier deliverables. It also describes how the project's work in this area is being taken up within EOSC and more widely, and provides further recommendations for taking the project's work forward. WP7's earlier outputs, and this document's relationship to them, are summarised in Table 1:

**Table 1: Relationship to earlier deliverables**

| WP7 Deliverable | How D7.5 builds on former WP7 deliverable |
|---|---|
| **D7.1: Skills landscape analysis and competence model[1]**<br><br>This deliverable provided an analysis of EOSC stakeholder policy priorities for skills development, surveyed current training provision and presented the first draft of the competence framework. | The first draft of the competence framework was further elaborated in D7.3, and is presented as a final version in D7.5. |
| **D7.2: Interim report and catalogue of EOSC skills training and educational material[2]**<br><br>This deliverable presented the first draft of the skills framework and related it to the current provision of training resources, identifying gaps in provision. It also provided the initial description of training as a service within a future EOSC and expressed initial views on the applicability of the FAIR notion to training resources and provision. | In D7.5 we present more developed ideas on training as a service, informed by the work carried out elsewhere in the EOSCpilot project and more widely, and recommendations on how EOSC can address gaps in training provision. |
| **D7.3: Skills and Capability Framework[3]**<br><br>This framework, now known as the FAIR4S framework, is the core output of this deliverable. It provides a set of core competences for data stewardship, relates topics to expertise levels and provides examples of capability and competence statements. | D7.5 updates the framework following feedback since its release in July 2018. We also describe how the framework has been taken up in Europe and elsewhere and provide recommendations for its further use and development. |
| **D7.4: Report on Training Workshops[4]**<br><br>This deliverable documents the delivery and outcomes of the four main WP7 training events held between November 2017 and August 2018. Three were consultative workshops; one, a data science training school, tested the applicability of the skills framework to a broad data science curriculum. | In D7.5 we summarise how the workshops influenced the development of the FAIR4S framework and show a further application of the framework to link it to materials from GridKa School and CODATA data science summer schools. |

---

[1] See: https://www.eoscpilot.eu/content/d71-skills-landscape-analysis-and-competence-model.
[2] See: https://www.eoscpilot.eu/content/d72-interim-report-and-catalogue-eosc-skills-training-and-educational-materials.
[3] See: https://www.eoscpilot.eu/content/d73-skills-and-capability-framework.
[4] See: https://www.eoscpilot.eu/content/d74-report-training-workshops.

## 1.4. Structure of this document

Section 2 presents our final conclusions and recommendations for future action. Busy readers are recommended to start with this section, then consult Annex B.5. Skills Tables, for a compact representation of which levels of expertise are expected of different roles in data-intensive research, finally consulting the other sections if they are seeking to understand how the conclusions and recommendations have been reached.

Section 3 of this document reiterates the case for a strategy and framework for skills development in EOSC that was made in the project proposal and in earlier deliverables, updating that case to take account of developments during the project lifetime. It describes some disciplinary approaches, illustrates how skills development fits into the broader context of EOSC and describes related developments on a broader, global stage.

Section 4 reviews the skills gaps and development priorities identified through the project, describes the training delivered to address these, and revises the FAIR4S skills and capability framework developed by the project, including consultation on the first version published in D7.3.

Section 5 describes the model for training as a service in EOSC and makes the case for an EOSC registry of training resources through a comparison of the costs and benefits of current models.

Annex A contains a glossary.

Annex B presents the current versions of the FAIR4S skills tables, templates for skill and role profiles, and multiple worked examples of both skills and role profiles. It also contains compact tabular representations of levels of expertise expected of various roles.

## 2.    CONCLUSIONS AND RECOMMENDED ACTIONS

### 2.1.  Introduction

The EOSC requires a concerted and coordinated effort to provide sustainable access to learning resources produced by EOSC projects.  This is not yet evident - there is much activity but it is still fragmented and siloed, although some points of connection are emerging. Section 3.3.3 of this document offers a brief, and likely incomplete, snapshot of current EOSC projects, including ESFRI clusters, that we believe to be producing new training materials or events. These are but a small fraction of the learning resource information that could be provided as a service to EOSC users from the full range of EOSC 'doers'.  The EOSC Implementation Roadmap[5]  comments, for example, that:

> *"Marie Skłodowska-Curie actions have been increasingly encouraging the incorporation of open science skills and knowledge training into the 'Innovative Training Networks' action (MSCA-ITN), both for early-stage researchers, and in the 'Co-funding of regional, national and international programmes' action (MSCA-COFUND), for excellent doctoral and postdoctoral R&I programmes across Europe. Such systemic activities foster a culture of data sharing and FAIR data management and help develop the human resource capital required for the EOSC to function." (p.13)*

Other sources could include, for example, research infrastructures, universities, other research producing organisations and third-party training organisations (SMEs) from anywhere.  However there is currently no overall plan or activity at EOSC level to ensure coordinated provision of the outputs of training or other forms of skills development (such as staff exchanges, or fellowships). Hence we recommend that:

| **Recommendation 1** |
| :--- |
| EOSC Executive should ensure the creation of a registry of learning resources and training opportunities, accessible to all players through the rules of engagement. |

The training resources and services offered via the EOSC portal should be described in a FAIR way and should encompass the many providers of free and chargeable services. Some of these training materials will be developed in EU funded projects - specifically via FAIRsFAIR and the disciplinary cluster projects such as ENVRI-FAIR - adding to the many online resources already developed in OpenAIRE, FOSTER and by many others initiatives surveyed in EOSCpilot deliverables D7.1 and D7.2.  These reports demonstrated that there is extensive existing delivery of online material within the scope of the EOSCpilot WP7 focus on the professional development of data stewardship skills in the context of open science and data science (the skills to make data FAIR and keep it FAIR).  This includes a variety of delivery formats, from materials resulting from workshops with a focus on skills development, to Seasonal School ('Summer School') courses over a number of days, to MOOCs, webinars, and online materials intended for self-directed discovery and learning.

Surveying all these forms of activity and discussing them with stakeholders we can identify a number of key points and themes that inform our recommended actions for stakeholders.

1.  Over the course of the EOSCpilot project, the Seasonal School ('Summer School') format has become a more widely established and evidently popular format for face-to-face learning and teaching; EOSCpilot's WP7 directly supported GridKA 2018 Summer School, and participated in the 2018 CODATA / RDA Research Data Science School. Similar and longer-established events have been supported by other EOSCpilot participants, e.g. PRACE and ENVRI.

---

[5] European Commission (2018) Implementation Roadmap for the European Science Cloud (Staff Working Document SWD(2018) 83) https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf.

2.  Digital research stewardship skills - those to make data FAIR and keep it FAIR - are still evolving. They encompass a set of competences that span the dynamic areas of data science and open science. There is strong interest in a framework that can serve both of these areas; one that can offer comprehensive guidance to individuals and organisations aiming to develop the competences and capabilities to implement FAIR principles and make effective use of EOSC services.

3.  The FAIR4S framework presented in this report offers a synthesis of preceding and parallel work, covering the very dynamic area of digital research stewardship. FAIR4S offers a basis for successor projects to scope their skills development actions. With further development of an ontology based on it, planned in GO-TRAIN, it will offer a tool for describing materials in a machine readable, and therefore more FAIR, way. In addition to the forthcoming Annex to this document the framework is currently maintained at GitHub: https://github.com/EOSC-FAIR4S/fair4s-framework and is showcased for the most relevant use cases at https://github.com/EOSC-FAIR4S/eosc-fair4s.github.io and can be accessed at https://eosc-fair4s.github.io.

4.  Learning resources on FAIR data and open science are provided globally, often free at point of use. The true extent of existing materials is difficult to establish as they are not made discoverable in a consistent way.  There are nevertheless perceived gaps in coverage of certain areas. These range from 'soft skills' in areas such as communication, to policy areas including data ethics and integrity, to more technical competences in cloud storage, workflow documentation, and computational notebooks.

5.  Communities of Practice for research teams, data stewards and training providers will be useful to provide opportunities for networking and continuing professional development following training. Indeed, the CODATA/RDA Schools have been particularly successful in assisting students to transition from attendees to helpers to course conveners to share their skills with broader groups. Communities of Practice should be framed broadly to extend across all stakeholders in the EOSC, and ideally beyond to support the transition to global platforms. Use of international fora such as the RDA will be particularly helpful in this regard and already support Interest Groups on Skills and Open Science Platforms.

6.  A culture change does not happen overnight. Trainers and coordinators have mentioned at all our workshops that they would like to see training catalogues to find training materials and training events more easily, yet there have been few initiatives to create such catalogues. As is described below this might be because creating and maintaining such a catalogue is quite costly (see Section 3.3.4) and the added value is not always clear. It first needs a common agreement on a standard for training metadata e.g. bioschemas.org in the case of the life sciences. When a standard is agreed it takes additional effort to make sure these data can be harvested, requiring effort centrally to provide the tools and effort by the provider of the training content to adjust their publishing procedure to the standard agreed.

7.  Many institutes, projects and research infrastructures publish their training materials and events in different ways and with different metadata. In some cases a domain specific catalogue has emerged (e.g. ELIXIR TeSS) that uses a domain specific metadata schema e.g. bioschemas.org in the case of TeSS. The life sciences with ELIXIR has extended knowledge on how to build and maintain the catalogue and what its added value is. Within a domain the added value is clear as the training can be on similar or overlapping topics. One can argue that a culture change in this regard is best reached within a domain as the added value of a training catalogue on similar topics is more obvious.

8.  Manually curating registries of learning resources is time consuming and therefore expensive.  The semi-automated harvesting approach adopted in ELIXIR TeSS appears, on the basis of our limited cost-benefit analysis, to be more efficient and sustainable, especially for event-related training information that is more subject to change.

9. Contributing to raising the awareness around improving the FAIRness to learning resources EOSCpilot has become a member of the Community of Practice for Training Coordinators[6] (CoP) that has been in place since September 2018. EOSCpilot presented at the communities' face-to-face meeting at the DI4R a presentation on the FAIR4S framework and how to use it and on improving the FAIRness of training resources. EOSCpilot also supported the CoP's activity of getting into contact with the Editorial Board of the EOSC portal for their strategy in providing information on learning services and resources.

10. EOSCpilot is not alone in improving the FAIRness of training resources. Recently the ELIXIR FAIR Training Working Group has assessed the FAIRness of some of their training resources. A range of techniques have been utilised to improve the FAIRness e.g. by adding 'licence' to the bioschemas metadata profile so that training providers are reminded to provide a licence. Such efforts build on longer-standing efforts in the OER (Open Educational Resource) community to develop metadata for open learning resources, which are utilised for example by the US-based Data Management Training Clearing House.

## 2.2. Actions for stakeholders

> **Recommendation 2**
>
> The EOSC Governance should engender a coordinated approach among stakeholders in skills development in order to share learning and training resources in a FAIR way, promote broader application of useful resources and avoid duplication of effort.

This coordination should include:

a. Working with the FAIRsFAIR project, RDA, CODATA, GO-FAIR initiative, and EOSC projects to ensure they apply and further develop FAIR4S towards a core curriculum in FAIR data skills for researchers, data scientists, data stewards, other data advisors and data service providers. This should be delivered both through professional development training for existing staff and model courses useable by Higher Education Institutions.

b. Engaging with the FAIRsFAIR project to produce a learning resource registry or knowledge base, aggregating metadata about training and learning resources, from training portals offered by existing (and future) organisations in the EOSC ecosystem, particularly the ESFRI cluster projects.

c. Through its Working Groups, engaging with a Community of Practice of Training Coordinators, comprising coordinators of skills development in infrastructures and institutions.

d. Further definition 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.

e. Brokering skills exchange between research infrastructures and institutions to help ensure training and development is effectively targeted to the users of the EOSC. In addition to the improved awareness of training materials, which a learning resource registry would facilitate, this should encompass staff exchanges for staff engaged in training data stewards - as the RItrain project has done for managerial staff[7].

f. Monitoring training provided, and determining where there are gaps that need filled to ensure data skills keep pace with changes in science and society.

g. Supporting 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.

---

[6] See: https://www.openaire.eu/cop-training
[7] See: http://ritrain.eu/staff-exchanges

CRediT) to improve the flow of information on FAIR/open science skills delivery, acquisition, and measurement of resulting contributions to open science.

The FAIR4S skills framework provides all stakeholders with a set of competences that describe the stewardship skills needed to make data FAIR and keep it FAIR, updating those in deliverable D7.3. The framework relates these skills to the capabilities that EOSC services will augment for their users, and also to the roles that researchers, data stewards, and other professional groups fulfil. We have the following recommendations relating to use of the framework.

---

**Recommendation 3**

Employers, learning providers and others should utilise the FAIR4S framework to describe learning resources and training opportunities, and to produce job descriptions in areas relevant to EOSC.

---

**Recommendation 4**

Further EOSC projects, including those funded under INFRAEOSC calls, and other stakeholder organisations among the 'coalition of doers' engaged in skills development should be supported to adapt and apply FAIR4S, e.g. to offer disciplinary adaptations of the skill and role profiles.

---

**Recommendation 5**

The registry should utilise metadata associated with the FAIR4S framework to assist in locating and describing the materials within it.

---

The GO-TRAIN Implementation Network could be one route to achieving some of the goals of Recommendation 5.

Finally, it is clear that some kind of monitoring action will be necessary to ensure that provision is made to ensure that we are aware of gaps in training provision for necessary skills, even though our consultation exercise did not produce unanimity on where responsibility lay for filling such gaps.

---

**Recommendation 6**

EOSC governance should assign responsibility for monitoring the provision of training and learning against the FAIR4S framework, allowing any gaps in provision to be identified.

---

## 3.     SUSTAINING SKILLS AND CAPABILITY DEVELOPMENT IN EOSC

### 3.1.  Introduction

The emerging EOSC ecosystem provides the context for our focus on developing the data stewardship skills of researchers and professional groups supporting research, including data stewards themselves. This chapter sets these against the broader landscape, with a brief recap and update of data skills initiatives. It takes as its focus support for 'open data science', and for implementation of the FAIR data principles to govern stewardship of the outputs of that science and scholarship activity. Figure 1 illustrates the core skills the work package is concerned with, which can be described as 'the skills to make data FAIR and to keep it FAIR'.
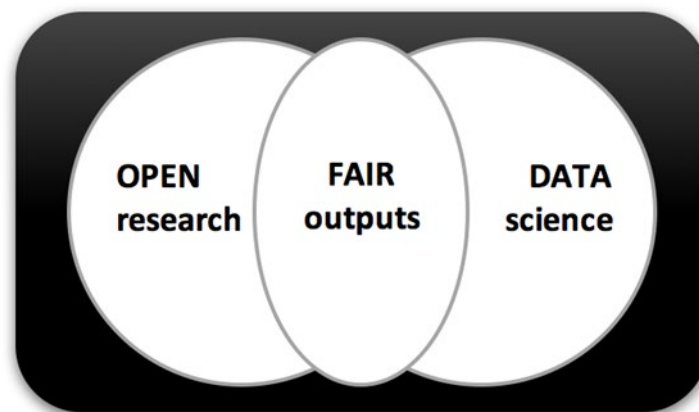


**Figure 1: Skills for FAIR outputs of Open Research and Data Science**[8]

Our work looks at skills in this context - the competences individuals need, and the capabilities that groups and organisations can deliver using individuals with those competences. To make effective use of the EOSC, individuals and organisations want to be able to find out how to acquire skills (together with the appropriate knowledge and attitudes to apply them), and how to demonstrate to others that they have them. They also want to know what kinds of EOSC services will help them to apply their skills and amplify their organisation's capabilities.

This chapter has three main sub-sections that cover the following aims:

- Recap the policy rationale for a sustainable skills and capability strategy, considering the major development in the policy landscape since the WP produced its first report D7.1;

- Provide an overview of the current EOSC projects also delivering training in areas related to our focus, as evidence of the need to make the training outputs FAIR in a coordinated manner;

- Recap the broader landscape of skills strategy and training in skills to implement the FAIR principles, including research data management and stewardship, as evidence of the likely demand from stakeholder organisations for services to make their training outputs FAIR.

To be sustainable, an EOSC strategy for skills and capability development must recognise and embrace the work being undertaken by stakeholders in this area both within Europe and globally. These include funding organisations, research producing organisations and research libraries. This chapter offers an overview of work being undertaken by one such body (LIBER). We also briefly highlight two examples of national strategies for research data management (those of the Netherlands and Sweden) that emphasise development of skills and capability.  As the EOSC further develops its strategy for skills and education, such

---

[8] Adapted from 'the relationship between FAIR and Open' CC-BY European Commission FAIR data Expert Group.

actions will need to be continually reviewed through the governance structure, including input via the EOSC Stakeholders Forum.

## 3.2. Why EOSC needs a strategy for sustainable development of skills and capabilities

A number of documents produced before and during the EOSCpilot project make clear that there is a significant skills gap to be closed if the vision of EOSC is to be made reality. Some of those documents, and the EOSCpilot project proposal, argued that closing this skills gap required the development of a skills and capability framework. We argued that this would offer a reference model and method for planning professional development in the stewardship of data and other research outputs. With the Framework, individuals can identify the skills they need and find training or other resources that will let them acquire them. Training providers can use the Framework to ensure that material is available to fill gaps in skills provision. Organisations can better define the skills they need from their researchers and professional staff.

We summarised much of the earlier work in previous deliverables (notably D7.1 & D7.3) and do not repeat the details here. We do list the documents again, some key points from them, and add information on some further initiatives that have produced relevant outputs in the last two years.

Our earlier work drew attention to documents specifically addressing EOSC, such as the first EOSC High Level Expert Group report[9], EOSC Declaration[10] and the EOSC Implementation roadmap[11]. We also noted the Digital Education Action Plan[12], with its broader scope of skills relevant to all areas of digital change. In addition, we noted that the Commission's revised Recommendation on access to and preservation of scientific information[13] identified a number of key expectations in the area of skills and competences.

The EOSC Declaration includes three paragraphs that set out three requirements for EOSC in the skills area:

1. On skills, the declaration calls for "the necessary skills and education in research data management, data stewardship and data sciences" to be provided throughout the EU "as part of higher education, the training system and on-the-job best practice in the industry".

2. Regarding data stewardship, the Declaration states that researchers need the support of adequately trained data stewards, and calls for investment in their education "...via career programmes delivered by universities, research institutions and other trans-European agents."

3. On rewards and incentives, the declaration notes these are essential for researchers who make research data open and FAIR for reuse, and/or actively reuse and reproduce data. As well as in project and career evaluation, these rewards should be reflected in "...other career policies in universities and research institutions (appointments, promotions etc.)."

The emphasis here on data stewardship and on skills to help address the FAIR principles is echoed in other reports such as that of the first High Level Expert Group (HLEG) report on EOSC. Similar emphasis appears in the April 2018 recommendations of the Open Science Policy Platform[14] which also includes broader recommendations on skills. Its fourth recommendation is to "Foster Open Science literacy as essential to European competitiveness at the global level, together with other digital and information competencies.

---

[9] First report of High Level Expert Group on the EOSC (2017): https://ec.europa.eu/digital-single-market/en/news/first-report-high-level-expert-group-european-open-science-cloud.

[10] European Commission DG research & Innovation (2017): EOSC Declaration.
https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

[11] European Commission (2018) Implementation Roadmap for the European Science Cloud (Staff Working Document SWD(2018) 83) https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf.

[12] European Commission (2018) Digital Education Action Plan COM(2018) 22 final.
https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:52018DC0022

[13] European Commission (2018) Recommendation on access to and preservation of scientific information.
http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51636

[14] See: doi:10.2777/958647.
https://publications.europa.eu/en/publication-detail/-/publication/5b05b687-907e-11e8-8bc1-01aa75ed71a1/language-en

Member States need to secure support for the development of an accredited curriculum for Open Science skills training…". Its recommendations on skills encompass actions for researchers, students and research support staff and it calls for the recognition of a "common set of baseline skills" - a strong argument for the need for a recognised skills framework.

We noted in earlier outputs that one problem with all these requirements is that they assume a common understanding of the skills needed to carry out these roles of data steward, data scientist, research data manager and so on. For the market to work effectively it is important that educators provide people with skills that are recognised as relevant by those who will later seek to employ them in these roles, for instance. Training that could provide professionals with needed skills must be easily identifiable as such, and research groups must be able to define what collective set of data skills they need to use EOSC to advance their research goals most effectively. It is for these reasons, amongst others, that EOSCpilot delivers a common Framework to be used in the widest variety of settings. We also proposed a definition of the term "data stewardship", repeated here in Box 1.

> *Stewardship for open data science = 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. In short, stewardship applies skills to make data FAIR, and keep it FAIR.*

**Box 1: Definition of stewardship for FAIR and open data science**

## 3.3. The emerging EOSC ecosystem

### 3.3.1. EOSC governance takes shape

The EOSC was officially launched by the EC in November 2018, establishing a governance structure to take forward implementation, encompassing three broad streams of activity illustrated in Figure 2:

- Governance Board representing all member states, associated countries and the EC, together with an Executive Board of eleven organisational representatives and individual experts will steer implementation, with operational support from an EOSC Secretariat[15].

- Stakeholders Forum and a series of Working Groups convened by the Executive Board will advise on implementation.

- A series of EU-funded projects, national initiatives, and wide-ranging projects and activities undertaken by the 'Coalition of the doers' which will contribute to the development and implementation of the core structures, tools and services of the EOSC.

---

[15] EOSCsecretariat.eu: https://www.eoscsecretariat.eu
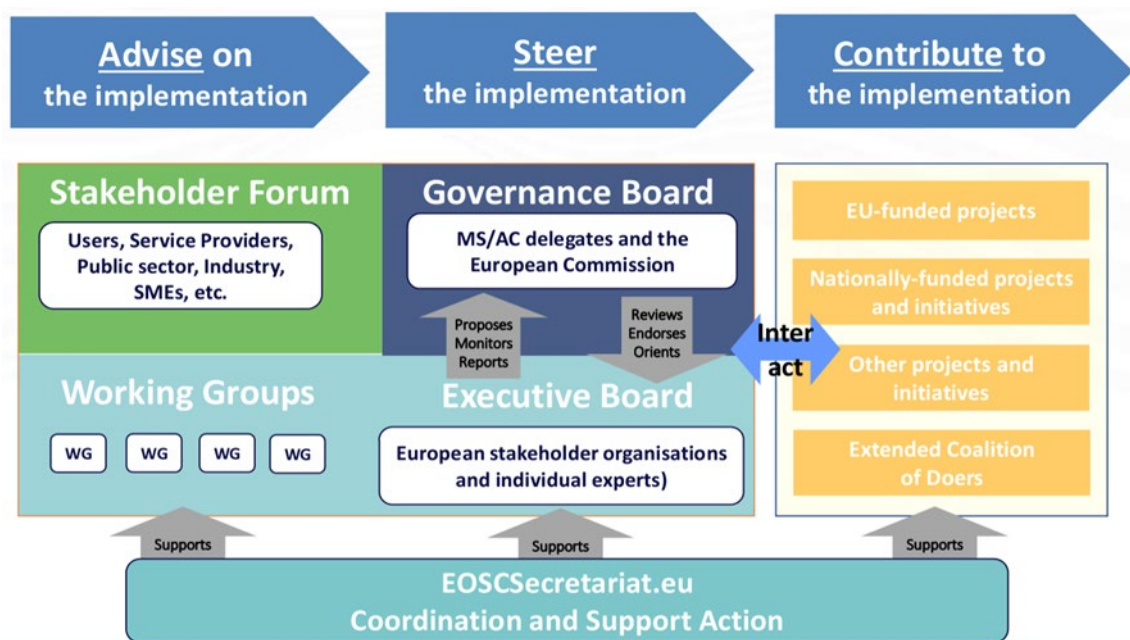
**Figure 2: EOSC Governance[16]**

Skills will be addressed in many aspects of the EOSC governance and implementation. To date, five Working Groups have been agreed by the Executive Board and Governance Board. These are: Landscape, Rules of Participation, Architecture, FAIR and Sustainability. It is expected to convene a dedicated Skills group later in 2019. Aspects related to skills will also be addressed in terms of the research culture and practice in the FAIR group, and in terms of the existing resources available in each country identified by the Landscape group.

### 3.3.2. Rules of participation for service providers are proposed

As EOSCpilot is the first European project in developing the European Open Science Cloud, one of its tasks is to develop the rules of participation for EOSC service providers, as recommendations to the EOSC Governance. Deliverable D2.5: Recommendations for a minimal set of Rules of Participation[17], describes these rules.

This set of Rules of Participation for all EOSC service providers consists of one main rule, complemented by a series of seven specific requirements. These requirements can be applied depending on the needs of each scientific field. The main rule is that "*EOSC services shall be registered in an EOSC-compliant or compatible service catalogue visible to the global EOSC gateway*". In each scientific community there are multiple service catalogues. With this main rule it is recommended that service providers register their services to community accepted and supported catalogues.

The seven specific requirements concern the function of the service (availability, functionalities, maturity, support, terms of use, contractual framework) or the characteristics of the service (quality, performance, sustainability, access policies, data portability and compatibility, amongst others).

The requirements are:

- **Machine readable metadata:** All services must be described in machine readable format and be identifiable by means of a common and persistent identification.

- **Terms of Use and Policies:** All EOSC services must have Terms of Use and other policies displayed publicly online and/or via the EOSC Service catalogue(s).

---

[16] From EOSC Secretariat; Sourced from https://www.slideshare.net/TheEOSChubproject/supporting-eosc-governance
[17] See: https://www.eoscpilot.eu/content/d25-recommendations-minimal-set-rules-participation

- **Accessibility:** EOSC Service Providers must describe how they ensure accessibility and interoperability, and provide information in the EOSC compatible service catalogue where they are registered, to allow for a more open market for users.

- **Portability:** Legal and technical infrastructures should enable the portability of data and services, as much as possible.

- **Access costs and charging model:** Service providers may apply user charges/fees, which could vary by type of service, type of service provider and location of users. This information must be made clear to users online and via the service catalogue(s).

- **Quality of service:** Service providers should adhere to a minimal set of quality guidelines that are being developed within the EOSCpilot project and the EOSC-hub project.

- **Relation to users:** Service providers should be transparent about the data management mechanisms they use to store-process-publish content, considering FAIRness and Reproducibility; Privacy and Security; and Easing usability.

These requirements are relevant to EOSC 'Training as a Service'. Training and other learning materials (e.g. for guidance) may be considered a 'resource'. The same may be said of information about events with a similar purpose. EOSC may also need services that add value for training users by making these resources more findable, accessible, interoperable and reusable (FAIR). This would naturally include training or guidance about other services, as well as training or guidance about implementing FAIR principles more generally.

Further on in this report we review a model developed earlier in the project (D7.1, D7.2, and D7.3) for EOSC 'Training as a Service'. Such a service could provide a comprehensive collection of training materials and events. This would include harvesting training information from a wide range of resources already available from relevant providers. These would include cross-disciplinary ('horizontal') resources, such as those on open science from the FOSTER project, and more discipline-specific ('thematic') resources from the Research Infrastructures and groups of these ('cluster projects' – described in Section 3.3.3.)

The main rule and requirements will naturally apply to training services. For example: the rule that all EOSC services shall be registered in an EOSC compliant or compatible service catalogue also includes that all training services should be registered, as (for example) life science training resources are in the ELIXIR TeSS service.

The seven specific requirements also apply to training and other skills acquisition services: for example the metadata should be machine readable and be identifiable by means of persistent identifiers in order to make an overview of the training materials and events. In order to harvest the metadata, the service provider has to ensure accessibility and interoperability. The implications are articulated further in Section 5 of this document.

### 3.3.3.    EOSC projects emerge with diverse skills activities

The EOSC Portal lists current EOSC projects, which are expanding in number and scope through INFRAEOSC calls, as summarised in the EC Implementation Roadmap for the EOSC.  The roadmap summarises resources the EC has committed to developing data management skills, starting with its 2014-15 work programme, through the EDISON Data Science Framework, FOSTER, and more recently FOSTERplus. The 2016-2017 programme funded OpenAIRE Advance, RDA Europe, FREYA and EOSC-hub to also address data-related skills development for e-infrastructure providers and users, and support for FAIR data management.

This sub-section of the report offers a snapshot of these projects, followed by those funded under the subsequent INFRAEOSC calls, which include a range of new horizontal and thematic EOSC projects. The aim is to illustrate the breadth of skills development activity, rather than to offer a synopsis of the much broader scope of these projects.

**eInfraCentral[18]**

eInfraCentral fosters an open and guided discussion between e-infrastructures in pursuit of its main objective of a common catalogue for their services. This has produced the eInfraCentral portal, which is a single entry point for end-users to browse a service catalogue. While eInfraCentral does not in itself develop training, it is relevant in this context as a contributor to the EOSC Portal, with entries include training services. The "Training" category in eInfraCentral's catalogue currently contains thirteen entries.

**EOSC-hub[19]**

EOSC-hub is the second major source of entries for the EOSC Portal, bringing together multiple service providers to create the Hub: a contact point for European researchers and innovators to discover, access, use and reuse a broad spectrum of resources for advanced data-driven research. 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. It closely collaborates with eInfraCentral, EOSCpilot, GÉANT 4.2, OpenAIRE-Advance and the RDA Europe 4.0 projects to deliver a consistent service offer for research communities across Europe. In this context, EOSC-hub hosts training events and develops training materials, listed under EOSC-hub resources[20].

**OpenAIRE Advance[21]**

OpenAIRE-Advance continues the mission of OpenAIRE to support the Open Access/Open Data mandates in Europe. It consolidates the current infrastructure, comprised of a human network and technical services, while working to shift the momentum among its communities to Open Science, aiming to be a trusted e-Infrastructure within the realms of the European Open Science Cloud. In this next phase, OpenAIRE-Advance strives to empower its National Open Access Desks (NOADs) so they become a pivotal part within their own national data infrastructures, positioning Open Access and Open Science onto national agendas. The capacity building activities bring together experts on topical task groups in thematic areas (open policies, RDM, legal issues, TDM), promoting a train the trainer approach, strengthening and expanding the pan-European Helpdesk with support and training toolkits, training resources and workshops. It examines key elements of scholarly communication, i.e. co-operative OA publishing and next generation repositories, to develop essential building blocks of the scholarly commons. OpenAIRE regularly hosts webinars and workshops, also in collaboration with other projects and initiatives.

**FREYA[22]**

FREYA aims to extend the infrastructure for persistent identifiers (PIDs) as a core component of open research, in the EU and globally. It improves discovery, navigation, retrieval, and access to research resources. New provenance services enable researchers to better evaluate data and make the scientific record more complete, reliable, and traceable. FREYA hosts a Knowledge Hub to help people understand what persistent identifiers are, why they exist, and how to use them for research. The Knowledge Hub is a space where users can come across comprehensive guides and webinars, as well as extra help, where needed.

**FAIRsFAIR[23]**

FAIRsFAIR's ambition is to deliver FAIR-aligned Rules of Participation in the EOSC. These rules will be designed for adoption by the emerging EOSC governance bodies to establish FAIR compliance of components and practices. FAIRsFAIR will also support EOSC by accelerating the realization of its goals and by opening up and sharing all knowledge, expertise, guidelines, implementations, new trajectories, courses and education needed to turn FAIR Principles into reality. The project will thus contribute to a culture change necessary to

---

[18] https://beta.einfracentral.eu/home
[19] https://www.eosc-hub.eu/
[20] https://www.eosc-hub.eu/training-material
[21] https://www.openaire.eu/openaire-advance-project
[22] https://www.project-freya.eu/en/about/mission
[23] https://www.fairsfair.eu/

achieve wide adoption of FAIR practices within the EOSC and beyond. FAIRsFAIR will use the FAIR4S framework.

**EOSC-synergy and EOSC-NORDIC**

EOSC-synergy and EOSC-NORDIC are two examples of regional projects successful in the Horizon 2020 INFRA-EOSC-5(b) call. The former, coordinated by IBERGRID, covers Spain and Portugal whilst the latter project covers Finland, Sweden, Norway, Denmark, Iceland, Estonia, Latvia, Lithuania, Netherlands and Germany. EOSC-synergy's WP6 will build on the work done within EOSCpilot WP7 with regard to creating guidelines on how to create online learning materials that are FAIR using FAIR4S and the deliverables as starting point.

### 3.3.4. Disciplinary projects – the ESFRI Clusters

Starting approximately at the same period of time, in the first months of 2019, a number of the cluster projects take into account the work done by the EOSC horizontal projects and will also attempt to uncover synergies between them, in order to align activities as much as possible. As the selection here shows, they are commonly conducting relevant training activities. EOSCpilot liaises with these projects through overlapping project partners, and by participating in common activities and meetings, such as the ESFRI RIs and EOSC workshop in London, January 2019.

**Astronomy and Particle Physics**

ESCAPE[24] (European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures), with a start date on February 1st, 2019, aims to address the Open Science challenges shared by ESFRI facilities, as well as other pan-European research infrastructures in astronomy and particle physics. The project will provide training and extension of FAIRness standards and tools for data access and data preservation.

**Physical Sciences**

PANOSC[25](Photon and Neutron Open Science Cloud) is a European project (financed by the INFRAEOSC-04 call) for making FAIR data a reality in six ESFRI Research Infrastructures (RIs), developing and providing services for scientific data and connecting these to the European Open Science Cloud (EOSC). User training is within the project's main goals: To train users how to write Data management plans, improve metadata, using DOIs, manage data better. A dedicated work package on staff and user training will be responsible for skills building.

**Earth Sciences**

ENVRI-FAIR[26] is the connection of the Cluster of Environmental Research Infrastructures (ENVRI) to the European Open Science Cloud (EOSC). The overarching goal is that at the end of the proposed project, all participating RIs have built a set of FAIR data services which enhances the efficiency and productivity of researchers, supports innovation, enables data- and knowledge-based decisions and connects the ENVRI Cluster to the EOSC.

A presentation[27] of Dr Maggie Hellstrom at the ESFRI-EOSC liaison workshop, highlights specific needs and requirements for training: a) the research data life cycle & basic research data management, including FAIR, b) legal & policy-related matters, including GDPR and ethics, c) best practice usage of common EOSC services by RIs and d) guidance towards offering ENVRI Community services via the EOSC Marketplace and similar channels.

---

[24] https://cordis.europa.eu/project/rcn/219246/factsheet/en
[25] https://panosc-eu.github.io/
[26] http://envri.eu/envri-fair/
[27] https://www.esfri.eu/sites/default/files/Margareta%20Hellstr%C3%B6m.pdf

**Social Sciences and Humanities**

SSHOC[28] is an EU funded H2020 project that aims to provide a full-fledged Social Sciences and Humanities Open Cloud (SSHOC) where data, tools, and training are available and accessible for users of SSH data. It will encourage secure environments for sharing and using sensitive and confidential data, will contribute to the Open Science agenda and the realisation of the European Open Science Cloud (EOSC). The intention is to create a European open cloud ecosystem for social sciences and humanities, consisting of an infrastructural and human component.

In this context, the project has a dedicated work package on fostering communities, empowering users and building expertise. Among this activities, SSHOC will identify existing, but will also develop new training materials and online learning paths. Materials developed in the task will be hosted open access in one of the existing platforms/training hubs. Together with the identified existing material externally hosted, the materials will be registered in and discoverable through the SSH Open Marketplace. The project will also build a Training Network to serve its target groups, which will be adopting a train-the-trainer approach and will be complemented by the development of a SSHOC train-the-trainer Toolkit.

**Life Sciences**

EOSC-LIFE[29] brings together the thirteen Biological and Medical ESFRI research infrastructures (BMS RIs) to create an open collaborative space for digital biology[30]. It started on March 1st, 2019 and is one of the EOSC cluster projects. The objectives of the EOSC Life Training WP is to develop and deliver training to enable effective data access and preservation for immediate and future sharing and re-use of data in the Biological and Medical Sciences. The training will address two target communities with distinct needs and roles: (i) the BMS Research Infrastructures (data) staff and (ii) the end-user of the (data) services offered.

## 3.4. Developments on the broader skills landscape

### 3.4.1. European stakeholder organisations' skills strategies – further examples

There are numerous European and other international initiatives whose skills development strategies are relevant to this report, and deliverable D7.1 provided a broad overview. In this sub-section, we describe one further organisational initiative from LIBER, which represents research libraries in Europe. The aim is to highlight the broad interest in developing resources for skills and capability development, and the need for EOSC governance bodies to include these (and other) organisations in implementing the EOSC strategy for skills and capability development.

LIBER (Ligue des Bibliothèques Européennes de Recherche) is a leading voice of Europe's research library community. LIBER aims to help research library professionals to stay at the cutting edge of their field and to lead their organisations. In order to do this, LIBER uses a number of delivery methods:

- Emerging Leaders Programme: an ambitious and international programme, designed to shape the next generation of senior leaders in European research libraries. The fifth edition of the Programme will start at LIBER's 2019 Annual Conference in Dublin;

- Journées Programme: An exclusive leadership programme for Library Directors;

- Annual Conference: Europe's biggest yearly gathering of research library professionals;

- Leading/participating in training activities in collaboration with international projects and initiatives;

- Organising a wide range of webinars on important topics, such as Research Data Management, Copyright, Open Access, etc.

---

[28] https://www.sshopencloud.eu/
[29] https://cordis.europa.eu/project/rcn/219199/factsheet/en
[30] https://www.esfri.eu/sites/default/files/Andrew%20Smith.pdf

LIBER's Board, Committees and Working Groups regularly produce factsheets, case studies and other resources.

LIBER has two dedicated working groups looking at areas relevant to EOSC's skills requirements:

- The Research Data Management working group collects good practices and lessons learned in the area of Research Data Management (RDM) in libraries. It also collaborates with other initiatives to evaluate and support skills development. At the moment, this working group has three main priority areas on publishing an annotated library of Data Management Plans, running webinars on key Research Data Management topics and producing a FAIR Data Factsheet. Examples of relevant webinars include TU Delft's Data Stewardship Project, GDPR and what it means for Researchers, Turning FAIR data into reality and Supporting Data Literacy.

- The working group on Digital Skills for library staff and researchers has two main priorities at the moment: developing an educational programme which will develop the digital skills of library staff members and researchers, and holding a workshop at the LIBER 2019 Conference focusing on Open Science skills, in collaboration with the Digital Humanities & Digital Cultural Heritage Working Group and the SSHOC project. Fostering collaboration with EU funded projects and other initiatives[31] is at the core of this working group.

The GO-FAIR initiative has also been attempting to coordinate European-level action on training. In GO-FAIR there are three GO-TRAIN implementation networks (INs). The Training Frameworks IN[32] is focusing on building skills capacity by developing and supporting the implementation of a train-the-trainer programme for FAIR data stewardship and a mechanism for endorsing training programmes for FAIR data stewardship. This is closely linked to the FAIR Data Stewardship Curriculum IN and the FAIR Seasons Schools IN.

### 3.4.2. One nationally coordinated approach with a focus on skills

In the Netherlands there have been many initiatives for installing data stewards in the universities, such as at the Delft University and Utrecht University. These data stewards meet regularly in person, and share news using email distribution lists. The national strategy is encapsulated in the National Platform Open Science[33] and the National Plan Open Science that was published in 2017[34]. The National Plan Open Science contains several references to data stewardship. Regarding adapting evaluation and award systems to bring them into line with the objectives of open science (reward systems) the following can be found in the report: "*In the present evaluation and reward systems the emphasis is often on the number of publications in prestigious journals with a high impact factor, often produced by well-established publishers and to which there is no open access, thus maintaining the culture of "publish or perish". Open science invites a broader set of evaluation criteria than just research output and research quality, including, for example, the quality of education, valorisation, leadership and good data stewardship.*" (p. 10) In this plan several ambitions can be found that relate to data stewardship e.g. '*ambition: to involve researchers in the transition towards open science*' (p. 27) where it is stated that further research is needed on the costs for building and maintaining infrastructure; the costs for training and deploying data stewards.

The National Coordination Point Research Data Management[35] has a pool of experts that form task groups. There has been a task group on data stewardship from October 2018 to February 2019. It has found that there is no uniform approach in the Netherlands, although many institutions provide tools, infrastructure and support to researchers with data. This support comes from a variety of professionals, with little

---

[31] For examples on collaboration, see: https://libereurope.eu/blog/2018/07/12/digital-skills-liber-2018/; https://libereurope.eu/blog/2019/03/04/liber-webinar-open-science-inspiring-cultural-change-in-your-library/; https://libereurope.eu/blog/2019/02/21/eoscpilot-liber-webinar-ethical-issues-skills-in-the-european-open-science-cloud/;

[32] https://osf.io/q98wt/

[33] https://www.openscience.nl/en

[34] OCW (2017, 9 February). National Plan Open Science. DOI: 10.4233/uuid:9e9fa82e-06c1-4d0d-9e20-5620259a6c65 https://www.openscience.nl/files/openscience/2019-02/nationalplanopenscience_en.pdf

[35] https://www.lcrdm.nl/en/

consensus about appropriate job profiles. Similar positions have different names: data stewards, data librarians, data managers, data officers, data consultants, data coordinators etc. To tackle these issues, the National Coordination Point Research Data Management (LCRDM) has initiated a nationwide survey to gain insight in what Dutch research institutions require, currently implement and what they are developing in the field of research data support. This survey aims to provide a better understanding of the current status quo and the data skills needed to professionalize the role and function of people working in data management. For the purpose of this survey, the range of different tasks in the data support field has been collected under the umbrella concept of data stewardship[36]. The results of the survey are not yet published.

### 3.4.3. Growing international coordination

Recently several organisations that also operate outside Europe have put emphasis on skills strategies. The 'Top 10 FAIR Data & Software Things Global Sprint' was organised on 29-30 November 2018. This global initiative was organised to develop resources by discipline and/or themes.[37] The discipline specific results can be found on Zenodo.[38]

The Carpentries provide coding and data science skills worldwide[39]. Carpentries instructors are addressing FAIR data and software and the skills that are associated in numerous training sessions such as at the RDA Plenary in Berlin. [40]

OECD's global science forum (GSF) has convened an expert group on digital skills for data intensive science[41] which is drawing on the framework produced by EOSCpilot amongst other things. It is due to report to GSF by Q2 2020.

---

[36] https://www.edugroepen.nl/sites/RDM_platform/LCRDM/Wikipages/Taakgroep%20Data%20Stewardship.aspx
[37] The global sprint was organised by the Library Carpentry, Australian Research Data Commons and the Research Data Alliance Libraries for Research Data Interest Group in collaboration with FOSTER Open Science, OpenAire, RDA Europe, Data Management Training Clearinghouse, California Digital Library, Dryad, AARNet, Center for Digital Scholarship at the Leiden University, and DANS.
[38] https://doi.org/10.5281/zenodo.2555497
[39] https://carpentries.org/
[40] https://uc3.cdlib.org/2018/07/24/a-carpentries-based-approach-to-teaching-fair-data-and-software-principles /
[41] https://www.innovationpolicyplatform.org/digital-skills-data-intensive-science-oecd-project

# 4. DEVELOPING SKILLS AND CAPABILITIES

## 4.1. Addressing skills and capability gaps

The project identified gaps in skills and capabilities through desk research and workshop activity between April 2017 and March 2018, reported in D7.1, D7.2 and D7.3. Key points of this activity included:

- A requirements workshop in May 2017, among the project partners.

- EOSCpilot-OpenAIRE joint workshop in Berlin (24 October 2017) identifying 'soft' skills and policy-oriented areas of knowledge.

- The Science Demonstrator projects, which piloted the application of EOSCpilot project partners' services towards scientific goals. From their reports on these activities and liaison with their 'shepherds' (project technical representatives), we derived skills gaps mostly of a technical nature from the challenges experienced.

- Further workshops to consult skills stakeholders from research institutions and infrastructures, at the EOSC Stakeholders Forum (2017, 2018), DI4R18, DI4R19, four RDA Plenaries, EUDAT2018, IDCC18, and several EDISON project workshops.

In D7.3 we summarised the identified skills gaps shown in Table 2.

EOSCpilot WP7 workshop effort was also partly about delivering training to address these gaps. This was primarily delivered through the GridKa School 2018 at KIT, which focused mostly on the more technical gaps. Several webinars were also provided, on digital skills development including the FAIRS4S Skills Framework, and on ethics in the EOSC context. These involved LIBER with DCC and DANS.

**Table 2: Skills gap analysis**

|  | Users' skills | Service providers/ operators' skills |
|---|---|---|
| **Abstract topics** | • FAIR and OA policy<br>• Data policy legal and funder requirements<br>• From traditional to cloud platform<br>• Pipelines (workflows for reproducibility)<br>• Research strategy and open research potential<br>• Research ethics | • AAI integration, management (security best practices)<br>• Personnel and skills development |
| **Specific technologies** | • IaaS basics; Linux administration<br>• Containers<br>• PaaS tools/services (inc. workflow tools)<br>• EOSC service portfolio (comparative)<br>• Jupyter<br>• FAIR metadata for interoperability<br>• Tools and domain standards | • OpenStack<br>• EOSC user support tools<br>• EOSC integration (minimum requirements)<br>• Secure storage integration |

### 4.1.1. GridKa School

The training provided in WP7 was primarily delivered through the GridKa School 2018, which we co-organised with the aim of addressing the competence and skills gaps. KIT, a partner of WP7, organises the Gridka School, a summer school held annually and traditionally covering topics about data management and analytics, Cloud

Computing, and modern programming, with special focus on applicability for distributed systems and Big Data. The GridKa School is well-established and one of Europe's largest summer schools in the context of Scientific Computing. It has evolved from an initial event to introduce Grid technology to the High Energy Physics community to encompass many learning objectives and topics targeting a large set of groups.

The programme of the GridKa School especially targets new technologies, tools and trends. It has a track record of attracting researchers from domains such as High Energy Physics, Astro-particle Physics, Environmental Science, Bioinformatics, and Informatics - data-intensive domains that were well represented in the EOSCpilot Science Demonstrators. This made the audience a suitable target for training on the skills gaps identified in EOSCpilot, and for testing the relevance to this audience of the broader range of topics in the Skills Framework. GridKa School participants range from graduate and PhD students, to researchers who are expected to be EOSC clients, whether as EOSC end-users, EOSC system managers, or EOSC suppliers.

The programme of the school was scoped so that for each of the identified gaps at least one training was offered, either in form of a lecture or hands-on tutorial. For most of the gaps we offered several trainings, e.g. for "Domain knowledge and software development to adapt workflows to cloud workflows" we provided eight lectures and nine tutorials focusing on different aspects of the identified skills gap.  Some of the lectures specifically targeted the adaptation of workflows to cloud workflows, e.g. 'ViCE: From desktop to cloud or HPC – Virtualized Research Environments' or 'Boosting productivity of researches integrating IT services into an interactive analysis platform'. Others focused on the software development for and in cloud environments, e.g. 'Scalable Scientific Analysis in Python using Pandas and Dask' or 'Collaborative Software Development.

The alignment of the programme to the identified gaps proved the applicability of the Framework in focusing and balancing a programme of relevant training for Data Scientists and Stewards. We evaluated the alignment by assessing the overall perceived relevance of the programme with the participants. The evaluation showed (see D7.4, Figure 6 for details) that lectures and tutorials were perceived highly relevant by the participants. There is some overlap between the topics in the EOSCpilot Skills Frameworks and those that GridKa School trainees identified as their own learning objectives. Of 59 topics listed in the Skills Framework, those that attracted the broadest interest from the GridKa School attendees were:

- Software prototyping.
- Math and statistical knowledge application.
- Machine learning methods.
- Data preparation and documentation for reproducibility.
- Critical thinking and theory building.
- Creative problem solving and flexibility.
- Cloud environment and storage management.

A small number of topics that were previously identified in EOSCpilot as gaps e.g. from Science Demonstrators and consultations, were identified as learning objectives by very few GridKa School attendees. These included ethical and legal compliance, data preservation topics, and some aspects of service management. This does not mean that these skills are not relevant, but that trainees assessment of their skills needs are not the only criterion for skills acquisition, and their awareness of these topics needs to be increased.

### 4.1.2. Webinars

Further discussion of the competence gaps and the plans of the EOSC-hub and OpenAIRE+ projects to address some of them led to agreement in WP7 to focus efforts on addressing remaining gaps through webinars. Two were held in the last quarter of the project; the first being on skills development and how the EOSCpilot skills framework can help address this in a specific group of research support staff. The second webinar was on ethics, a topic whose importance had been made clear by many stakeholders but which had received little attention from a skills perspective. Both webinars were delivered through WP7 partner LIBER's webinar platform but were available to any attendee, not just those in LIBER's member network.

**Skills strategy webinar**

This webinar took place on 14th December 2018 and brought together an existing LIBER working group on digital skills for librarians with the creators of the FAIR4S framework in EOSCpilot. Panellists addressed the role of libraries in fostering open science skills and the ways in which the framework could assist. There were 74 attendees, and the slides and recordings have been made available for later viewing[42].

**Ethics webinar**

This webinar took place on 18th March 2019 and served to promulgate both the ethics-related recommendations produced by EOSCpilot WP3 and the skills-related implications for research-producing organisations. There were 123 attendees; slides and recordings are available for later viewing[43].

## 4.2. Consultative workshops to identify priorities for skills development

### 4.2.1. Workshops at stakeholder events

Workshops conducted throughout the project identified a number of consistent priorities from stakeholders for skills development in EOSC. They are described in detail in deliverable D7.4, and we summarise the key points here. Three priorities had consistent support:

1. Support train-the-trainer approaches.
2. Catalogue existing provisioning of training.
3. Define minimum standards for training.

One attracted more diverse views, with strong opinions for and against. This was that EOSC should commission the development of material to fill gaps identified in training provision. There was also consistent support for recommendations on skills development that EOSCpilot puts forward:

- That EOSC should offer information on training across Europe via a 'marketplace' of IT and soft services, and that this should be done through a single point for EOSC, rather than just through project-specific web sites.

- That a global perspective is needed for skills development, just as it is for research. Therefore, EOSC should include skills-related issues in its global engagement with similar infrastructure initiatives.

- That EOSC should promote standards that aid description, discovery and reuse of learning resources and that compliance with these standards should be a condition for inclusion in the EOSC training registry.

### 4.2.2. Open Science Skills Workshop in collaboration with 4TU Research Data Service

The September 2018 workshop at TU Delft in the Netherlands 'Time for Open Science Skills to Count in Research', was carried out in collaboration with data stewards, particularly the team at TU Delft, from the 4TU Research Data Service. This service is shared by Delft and three other Technical Universities in the Netherlands.  The TU Delft setting was useful as a testbed of the framework, as the team has been a leading exponent of an approach to embedding early career researchers with disciplinary data skills in relevant faculties, and coordinating their professional development as data stewards through the research library.

The workshop involved researchers and professional support staff from a number of institutions - primarily in the Netherlands and Belgium. It was motivated by shared interest of the TU Delft team and the EOSCpilot WP7 team in matching skills to researcher career stages, and in linking skills to the open science behaviours that were identified as desirable by a working group of the Open Science Policy Platform (OSPP). The OSPP Rewards WG had published a draft Open Science Careers Assessment Matrix (OS-CAM) that was used a source for the EOSCpilot framework in D7.3.

---

[42] Youtube recording: https://www.youtube.com/watch?v=2GZ7sy-oX2c; slides: https://zenodo.org/record/2247650
[43] Youtube recording: https://www.youtube.com/watch?v=ArlYY3r1Qlg; slides: https://zenodo.org/record/2597309

The workshop offered an opportunity to further focus the skills listed in the EOSCpilot framework, by identifying those considered a priority by researchers, data stewards and other support staff in breakout sessions. Breakout groups were split according to each of the four stages in the European Research Careers Framework[44], which identifies the following standard terms:

- R1 Up to the point of PhD.
- R2 PhD holders/equivalent not yet fully independent.
- R3 Established researcher.
- R4 Leading researcher.

In preparation for the workshop the TU Delft data steward team selected nine key skills for researchers from the 59 listed in the EOSCpilot D7.3 Annex. The group activity for each of the groups was divided into two parts. In the first part, the groups were asked to shortlist a maximum of 4 skills most relevant to their respective career level (R1-R4). Subsequently, for each skill, they wrote down on post-it notes:

- Why is this skill relevant to researchers at this career level?
- What would be the evidence that researchers at this career level have these skills and can apply it in practice? (in other words: what does a person applying the skill do?)
- What support (from support staff, service providers) will researchers need to apply this skill?

The responses to these questions and the groups' selection of skills have informed the revisions to the framework, described below, after the consultative survey.

Lively and productive discussion was preceded by opening talks from Anne de Vries (PhD students Network Netherlands), Bartel Van de Walle (TU Delft) Rinze Benedictus (UMC Utrecht), and Simon Kerridge (University of Kent and CASRAI). These set the scene for the discussion, framing the need for PhD students to acquire skills in open science and FAIR practice in terms of the need for universities to change the reward culture. Anne de Vries introduced work done by the PhD network and broader EURODOC organisation to identify skills. Bartel Van de Walle advocated for universities to move from measuring publication impact towards evaluating research practices from a more holistic view of the research cycle. Rinze Benedictus described steps taken by UMC Utrecht to do this, and Simon Kerridge introduced the CreDiT taxonomy.

This taxonomy is a standard being developed by CASRAI to describe authorship roles in such a way that these can be credited through research evaluation systems. Submitting authors of research publications can identify which authors have contributed to any of fourteen tasks spanning the research cycle, e.g. writing, supervision, review, data analysis, project management etc.  The closing talks by Kevin Ashley and Annnette Björnsson reflected on the current initiatives within the European Commission towards changing academic rewards, highlighted the importance of OSPP working group and their outputs including the Rewards WG 'Open Science Careers Assessment Matrix', which would be the subject of further work in light of the sister report on Next Generation Metrics[45]. She also highlighted the parallel work of the Skills Working Group on Evaluation of Research Careers fully acknowledging Open Science Practices[46], and the FAIR Data Expert Group on Turning FAIR Data into reality[47].  A full report of the workshop is available in blog posts on the TU Delft site[48].

---

[44] https://euraxess.ec.europa.eu/europe/career-development/training-researchers/research-profiles-descriptors

[45] https://ec.europa.eu/research/openscience/pdf/report.pdf

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

[47] https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283

[48] https://openworking.wordpress.com/2018/12/03/its-time-for-open-science-skills-to-count-in-academic-careers-part-1-talks

## 4.3.  The FAIR4S Skills and Capability Framework

### 4.3.1.  Introduction and recent changes to the framework

The first version of the FAIR4S Framework was published in EOSCpilot D7.3 Skills and Capability Framework for the EOSC. The update described below and in the Annexes to this report follows a pilot application of the framework in a September 2018 workshop intended for its formative evaluation and further development, plus a consultative survey conducted between October and December 2018. The workshop is described below as motivation for the changes introduced to the framework. The name 'FAIR4S' is an abbreviation of 'FAIR Stewardship Skills for Science and Scholarship' and was adopted after D7.3 was published to help communicate the draft framework to stakeholders, who had been consulted on it.

The FAIR4S framework has a similar target community to that of the EOSCpilot as a whole i.e. publicly funded researchers and the professional groups that support them in the stewardship of research outputs are the intended beneficiaries of skills development. The framework itself is intended for use by two main target groups:

- Research Producing Organisations and the coordinators of their RDM services responsible for developing skills of data stewards and providing data management training and support to research staff, generally of a generic cross-domain nature and targeting culture change, policy awareness and compliance, and knowledge of ICT support services.

- Research Infrastructures and e-Infrastructures and their coordinators of training for researchers and support staff, generally on domain-specific data management or on the technical skills required to deploy networking and cloud compute services that enable data management.

For coordinators of data stewardship skills development and training, or research and support staff looking to improve their own skills, and for learning resource providers supporting either group, the main use cases are:

1.  To plan a professional development course or curriculum, by considering suitable topics for specifically targeted roles and for the level of expertise they need, taking into account:
    a.  the skills, knowledge and attitudinal requirements and desirable learning objectives relating to the competence topic;
    b.  the local availability of other roles that may have (or require) related competences;
    c.  capabilities the organisation intends to develop, and services it plans to use to enhance them.
2.  To tag items in a knowledge base or registry of learning resources, or use such a registry to find learning resources that match a selected topic, also taking into account the level of expertise the resource aims to cater for, and the roles targeted.

There are at least two other groups to whom the Framework is relevant who are not part of that more general audience. The first such group is Human Resource (HR) professionals, who are involved in specifying job descriptions, in recruitment against those job descriptions, and in the processes for staff assessment for reward and promotion. The second such group is independent training providers who can provide courses and other learning resources that allow people to develop the skills described in the Framework.

Many of these organisations are SMEs and the training services and resources they provide need to visible in EOSC's service catalogue along with those from universities, research organisations and EOSC infrastructure service providers who are already engaged in EOSCpilot's work more broadly.

In the next section we describe the reshaped FAIR4S framework, followed by the consultation survey and a summary of the changes made in response to it.

### 4.3.2.  Revised FAIR4S framework

FAIR4S draws on pre-existing competence frameworks developed for data literacy, research data management, and for data science.  These four in particular have shaped the structure and content:

1. A Pilot Competency Matrix for Data Management Skills: A Step toward the Development of Systematic Data Information Literacy Programs[49].
2. Research Data Alliance Interest Group on Education & Training in Data Handling (ETDH-IG) Task Force on Defining data handling related competences and skills for different groups of professions[50].
3. EDISON Data Science Competence Framework v.09, and v10[51].
4. Open Science Careers Assessment Matrix (OS-CAM)[52].

The fourth item above is not a competence framework but, as a list of OSPP recommended behaviours and indicators of open science activities researchers should undertake, it offered a very useful basis to sharpen the focus of the Framework on 'core' skills that support these activities. The Delft workshop described in Section 4.2.2 took further the initial steps made towards this in D7.3.

**Professional group roles targeted by FAIR4S**

FAIR4S aims to define competences relevant to the training of professional data stewards as a distinct role that is emerging in response to the challenges identified in Section 2, which also intended to help develop and reward stewardship skills for researchers and support staff in other roles, who will share responsibility for stewardship. FAIR4S treats stewardship as an emergent data advisory role whose competences overlap between these established professional support roles, as shown in Figure 3.
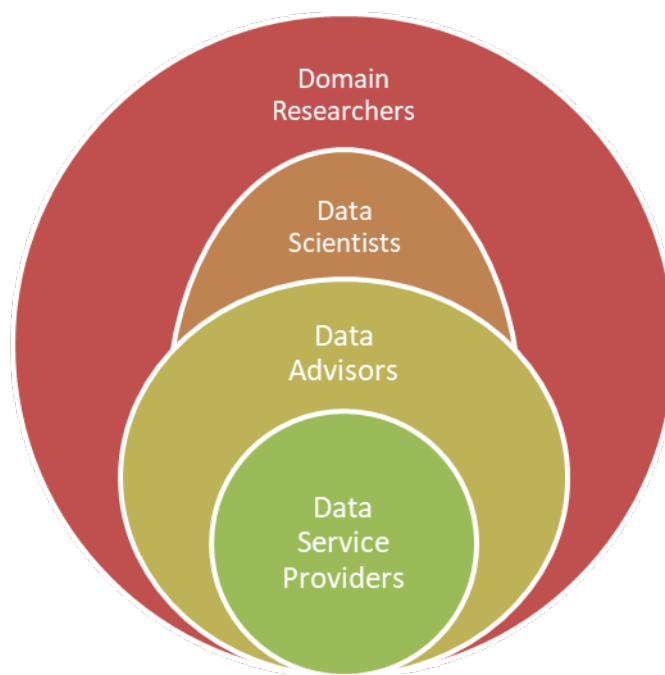


**Figure 3: Professional group roles represented in FAIR4S**

**Skills groups described by FAIR4S**

The starting point – what the Framework assumes as the basic level knowledge needed to work in the EOSC environment – is a core group of data handling competences identified for undergraduate-level information professionals, and collectively termed 'digital information literacy' skills[49]. These core competences have been mapped to the six 'skills groups' shown in the inner circle of Figure 4.

---

[49] Purdue University, & Sapp Nelson, M. (2017). Journal of EScience Librarianship, e1096. https://doi.org/10.7191/jeslib.2017.1096
[50] Molloy, L. Demchenko, Y, Jung, C. et al. (2016) https://www.rd-alliance.org/group/education-and-training-handling-research-data-ig/wiki/task-force-defining-data-handling
[51] Demchenko, Y. Belloum,A. and Witkorski, T, (2016, 2018) https://github.com/EDISONcommunity/EDSF
[52] OSPP Working Group on Rewards Under Open Science (2018) Evaluation of Research Careers fully acknowledging Open Science Practices https://ec.europa.eu/research/openscience/index.cfm?pg=rewards_wg

Certain groups of skills will be applied consistently across a number of research projects because they relate to common infrastructure services. An example is the ability benefit from single sign-on to EOSC services through an underlying AAI (authentication and authorization infrastructure) service. FAIR4S differentiates between such generic capabilities that will normally be applied across projects from those more likely to be applied in a project-specific manner. Figure 4 introduces these cross-project skills groups, and Table 3 describes the scope of these and the core skills groups.

Figure 4 shows these two types of skills: the six in the inner circle follow a research data lifecycle, and represent the competences and capabilities applied in project- specific ways. They are adapted from lifecycle models commonly used in research data management training[53 & 54]. The three in the outer circle represent those that are applied consistently across multiple research projects. These are adapted from RISE, a capability framework for research data service development based partly on the CoreTrustSeal standard.



Figure 4: Core Skills Groups

**Key Skills**

Table 3 describes the scope of each skills group in more detail. The groups each include five or six competence/capability areas that skills are applied to, and at least one 'key skill'. Key skills are those that are a priority for at least one of the roles, and there is at least one key skill in each group.

**Three dimensions of skills and capability development**

FAIR4S considers skills and capabilities to have three dimensions: expertise, organisation, and responsibility.

The expertise dimension has three levels[55]:

- Basic - comprehend knowledge and, attitudes or aptitudes needed to apply the topic.

---

[53] University of Edinburgh (2012) Research Data MANTRA Training https://fr.slideshare.net/edinadocumentationofficer/mantra-poster2

[54] Research Data Netherlands (2017) Essentials 4 Data Support and University of Bath https://datasupport.researchdata.nl/en/start-the-course/i-definitions/research-lifecycle/

[55] Derived from Bloom, B. S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H.; Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: David McKay Company.

- Intermediate - demonstrable ability to apply the topic.
- Expert - acquisition of sufficient experience to advance knowledge on the topic.

The 'organisation' dimension has three levels:

- Individual - competences acquired and applied by individuals through professional development training and work experience.
- Team - competences applied through the capabilities of a research team or organisational unit, such as department.
- Organisation - competences applied through the capabilities of an organisation-wide function.

The terms 'capability' and 'competence' are sometimes used synonymously in skills development, in FAIR4S the intention is to link skills development to organisational culture change and service deployment. Using 'competence' to refer to practices of individuals and 'capability' to refer to collective practices of organisations makes sense in this context. It is also consistent with the practice of applying 'capability models' to assess practices that organisations carry out or want to improve on, and 'competence models' to assess practices of individuals, or to improve their deployment across roles.

**Table 3: Skills Groups and Key Skills**

**Plan and design:** Planning and design of data, research software and other outputs, including the associated documentation. This will include all relevant steps including identifying requirements of research output users, the organisation and research funders, establishing effective approaches to meet their requirements, then reviewing this planning.

**Key skill: Plan stewardship and sharing of FAIR outputs**. Related skills: Model data structures and define database needs; Specify metadata and persistent identifier standards; Analyse requirements for services or software; Appraise and select repositories for FAIR sharing

**Capture and process:** Capturing and processing of data or related materials to enable research evidence to be prepared for analysis; provisioning of secure managed access to networked storage, scalable to meet demands, plus resources, tools, standards and workflows for collaboration between research team members, and relevant third parties

**Key Skill: Reuse data from existing sources.** Related skills: Manage databases, Software prototyping, Set up and document workflows, File naming and organisation, Data cleaning, processing and software versioning

**Integrate and analyse:** Developing and applying appropriate methods to enable lines of enquiry to be formulated and pursued towards the research objectives, by assembling and integrating selected data, software, systems, or other resources, and enabling relevant knowledge and techniques to be applied in their analysis and transformation into research outputs.

**Key Skill**: **Use or develop FAIR research tools or services.** Related skills: Critical thinking and theory building; Creative problem solving, flexibility; Math and statistical knowledge application; Data transformation and integration; Data mining, querying, interpretation; Predictive modelling and analytics; Machine learning methods

**Appraise and preserve:** Developing and applying appropriate methods to appraise research outputs for their compliance with ethical, FAIR and research integrity principles, their value to the organisation and to research reproducibility, and their potential to serve new purposes or communities; planning and taking action to mitigate risks to long-term access for further appraisal.

**Key Skill: Prepare and document data/code to make outputs FAIR.** Related skills: Data quality assurance using open standards; Assess FAIRness and check policy compliance; Data transfer and long-term storage; Format and media migration; Software review and preservation

**Publish and release:** Describing research products and their inter-relationships, providing access to meet the needs of their providers, users, and other stakeholders, in order to maintain or enhance their value and comply with ethical, FAIR and research integrity principles and policies.

**Key Skill: Publish FAIR outputs on recommended repositories.** Related skills: Access control and management; Ethical application of patents, licences; Open access publishing and self-archiving; Documentation for public use, lay understanding; Engaging in open innovation beyond academia

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

**Key Skill: Recognise, cite and acknowledge contributions.** Related skills: Metadata and persistent id. Exposure; Visualisation and presentation of results; Evaluation of repository and publishing platforms; Searching repositories and scientific databases; Vocabulary/ ontology application

**Govern and assess:** Developing and maintaining the organisation's strategies, policies, and processes on FAIR/open research outputs, and associated documents and processes that enable these to be implemented, and relevant laws or regulations to be complied with. Continually reviewing these strategies, policies and processes through stakeholder consultation, communication, and impact monitoring

**Key Skills: Apply policies to comply with legal, ethical and FAIR principles**; Develop open research strategy and vision. Related skills: Research integrity, attribution, impact awareness; Information security and risk management; Data governance, handling third-party data; Storage security management; Contributing to quality assessment or peer review

**Scope and resource:** Identifying the scope of research data services and stewardship activities and securing the resources to sustain these. Continually reviewing the business case considering the service value propositions, processes, and relevant costs and benefits, taking into account governance processes and timelines, and the need for cost recovery mechanisms to comply with funder requirements.

**Key Skill: Secure funding for open science /support**; Related skills: Perseverance delivering diverse open research projects; Service level management; Change management; Workflow set-up and provenance information mgmt.; Cloud environment and storage management; Authentication and authorisation (AAI) management; Costing of data management and preservation.

**Advise and enable:** Management of services that enable data stewardship and open research, and the online or face-to-face training and support that service customers need to make effective use of them. This will include interacting with relevant professional service units, managing stakeholder relationships, and making continuous improvements to these.

**Key Skill: Lead good practice by example**; Related skills: Engaging with research users and stakeholders; Developing a profile of open research; Training in open methods, services; Contributing to education, professional development; Contributing to open research, networks, standards bodies; Building open inter-disciplinary collaborations.

The 'responsibility' dimension corresponds to the four career levels defined by the European Research Careers Framework:

- R1 Up to the point of PhD.
- R2 PhD holders/equivalent not yet fully independent.
- R3 Established researcher.
- R4 Leading researcher.

The first version of the model proposed in D7.3 described the 'responsibility' dimension in terms of seven levels, inspired by the SFIA model 'Skills Framework for the Information Age'. These responsibility levels were not directly mapped to any of the professional support group roles in the framework, on the basis that this mapping would depend on the organisational and disciplinary context. Our consultation survey respondents found this confusing however. Therefore, the dimension is retained, but the seven levels are now replaced with the four research career levels. Although these four levels do not apply to data stewards or other research support roles, we comment in the conclusions on the need to further define responsibility levels for these roles.

**Role profiles - matching key skills to roles**

For educators or trainers planning professional development in data stewardship, Role Profiles help to identify the appropriate key skills to focus on. Each role profile identifies three or four key skills for the role and, for each, answers the following questions:

1. Why is this a key skill for this role? The profile identifies competences that the role would be expected to demonstrate.
2. What level of expertise is expected? The profile specifies one of the three levels: basic, intermediate, or expert.
3. Who may help? The profile identifies which other roles could offer support in applying the competences.
4. Where can you learn? An illustrative selection of training resources are listed, e.g. from CODATA-RDA Data Science Summer School, and a variety of other sources.

At the end of this section examples are given of the Role Profile and Skills Profile formats. Role profiles are also intended to cover the following groups:

- Researcher:
  - Career levels R1 to R4*.
- Data scientist:
  - Career levels R1 to R4*.
- Data advisor - 'front-office' specialisations
  - Data steward*;
  - Research manager;
  - User support, training and outreach;
  - Ethics and data protection advisor;
  - Commercialisation advisor.
- Data service provider - 'back office' specialisations:
  - Data librarian / institutional level data steward*;
  - Service manager/ project manager;
  - Research software engineer;
  - Data service architect;
  - Archivist.

Annex B (FAIR4S Role and Skill Profiles and Tables) provides examples of those marked '*', and also gives summary tables mapping each role to the expected expertise level on each skill.

The role profiles draw on various sources as well as the competence frameworks already mentioned - including the EDISON Professional Profiles[56] and the Life sciences data steward function matrix[57].

**Skill profiles - linking competences to capabilities and services**

When applying FAIR4S it may be useful and sometimes necessary to consider the capabilities the team and organisation has already, or wants to develop. These in turn may be influenced by the services it has or wants to develop. Skill profiles describe the key skills identified in Table 3 and link them to this broader context. Each Skill Profile answers the following questions:

1. What knowledge, skills, and attitudes/aptitudes are needed?
2. What counts as success in applying the skill?
3. Whose roles apply this skill?
4. What related competences should the research team have?
5. What capabilities and services would amplify this skill?

Profiles cover each of the key skills identified in Table 3, and Annex B provides examples of three of these.

---

[56] https://github.com/EDISONcommunity/EDSF
[57] https://doi.org/10.5281/zenodo.2561723

FAIR4S uses the term capability to refer to the collective competence of a group, as mentioned earlier. The group can be a research team, an organizational level, or a third party. We make the distinction because the group as a whole is more than the sum of its parts; we cannot simply say that its competence is the sum of competences of its members. There is, however, a strong relationship between them and capabilities and competences share the same content at the top level of the Framework (skills groups).

Organisations can amplify their capabilities by using or providing services. Skill profiles aim to help relate competences to the organisational dimension and to the categories of service that EOSC provides. These categories are broad, and ideally, organisations will use these profiles as templates, replacing broad categories with locally relevant services. Table 4 shows service categories based on the EOSC Portal and EC Staff Working Document - Implementation Roadmap for the EOSC[58].

**Table 4: EOSC Service Categories**

| Service category | Description |
|---|---|
| 1. Service catalogue<br>2. Security and operations | A unique identification and authentication service and an access point and routing system towards the resources of the EOSC. |
| 3. Networking<br>4. Compute<br>5. Sharing and discovery | A protected and personalised work environment/space (e.g. logbook, settings, compliance record and pending issues). |
| 6. Training and support | Access to relevant service information (status of the EOSC, list of federated data infrastructures, policy-related information, description of the compliance framework) and to specific guidelines (how to make data FAIR, to certify a repository or service, to procure joint services). |
| 7. Data management<br>8. Processing and analysis | Services to find, access, re-use and analyse research data generated by others, accessible through appropriate catalogues of datasets and data services (e.g. analytics, fusion, mining, processing). |
| 9. Storage | Services to make their own data FAIR, to store them and ensure long-term preservation. |

**Examples of Role and Skill Profiles**

Table 5 and Table 6 provide examples of profiles, and Annex B provides further examples.

**Table 5: Role Profile for Data Steward**

| Role Profile | Data Steward  - Operational level e.g. research group or department/ faculty | | |
|---|---|---|---|
| **Key skills** | | **Level** | **Who may help?** |
| **Plan for the stewardship and sharing of FAIR outputs** | | Expert | Data Librarian |
| You have in-depth understanding of your organisation's RDM policy, and of the policies of relevant research funders in relation to planning for FAIR outputs, including Data Management Plans. You have in-depth knowledge of the legal and ethical frameworks applicable to the research communities you support. You are also familiar with these communities' research methods, tools, standards and services used locally.  You are able to translate this knowledge and understanding into Data Management Plan (DMP) templates for the researchers you support, providing expert | | | |

---

[58] http://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf

advice appropriate to all career stages.  You can articulate how planning benefits the researcher as the 'first reuser' of their outputs, and seek help where needed from other professional services**.**

| Use or develop open research tools or services | Intermediate | Data Librarian<br>Research Software Engineer<br>Data Service Architect |
|---|---|---|

Aware of relevant (inter)national data and e-infrastructure services you identify and respond to the needs of researchers and stakeholders in the areas you support for FAIR tools, services or resources. You provide guidance and instruction on discovery, acquisition and (re-)use of data, including local collection and those in the public domain. You ensure these are as open possible to reuse, according to organisational policy and legal obligations. Liaising with relevant professional services, e.g. research software engineers and data service architects, you develop and implement standard solutions for recurring issues and deliver training on their effective application, also ensuring that the organisation's service management processes are followed.

| Prepare and document data/code to make outputs FAIR | Expert | Data Librarian<br>Research Software Engineer |
|---|---|---|

You provide researchers in fields you support with examples of FAIR and open data or software that advances research, and describe the relevance of FAIR principles to the norms for validation or reproducibility in their domains. You provide expert advice on making data findable, and on cleaning data to make it actionable and interoperable for reuse. Knowledgeable about ethical and commercial constraints on data access, you liaise with relevant professional services and systems.  You provide expert advice on appraising and selecting data of value for reuse, and on good practice in managing software code, including versioning and documentation.

| Publish FAIR outputs on recommended repositories | Expert | Data Librarian<br>Project/ Service Manager |
|---|---|---|

Using your in-depth knowledge of repositories for data, code or other outputs (such as samples or protocols) , you direct researchers in your area to these repositories and other services that are a good fit to their domain, the characteristics of their data, and their data management processes.  You provide expert advice to researchers in your area on data formats and vocabularies relevant to their research domains, and on the deposition processes of repositories, whether these are locally provided or external. You liaise with relevant professional services to ensure appropriate constraints on how FAIR the research outputs will be, depending e.g. on assessment of ethical impacts, data protection, or licensing and IP rights.

**Where can you learn?**
**Key Skill 1**
Basic level:
- Introduction to Research Data Management and Data Management Plans (CODATA)
- FAIR Sharing (CODATA)
- Managing and sharing research data (https://www.fosteropenscience.eu/node/2328)

Intermediate level:
- How to get your article published (https://www.egsh.eur.nl/doctoral-education/phd-course-guide/how-to-get-your-article-published/)

**Key Skill 3**
Basic level:
- Tutorial: Variable Detection and Linking in Social Sciences Publications (https://www.fosteropenscience.eu/node/2315)
- Introduction to Text and Data Mining (https://www.fosteropenscience.eu/node/2263)
- R (CODATA)
- Github (CODATA)

**Key Skill 5**

Basic level:
- Practicalities of RDM (CODATA)
- Visualization (CODATA)
- R (CODATA)

**Key Skill 6**
Basic level:
- FAIR sharing (CODATA)
- Practicalities of RDM (CODATA)
- Publishing to GitHub and Rstudio (CODATA)
- Be Persuasive: Write a Convincing Position Paper or Policy Advice
  (Project-Centered Course from Coursera, https://www.coursera.org/learn/persuasive-writing)

**Table 6: Skill Profile example**

| Skills Profile | 1 PLAN - Plan the stewardship and sharing of FAIR outputs |
|---|---|

*What knowledge, skills and attitudes or aptitudes are needed?*

**K- Basic:** Show awareness of FAIR principles, relevant data policies of funders or journals, and legal and ethical requirements. Understand why funders have these expectations, why scientific organisations and research institutions support them, and what should be covered in the Data Management Plan (DMP).
**S - Basic:** Describe guidelines from relevant funders, journals or other organisations representing your domain, and identify what data, software or other objects will be reused or created, and how in general terms the FAIR principles will be applied to them.
**A - All levels:** proactively give advice or seek it from appropriate colleagues, data advisors and data service providers.

**K - Intermediate:** Describe how the research, thesis or dissertation objectives will draw on sources of evidence or materials, the methods that will be applied, and how outputs can be made FAIR.
**S - Intermediate**: Following guidelines from relevant funders and domain organisations, describe how relevant standards and services will be used to make data or code findable, accessible, interoperable and reusable, and who is responsible for this.

**K - Expert:** Describe appropriate criteria to meet user and stakeholder expectations regarding implementation of FAIR principles, demonstrating in-depth understanding of data-driven research methods in your domain and the data sources, formats, and disciplinary norms applicable in your research field.
**S - Expert -** Determine appropriate criteria to meet stakeholder expectations regarding implementation of FAIR principles. initiate process for managing data, code, or other outputs across the research team. Develop templates and contribute to local guidance, aligning this with data policies and processes of the organisation and funders. Give support and supervision to team members, and contribute to support services through your peer network.

*What counts as success?*

Each research study has a Data Management Plan (or similar, e.g. Data Stewardship Plan, Software Management Plan, or Outputs Management Plan) produced and reviewed by the Principle Investigator, supervisor, and/or funder. It is referred to throughout the study it relates to, and updated if necessary. Updates include any significant changes in methods, standards, or services that have been or will be used to make outputs FAIR, or in responsibilities for these.

*Whose roles apply this skill?*                                  *(bold= directly involved, \* indicates a key skill for the role)*
Researchers:                 - **R1** \* - **R2** - **R3**\* - R4
Data scientists:             - **R1** - **R2** \* - **R3** \* - R4
Data advisors 'front-office': - **Data Steward** \* - **Research Manager** - **User support & Training/Outreach** \*
                             - **Ethics & Data Protection Advisor** - **Commercialisation Advisor**
Data service provider: - 'back office': - Data Librarian - **Service Manager/ project manager** - Research Software Engineer - Data Service Architect - Archivist

*What related competences should the research team have?*

| | |
|---|---|
| Application of policies to comply with legal requirements , ethical, and FAIR principles<br>Costing of data management and preservation<br>Modeling data structures and defining database needs | Specifying metadata and persistent id. standards<br>Analysing requirements for services or software<br>Appraising and selecting repositories for FAIR sharing<br>Training in open methods and services |

*What capabilities - and services - would amplify this skill?*

| | |
|---|---|
| **Prepare and document data/code to make outputs FAIR :** Supports research groups to determine and fulfil appropriate criteria to reuse, manage, and share FAIR research outputs, and fulfil these expectations according to domain norms and standards. | Data management |
| **Apply policies to comply with legal requirements , ethical & FAIR principles**: Develops local policies and guidelines for publishing research data and related objects, and for selecting repositories that comply with relevant regulatory and policy frameworks. | Training and Support |
| **Publish FAIR outputs on recommended repositories :** Provides access to repository functions to ensure preserved research objects remain FAIR for as long as required, enabling these functions to interoperate with other systems or processes that provide or extract information to maintain or enhance FAIRness of the research objects held. | Storage Sharing and discovery |

## FAIR4S from the service providers' perspective

The EOSC will not consist of a static set of services and infrastructure but will instead be subject to continual evolution. Services will be improved and altered, new services will appear and other services will be retired. Since many services are not used in isolation, the complexity of the service landscape and the possible combinations of services to achieve research goals will be significant. Some service changes will be incremental and require little or no change in the competences and capabilities required to use them. Others will be significant; service simplification may reduce the skills barrier for use, whereas some new innovative services or service combinations will require new skills to make effective use of them.

As the service landscape evolves, the research team or organization will need to consider what capabilities they want to improve, what competences are no longer required, whether the individuals involved have the relevant competences, and how the new services may augment their competences to bring new levels of skills. The Framework permits periodic reassessment of all these requirements using a common language, which will continue to be applicable as services evolve.

The current EOSC services landscape is heterogeneous in many aspects. These range from the variety of providers (and users), and differences in purpose, technology, costs and funding mechanisms to the disciplinary and geographical remit of the services. They can be regional/national or international, they can be e-infrastructure or research services, and they can be services that process, visualise, integrate, augment, transport, store and/or archive data.

Within the EOSC implementation roadmap, service classes are described in the context of "A possible EOSC Model". The model is currently under development through the EOSC-hub, FREYA, OpenAIRE-Advance and other H2020 projects. According to this model, In terms of architecture, the EOSC will essentially comprise a federating core and a variety of federated research infrastructures committed to providing services as part of the EOSC.  The main service classes, their connections and the user access channels are currently described as follows:

- The EOSC federating core, understood to be constituted by EOSC shared resources and by a compliance framework including the Rules of Participation. The core will provide the services that research infrastructures can use/are required to use to become part of the EOSC federation. The EOSC-hub project is tasked in H2020 to define and put in place the EOSC federating core. The

federating core of EOSC will relay the resources and the services of data infrastructures funded at EU, national and regional level.

- Services and data contributed to EOSC by various research infrastructures and other initiatives. The service can be both generic[59] and thematic.

  o An initial set of generic services around the core will be federated from the EOSC-hub project. These generic services can provide value-added capabilities to thematic services, or can be relevant for scientific users as well. Generic services include Infrastructure as a Service clouds; Service Monitoring; Resource Usage Accounting service; Helpdesk; Scalable storage.

  o Thematic services and data would enter the federation on a voluntary basis based on the commitment of resources and on the capacity to comply with its rules; minimum commitments would be set in the Rules of Participation to ensure fairness. The first set of thematic services and data will be brought into by EOSC-hub in 2018. The INFRAEOSC-04 RI Cluster projects will from 2019 federate further services and data from disciplinary areas covered by the ESFRI roadmap.

Based on the architecture and service classification in Table 7 (originally presented in D7.3) we can derive the following roles needed to apply the services, the typical activities EOSC stakeholders perform in these roles, and the knowledge they need to be successful.

The 'required knowledge' in Table 7 is a high-level description that can be used to locate relevant competences and capabilities described in more detail in the Skills Framework in Annex B .

**Table 7: Service roles, activities and knowledge**

| Service role | Main activities performed | Required knowledge |
|---|---|---|
| Owners and operators of the federating core | <ul><li>Fit service into the core</li><li>Monitor proper operation of the core as well as the federated services</li><li>Respond to errors/malfunctioning directly or via peers</li><li>Support Generic and Thematic providers federate into EOSC</li></ul> | <ul><li>Architectural understanding of the EOSC core</li><li>Expert-level knowledge of EOSC core services</li><li>Deep technical skills in the operation of core services in scope for the operator</li><li>User/customer support skills</li></ul> |
| Owners and operators of Generic services | <ul><li>Federate the service into EOSC via the federating core</li><li>Maintain operational continuity of the service</li><li>Upgrade the service as required (based on new releases from developer team)</li><li>Recognise and respond to service errors and malfunctioning</li><li>Support and train users of the service</li></ul> | <ul><li>Deep technical skills in the deployment, operation and monitoring the Generic service in scope</li><li>Application of the EOSC 'Rules of Participation' to the Generic service</li><li>User-level knowledge of EOSC operational support tools</li><li>User/customer support skills</li></ul> |
| Owners and operators of Thematic services (incl. Data as a service) | <ul><li>Federate the service into EOSC via the federating core</li><li>Maintain operational continuity of the service</li><li>Upgrade the service as required (based on new releases from developer team)</li><li>Recognise and respond to service errors and malfunctioning</li><li>Support and train users of the service</li></ul> | <ul><li>Deep technical skills in the deployment, operation and monitoring the Thematic service in scope</li><li>Application of the EOSC 'Rules of Participation' to the Thematic service</li><li>User-level knowledge of EOSC operational support tools</li><li>User/customer support skills</li></ul> |

---

[59] The word 'generic' in this context indicates services which are relevant for any scientific discipline, rather than 'thematic services' which are specifically designed and implemented to serve users from a given disciplinary area.

| Users of Generic and/or Thematic services and data (incl. scientific 'end users' as well as local institutional supporters for scientists) | <ul><li>Develop user-level knowledge of the relevant services of EOSC (via documentation, training, consultancy)</li><li>Apply the services and data to generate new scientific findings</li><li>Share the tangible outputs in a way that makes them visible in EOSC (data, workflows, new algorithms, papers)</li></ul> | <ul><li>Navigating the EOSC catalogue and identification of relevant services</li><li>User-level knowledge of relevant services of EOSC</li><li>Generic data science skills (DMP, stewardship, statistics, etc.)</li><li>Data science skills depending on the discipline (data formats, validated algorithms, ethics, etc.)</li></ul> |

### 4.3.3.  Survey of training coordinators in research infrastructures and institutions

The survey comprised eight main questions, employing a mix of rating scale and open question formats, and covering the following topics:

1. Utility of the definitions used in the briefing document, which summarised definitions used in the EOSCpilot D7.3 Skills and Capability Framework.

2. How far the professional group roles identified in the framework matched expectations of those who should be involved in stewardship.

3. To what extent respondents found the framework's 'key dimensions' relevant to planning skills development.

4. To what extent they agreed that researchers need guidance on the FAIR/ open science skills relevant to their career level.

5. Any changes needed to the dimensions or scales used in FAIR4S.

6. Applicability of the framework to each of three use cases:
    a. For EOSC Service providers/operators to specify the competences and capabilities involved in using their service;
    b. For EOSC users to identify learning resources for their own professional development, relating to specific services and more generally to FAIR/open research;
    c. For research leads/ human resource professionals to plan skills development.

7. Extent of agreement with draft conclusions, which were based on previous workshop discussions and landscape analysis summarised in D7.4 Workshop Report.

8. Actions that stakeholder groups need to take to develop data stewardship as a profession.

The survey was delivered to a pre-identified list of 100 people in the two main target groups for the framework:

- Training coordinators in Research Infrastructures were identified by writing to the contact addresses of RIs selected to cover a range of disciplines, resulting in 43 contacts.
- RDM coordinators in Research Producing Organisations (universities) were identified through desk research of a sample of institutions selected to cover a range of EC member states - 57 individuals were identified through their websites and from known contacts of the partners.

Survey delivery was online through the EUSurvey platform, with a covering email invitation linking to a ten-page briefing paper. This email summarised definitions used in the framework and the main dimensions of the framework, pointing to D7.3 for further details. The survey link and briefing paper were also made available online [60]. The survey was open from mid-October to mid-December 2018. A summary of the responses about the framework follows below. Responses on the draft conclusions and actions expected of stakeholders are summarised earlier in the conclusions to the report.

---

[60] https://eoscpilot.eu/media/eoscpilot-policy-and-skills-frameworks-influence-draft-proposals

**Respondent characteristics**

Individuals who completed the survey were asked to self-identify on a number of characteristics (they could choose more than one characteristic). Of the 32 respondents, two thirds indicated they 'do research, consultancy, or policy development in skills, education or training'; 19 (59%) indicated that 'Research infrastructure or e-infrastructure services are the target for my role' and the same number that 'Research producing organisations/ institutions are the target for my role'. A few more had responsibilities for skills covering a single organisation (16) than in multiple ones (14). Asked whether their organisation participates in any current EOSC project, 11 said yes, 8 no, 5 did not know, 1 preferred not to say, and 7 did not answer.

**Framework terms, definitions, and dimensions**

There was positive feedback from two-thirds of respondents, who found the definitions useful. Others encountered some difficulties, the most common being that:

- The roles of 'professional groups' were not clearly defined. In particular 'data service engineering' was not understood (the term was adapted from the EDISON competence framework term 'data science engineer').

- The application of the 'responsibility' dimension was not clear.

Suggestions were made to expand the roles to cover others important in data management support, e.g. IT service providers / managers, IP and commercialisation staff, Data protection staff, local research support staff (e.g. project coordinators). The lack of 'data steward' as a distinct role (rather one shared among the others) was commented on.

Other points were:

- Lack of distinction between 'competence' and 'skill'. Although the first iteration of the framework (D7.1) had defined data stewardship competence as "demonstrable ability to apply the knowledge, skills and attitudes needed to meet responsibilities for data stewardship", the subsequent reports D7.2 and D7.3 had begun to use competence and skill interchangeably.

- Some imprecision in the terms used for individual competences, and under-emphasis of ethics.

- The definition of stewardship as "formalisation of roles and responsibilities to ensure that data is managed for long-term reuse, and that roles are performed in accordance with FAIR data principles" was queried[61].

**Usability for the main use cases**

The survey responses were generally positive about the value of the framework. Almost all respondents (91%) agreed with the statement that "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". Two thirds (66%) agreed that "the FAIR4S framework will help promote education for data stewardship experts, and their reward and recognition', but 19% did not know and 9% disagreed.

Usability for the main use cases was rated disappointingly low, with fewer than a third agreeing that the framework was easily applicable for them. The presentation in the briefing was felt to be over-complex, lacking clarity in how to apply the tables and definitions. There were suggestions to make the presentation simpler and more appealing, and reduce the complexity by reducing the number of dimensions. The following comment covers points that were also made by some six individuals "*For users some examples/use cases will be crucial, for HR professionals FAIR4S involves some understanding of the data lifecycle etc which may not*

---

[61] This was introduced in D7.1, adapting an industry definition: "*The formalization of accountability for the management of data resources.*" Robert Seiner, TDAN.com http://tdan.com/the-data-stewardship-approach-to-data-governance-chapter-1/5037

*be the case at large research organisations, so transposing these expectations onto more familiar traditional roles may be necessary*".

**Steps taken in response**

The survey offered clear evidence of the need for changes to make FAIR4S more fit for purpose. The success of the TU Delft workshop in framing discussion of researchers' skill sets around a subset of the competences encouraged a rethink of the structure, to respond to the shortcomings pointed out by the survey participants. The steps taken to do that were as follows:

1. Identifying 10 key skills from the 59 competence and capability topics listed in version 1.0 of the EOSCpilot skills and capability framework (D7.3), based on those selected by the TU Delft data stewards for the Sept 2018 workshop at TU Delft 'Time for Open Science Skills to Count in Academic Careers'.

2. Professional group titles ('roles') have been changed, aiming to better reflect current institutional roles. Roles now distinguish between career stages for researchers and data scientists, using the standard R1 to R4 classification of the European Research Careers framework[62].

3. Some competence topics were reworded for clarification.

4. Key skills-profiles and role-profiles were added to the framework. These added text to define competence topics and roles, intended to improve usability of the framework for matching service capabilities to relevant roles, and to help plan professional development training.

5. Each key skill-profile lists competences, knowledge and attitudinal aspects for the 'individual' level. The terms 'skill' and 'competence' were used interchangeably in Version 1.0. This was seen as confusing by some survey respondents, and has been changed to reflect the common understanding of competence, as comprising distinct knowledge, skill, and attitudinal aspects.

Basic terms such as 'competence', 'skill', and 'capability' are re-defined in the Glossary at the end of this report to explicitly include knowledge and attitudinal (or where appropriate aptitudinal) aspects of competence as well as skill.

---

[62] https://euraxess.ec.europa.eu/europe/career-development/training-researchers/research-profiles-descriptors

# 5.    MAKING THE CASE FOR A FAIR DATA SKILLS RESOURCE REGISTRY

## 5.1.  Training-as-a-service in EOSC

In the deliverable D7.1 we introduced the model of Training as a Service (TaaS) and a possible infrastructure as well as concepts to enable the implementation of TaaS in EOSC. We advanced the model in D7.2 and D7.3. The TaaS model puts the (EOSC) user into the focus to help them keep up the pace of advancements of the EOSC and its services to ensure effective use of EOSC. For this, we require a tight integration with existing EOSC Services and Components and the reuse of external solutions.

The requirement for a tight integration with existing EOSC Services and external solutions is threefold:

1.  Supplementing services in the EOSC Service Catalogue with relevant training resources for users based on the services they use.

2.  Enabling TaaS services, e.g. repository services for storing learning resources, registry services for storing relevant metadata, or Notebook services to deploy learning materials for exercises and hands-on trainings (see D7.2 for further details).

3.  As a prerequisite for each of the above, applying rules of participation consistently to the provision of learning materials to ensure that minimum criteria for quality and relevance are applied.

Ensuring that training providers conform to rules of participation (see Section 3.3.2) that include criteria for FAIR learning resources would benefit stewardship of these resources, by making it more straightforward to automatically harvest the metadata on these resources into a registry. This would entail asking learning material providers to provide information on the competence topics and expertise level each learning resource is relevant to. In the long term, the TaaS model also foresees a certification and progress tracking to have a sustainable, user-centric service that acts as an active recommender system for EOSC users for relevant learning materials.

In this section, we:

- Discuss and summarize the case for an EOSC training registry.
- Highlight the importance of metadata to make learning materials findable.
- Present a cost-benefit analysis of different solutions for implementing TaaS in EOSC.

## 5.2.  Evidence of the need for an EOSC-wide registry

It should be possible for the EOSC to offer a level of automation and precision to the researcher looking for relevant learning materials. The Training-as-a-Service concept proposed in EOSCpilot deliverable 7.2 identifies a case for providing a training registry service. This would aggregate metadata about learning resources from the individual training catalogues and portals that Infrastructures (and others, such as research institutions) are likely to offer as services in the Service Catalogue.  A similar registry concept is already operational at the domain level in ELIXIR, through the TeSS Training Portal.

As discussed in D7.2, an EOSC Training-as-a-Service should include a cross-domain training registry, based on a minimal set of discriminators to describe learning materials. The D7.3 report builds on that idea, and proposes that services are linked to relevant learning materials via a skills specification in the Service Catalogue. To make this linking work, the metadata of learning materials require the association of competence topic and expertise level.

D7.2 also foresaw the automatic harmonisation and enrichment of EOSCpilot training metadata during the harvesting process, e.g. the automatic recognition of the competence topic and expertise level. Implementing this idea is not (as far as the authors are aware) part of any current project to deliver the EOSC, so we must refer to this a long-term scenario.

The demand for a registry and the training-as-a-service concept has been reinforced through workshop discussions and consultation surveys conducted during the course of the EOSCpilot project. A similar

requirement was also articulated in a report from the Knowledge Exchange[63] in November 2016, shortly before EOSCpilot began. That report merely sought a repository for open training materials, and envisaged administration by a collaboration of national infrastructure providers, of which the Knowledge Exchange itself is one example. It contained concrete recommendations on ensuring that training materials in the repository would be FAIR. Our recommendations can be seen as building on and extending this in the light of developments in the following 2 years.

## 5.3. Collecting FAIR learning resource metadata across communities

As described in Deliverable 7.2 a training-materials and events catalogue helps users acquiring the skills to apply the EOSC services and enable new data science in accordance with FAIR principles. We specifically focus on the FAIR data principles on different levels: First, to make users aware of the FAIR principles by ensuring coverage of appropriate training on the FAIRness of data itself and second, to enable the provision and exposure of FAIR training resources for the EOSC.

The FAIR data principles provide a template to enable FAIRness also for training resources. One of the principles of FAIR is the description of data objects with rich metadata. Overall, we found (D7.2 Section 3.3.1) that there is a lack of metadata (description, keywords) that is required to ensure findability by users. Without descriptions and more specific keywords, it is difficult for users to find the appropriate materials. For accessibility and reusability we discovered a better coverage than we anticipated, while two thirds of the materials provided clear licensing information. We also found a welcome trend towards using persistent identifiers. Overall, we concluded that a dedicated culture-change may be needed by training authors, providers and consumers by adding missing critical information to improve the findability.

The proposed Training as a Service infrastructure relies on structured training information, in order to discover, filter and process available training materials, and also to properly expose the available training materials for the users. In D7.2 we therefore proposed a set of minimum and recommended metadata for training resources. The fields of the proposed metadata were identified via a crosswalk of selected training portals and catalogues from different domains to finally derive a domain-independent set of metadata.

To support pre-existing domain-specific data models and implementation, we explicitly support and call for extension of the proposed metadata. The proposed metadata further integrate specific fields required for finding training materials in context of the Framework. This enables matching of required competences of services listed in the EOSC service catalogue.

D6.3: 1st Report on Data Interoperability[64] highlights in section "Metadata catalogues and datasets in EOSC" the role of data resources and datasets in EOSC and emphasises the importance of relying on a common minimum information guideline. For this, they propose the EDMI guideline that encourages the use of existing standards for generic datasets. We, therefore, advanced our proposed set of metadata for learning materials to be EDMI compliant. The updated set of metadata is shown in Table 8 and Table 9.

---

[63] Recommendations for a Repository for training Materials on RDM http://repository.jisc.ac.uk/6379/19/2016-11-11_KE_Recommendations_for_a_Repository_for_Training_Materials_on...pdf

[64] D6.3: 1st Report on Data Interoperability: Findability and Interoperability https://www.eoscpilot.eu/content/d63-1st-report-data-interoperability-findability-and-interoperability

**Table 8: Updated minimal set of properties for learning materials (EDMI compliant)**

| Proposed property | Type | Multiplicity | Description |
|---|---|---|---|
| Title | Text | one | The title of the learning material |
| Description | Text | one | A short summary describing the learning material |
| PID | Text, URL, PropertyValue | one | Persistent identifier of learning material, e.g. DOI |
| Url | URL | one | Link to the learning material |
| Author | Person, Organisation | many | The author of the learning material |
| Date modified | Datetime, Date | one | Date/time of most recent change of learning material excluding metadata |

**Table 9: Updated recommended set of properties for learning materials**

| Proposed property | Type | Multiplicity | Description |
|---|---|---|---|
| Keywords | Text | many | Keywords describing the learning material |
| Licence | CreativeWork, URL | one | The licence of the learning material |
| Domain | Text | many | A field of science or expertise |
| Competence | Text | many | The expertise level (comprehend, apply, synthesize/ evaluate) and competence area of the FAIR4S framework |
| Target professional group | Audience | many | Any appropriate professional profiled defined in the EDISON project: data stewards and more established professional groups that support research, particularly data managers, data service engineers, and data scientists/analysts |

These metadata terms have subsequently been adopted by the EOSC-hub project. According to its deliverable D11.1: Training materials about common services and thematic services[65] (p.15): "*The first release of the EOSC-hub online training catalogue was announced during the EOSC-hub All Hands Meeting (AHM) in Malaga in April 2018. The online catalogue allows users to register new training events and upload training materials developed by project members. The overall architecture of the training catalogue is based on Deliverable 7.2 of EOSCpilot. The selection of the metadata fields applied in the EOSC-hub registry is also based on D7.2*".

---

[65] https://documents.egi.eu/public/ShowDocument?docid=3416

The above metadata terms proposed in EOSCpilot D7.2 are consistent with existing standards, although they have not been mapped to them. A desirable next step would be to define mappings between the terms and the following standards for producing machine-readable description of educational materials.

The Learning Resource Metadata Initiative (LRMI) is a mature standard for describing educational resources, initially developed by Creative Commons and the Association of Educational Publishers, and now maintained by the Dublin Core Metadata Initiative. LRMI is also endorsed by schema.org, the search industry standard for describing online content mark-up in machine-readable terms[66].

The bioschemas.org standard is a more recent initiative in the life sciences[67]. This standard aims to improve data interoperability in life sciences by encouraging life science organisations to use schema.org markup in their websites and services. It does this by producing specifications that describe life sciences content using terminology that maps to schema.org terms, including for example training materials and events.

The bioschemas.org profiles have significant potential for application in an EOSC training registry. Despite their origin in a single domain, the specifications for training materials and events are not domain-specific, and current work is focused on defining terms that content providers may apply to improve the FAIRness of training materials.

The LRMI and bioschemas standards are deployed in two of the services (DMTC and TeSS, respectively) that we highlight in the next section as current exemplars for an EOSC registry of training materials.

## 5.4. Providing a Training Registry – alternative models

A range of examples are available to inform the case for an EOSC-wide registry of training materials. Those that follow are not the only ones available – FOSTER Plus is a notable omission – but are useful for comparison.

### 5.4.1. Catalogue of training services (e.g. EOSC portal)

Key points:

- Users can find training services and resources using generic service catalogue metadata.
- Manually curated - providers nominate own resource, with editorial control.

**Example: The EOSC portal**

The EOSC Portal was launched in November 2018 to coincide with the launch of the EOSC itself. This portal has been created by the projects E-InfraCentral, EOSC-hub and OpenAIRE Advance. The content of the portal is selected by an Editorial Board with representation from these projects. The listings are not yet extensive, and are derived from the three contributing projects' portals, which have similar content. Listings include individual training instances as well as references to training catalogues.

The future of the EOSC portal is currently unclear and further implementation, as we understand, depends on the outcome of the INFRAEOSC-06-2019-2020 call - Enhancing the EOSC Portal and connecting thematic clouds.

A logical extension of this 'catalogue of training services' approach would be to provide more specific metadata relating to training sites specifically, as for example re3data does for data repositories. In the case of training services, this might include characteristics such as the domains covered, content providers, metadata standard used, any certification of the training or support for learners to obtain badges for completing it.

---

### 5.4.2. Manually curated registry of learning resources (e.g. EOSC-hub, DMT Clearing House)

Key points:

- Content providers are selected project partner sites.
- Metadata standard is not domain specific.

**Example 1 - EOSC-hub training catalogue of training materials and events**

The EOSC-hub training catalogue became operational in December 2018. It is created for the projects' service and training providers. The development of the EOSC-hub training catalogue (both training events and training material) is described in detail in EOSC-hub D11.1. The catalogues are based on description fields provided by EOSCpilot. The catalogue currently does not have a harvesting / aggregation function. The development of the service consisted of a requirements analysis, an evaluation of alternatives to implement the service (using existing solutions vs own development), building the service and reporting on it. These activities cost about 5 PM. Operation and maintenance of the service is 0.5 - 1.0 PM per year. Training materials and events can be added by the service providers using a metadata entry form (taking 10-15 minutes when e.g. licences are clear to the depositor). The metadata that is required is based on recommendations by EOSCpilot.

The most challenging issues in aggregating skills content are quality assessment and heterogeneity of the content providers and target groups addressed by the training catalogue. Quality assessment concerns to what extent the training material is up-to-date. For this, an assessment of an expert is required. The heterogeneity issue is related to the fact that in EOSC-hub a wide range of scientific disciplines are represented that do not share a common understanding on all aspects of data management. EOSC-hub tries to cope with these issues to consult content providers on a regular basis asking them whether training material is still current or has to be updated. The heterogeneity issue is tackled by putting the target group in the training documentation, so the target group can assess to what extent a training is relevant for them.

**Example 2 - Data Management Training Clearing House**

The Data Management Training Clearing House is a searchable and browsable registry of training materials, increase discoverability and accessibility of learning resources to inform and train scientists about research data management. The DMTC collates these resources and adds metadata intended to help research scientists, students, or teachers assess whether resource are appropriate and useful for their needs. Initially funded through the US Geological Survey and supported through ESIP (Earth Sciences Information Partnership), the DTMC is currently funded through the US Institute of Museum and Library Studies (IMLS), broadening its scope internationally and beyond Earth Sciences. Its coverage currently includes Education, Engineering, Life Sciences, Medical and Health Sciences, Physical Sciences and Mathematics, and Social and Behavioural Sciences. The Clearinghouse uses the Learning Resource Metadata Initiative (LRMI) metadata schema for the resources included in their registry.

### 5.4.3. Semi-automated curation of domain-based registry (e.g. TeSS)

Key points:

- Users can search for skills resources held on a number of content provider sites, each affiliated to a single research domain (or Research Infrastructure).

- The service holds metadata about skills resource (training or other learning material, or event) that may have been entered manually by an authorised provider, or has been scraped from contributing sites.

- Users may search on metadata describing learning resources e.g. target audience, difficulty level.

- Employs a metadata standard that includes domain-specific subject terms drawn from a machine-readable vocabulary.

**Example: TeSS, ELIXIR's Training Portal[68]**

TeSS enables browsing and searching in training content provided by life sciences organisations. The TeSS training events and materials portal was developed from 2015, when infrastructural test versions and systems were initiated. Since then building the portal has been an iterative process, enhancing TeSS and adding continuously new functionality. TeSS has been funded for five years by BBSRC and ELIXIR-EXCELERATE, and now has a further year of funding by ELIXIR, and potential take-up in EOSC-Life Cluster Project.

TeSS aggregates and integrates life science training materials and event metadata from ELIXIR nodes, collaborating with participating sites to do this automatically. ELIXIR collaborates closely with EMBL-EBI and with GOBLET, the Global Organisation for Bioinformatics Learning, Education & Training. TeSS employs scraper software to derive metadata records from sites employing the bioschema.org metadata profile, based on schema.org, in their content markup. Increased adoption of the bioschemas standard has improved the efficiency and effectiveness of the content scraping approach.

## 5.5. A straw-man comparison of costs and benefits

Each of the example services was asked to complete a questionnaire characterising the cost-benefit factors in developing and maintaining their training resource registries, providing estimates from project budgets and time recording. The questionnaire was based on a simplified cost model using the following cost categories[69]:

1. General cost factors:
   - Project duration, research domains served, provider sites, number of items registered.

2. R&D and maintenance:
   - Requirements analysis, Set metadata standards, Negotiate with the providers, Facilitate metadata exposure and discovery on provider sites, Develop infrastructural systems, Develop and test new functions, System Design,Web design, Generate scripts for partner site scraping, Establish service management (e.g. help desk, change management);
   - Day to day service operation, Provide trainings and guidance materials.

3. Content provision: time to complete metadata entry form, and/or time to set up provider site to expose metadata for harvesting/ aggregation:
   - Content acquisition: outreach and depositor support, promotion, agreeing on metadata, data formats access rights;
   - Metadata management to describe skills resources, populate registry.

### 5.5.1. General cost factors and development costs

**Data Management Training Clearing House (DMTC)**

Domains served: Education, Engineering, Life Sciences, Medical and Health Sciences, Physical Sciences and Mathematics, and Social and Behavioural Sciences.

Content scope: RDM.

Providers: ICSU - World Data System, USGS, DataONE, the Digital Preservation Network, ESIP.

**EOSC-Hub**

Domains served: all.

Content scope: EOSC service operation, RDM.

---

Providers: EGI, EUDAT, Indigo Data Cloud, CERN, ENES, CLARIN, DANS.

**ELIXIR Training e-Support System (TeSS)**

Domains served: Life Sciences.

Content scope:  Bioinformatics, RDM.

Providers: 65 providers - ELIXIR nodes and bioinformatics content providers.

**Table 10: Comparison of training resource registry models**

|  | DMTC | EOSC-hub | TeSS |
|---|---|---|---|
| Months duration to March 2019 | 48 | 14 | 60 |
| Months operational | 26 | 12 | 32 |
| Current #items - training events | n/a | 45 | 9505 |
| Current #items - training materials | 315 | 50 | 1208 |
| Development time (Person-Months) | 8.7 | 5.0 | 22.5 |
| ● Requirements analysis | 0.9 | 1.0 | 1 |
| ● Setting metadata standards | 1.0 | 0.0 | 0.25 |
| ● Negotiating with the content providers | 0.7 | 1.0 | 0.5 |
| ● Developing infrastructural systems | 0.3 | 1.0 | 8 |
| ● Developing and test new functions | 0.7 | n/a | 8 |
| ● Systems Design | 0.2 | 1.0 | 2 |
| ● Web design | 0.4 | n/a | 2 |
| ● Establishing service management (e.g. help desk, change management) | 0.1 | n/a | n/a |
| Day to day service operation | 0.8 | n/a | 0.5 |
| Providing trainings and guidance materials | 0.9 | 1.0 | 0.25 |

### 5.5.2. Comparing the approaches

The time figures should be compared with a high level of caution for a number of reasons, including differences in time counting approaches. These rule out any quantitative assessment of cost efficiency. Instead they serve to highlight the differences in project context, scope, functionality, community engagement, and the model for populating the registry.

**Project context differences**

Unlike DMTC and TeSS, the EOSC-hub training registry is an integral part of a larger service catalogue, in turn one stream of work in the larger EOSC-hub project. This makes project management, service management and marketing costs difficult to apportion to the training material alone. These have been excluded from the figures in Table 10.

**Scope**

Although broadly similar in terms of providing material and events on data-related topic, DMTC has a more specific focus on RDM. Despite a current focus on a single domain, TeSS is broader than data in covering bio-informatics more generally, and EOSC-hub broader in covering more technical aspects of cloud service operation and management, as well as data management issues that are not service-specific.

**Functionality**

TeSS development effort spans a broader range of functions than DMTC and EOSC-Hub. These support content acquisition and curation, and the application of training materials in training workflows.

**Community engagement**

DMTC has established a broadly based cross-disciplinary network of provider organisations and content contributors. As a consequence, much of the content acquisition effort is counted as voluntary by the project (estimated at approximately one-third).

**Metadata maintenance – populating the registry**

DMTC estimates (not included in the table) are 4.5PMs to create 440 records, of which 315 have been published and the rest queued. The cost of manually entering and curating the registry entries (cleaning, enriching, updating) across the three examples is difficult to establish, partly due to limited time for this exercise, but also because of the varying extent to which time is externalised from the project costs.

EOSC-hub estimates each entry takes 15 minutes to complete. Time for additional information gathering e.g. on licences is not included. This can be very time consuming. Quality assurance is also an expensive activity. Each entry in the catalogue has to be evaluated in communication with the depositor. The depositor in most cases is the expert that can determine whether the training material is outdated, has to be removed from the catalogue or replaced by a new version. A rough estimation: this communication can take up to one day per training material item.

### 5.5.3.  Automating the approach – TeSS content scraping

The approach taken in TeSS requires initial set-up and maintenance of scraper software scripts for each content provider (39 currently). Maintaining these takes approximately 6 hours per week. It should be noted that not all TeSS content is automatically aggregated. Currently this applies to two-thirds of the entries referring to training materials and 95% of those referring to training events.

The process for adding a site involves the following three main steps:

1. Planning:
- Identify what training resources are on their website and the URLs for them;
- Determine a programmatic way to access each of the relevant pages (e.g. parse sitemap.xml, or parse an index page);
- Look at what metadata properties are being presented for each resource;
- Map these to the Bioschemas specifications.

   Amount of time - 1 hour.

2. Adoption: In their codebase,
- Create a snippet of HTML that can be included in each training page;
- Add a template of the schema.org JSON-LD;

- Fill in the values of each schema.org property with the corresponding variables from the table;
- Test and release.

Amount of time - 2 hours.

3. Aggregation:
- In TeSS, create a scraper that accesses each of the pages and parses any schema.org found, and pushes it to TeSS. This typically re-uses lots of code from other scrapers.

Amount of time - 1-3 hours.

### 5.5.4. Conclusions from cost-benefit comparison

The three examples described each illustrate characteristics that any EOSC-wide registry should emulate; the broad engagement of DMTC, the breadth of scope and service-related coverage of EOSC-hub, and the adoption of community-supported machine-readable standards by both DMTC and TeSS.

The semi-automated approach that TeSS illustrates is likely to be essential for any registry whose scale includes thousands of items, particularly if these are likely to change, due to the costs and complexity of synchronising manual updates.

It seems beyond doubt that an EOSC-wide registry should have such a scale, based on the likely number of content providers (at least in the hundreds), disciplinary variations of training content (at least in the order of six), the need for variations to deal with different levels of expertise, and the number of topics that the content deals with (e.g. FAIR4S has 59 topics).

While bespoke suites of training material will and should continue to be produced on the individual sites of Research Infrastructures and institutions, there is a clear need to support their discovery in a consistent, standardised and automated manner.

## ANNEX A.    GLOSSARY

| Term | Explanation |
|------|-------------|
| **Capability** | *Competence* applied at a research team or organisational level, with a defined level of expertise, to perform a service role or work in the EOSC environment. |
| **Competence** | Knowledge, skill and aptitude or attitude relating to a topic e.g. 'workflow set-up and management', combined with an *expertise level.* |
| **Expertise level** | A description of level of *competence*; e.g. to indicate whether someone has a basic awareness of the topic, or an intermediate-level ability to do it, or expert competence sufficient to contribute something new to theory or practice. |
| **Professional group** | A person's domain of responsibility defined by a set of *competence*s, e.g. domain research, data science/ analytics, data management, data service engineering. |
| **Responsibility level** | Application of a *competence* to the delivery of one or more *capabilities* in a specific research team or organisational context. Levels are, from low to high: support, discretion, substantial discretion, objective setting, or full accountability. |
| **Service** | A *service* described in the EOSC service portfolio, or service catalogue, that offers value or reduces risk to a *professional group* or other designated community. |
| **Service class** | A broad type of *service*, identified in the EOSC Implementation Roadmap. |
| **Skills group** | Skills groups are of two types: those applied in a research project-specific way e.g. 'capture and process', and those applied consistently across (research) projects, e.g. 'govern and assess'. |
| **Skill** | Demonstrable application of the knowledge and aptitude or attitudes needed to deliver a *competence* or *capability* to a specific context. Skills may be specified in the form of a *skills user story.* |
| **Skills user story** | A method for specifying skills in a form that combines into a sentence the terms used as search parameters for a training materials catalogue, using the following convention:<br><br>As a [*professional group*] needing to [*deliver a capability*] I should have [*expertise level*] in [*competence*]".<br><br>The [deliver a capability] element of the user story should be expressed in a form appropriate to the service and the high-level requirements it fulfils. For example:<br><br>As a [*data manager*] needing to [*support colleagues in selecting repository services that are relevant to the research domain, and comply with relevant regulatory and policy frameworks*] I should [*comprehend*] [*repository/database evaluation and options appraisal*]. |

## ANNEX B.    FAIR4S ROLE AND SKILL PROFILES AND TABLES

This annex presents templates for skills profiles (B.1) and roles profiles (B.2) followed by worked examples of each in Sections B.3 and B.4. Finally, Section B.5 contains tabular representations of the expected levels of expertise for researchers (at levels R1 to R4), data scientists (at levels R1 to R4), and for various different professional roles amongst data advisors and data service providers.

### B.1.    Skills Profile Template

| Skills Profile | Topic of competence or capability – from Skills Tables |
|---|---|

*What knowledge, skills and attitudes or aptitudes are needed?*

**K- Basic:** describes what individual should know to have a basic comprehension of topic (30-60 words approx.)
**S - Basic:** describes actions the individual should be able to perform at a basic level
**A - Basic:** describe attitude or aptitude the individual should display at basic level

**K- Intermediate:** describes what individual should know to have intermediate comprehension of topic (30-60 words)
**S - Intermediate:** describes actions the individual should be able to perform at an intermediate level
**A - Intermediate:** describe attitude or aptitude the individual should display at an intermediate level

**K- Expert:** describes what individual should know to have expert comprehension of topic (30-60 words)
**S - Expert** describes actions the individual should be able to perform at an expert level
**A - Expert:** describe attitude or aptitude the individual should display at an expert level

*What counts as success?*

Describes what activities should be performed and what evidence should be available that this has happened, in 30-60 words approx..

*Whose roles apply this skill?*                              *(bold= directly involved, \* indicates a key skill for the role)*
Researchers:                    at which career levels (indicate R1-R4)
Data scientists:                at which career levels (indicate R1-R4)
Data advisors 'front-office':  identify which support roles directly involved
Data service provider: - identify which support roles directly involved

*What related competences should the research team have?*

| From the Skills Tables, list those competences that are in the same skills group, or may otherwise be required of at least one team member | From the Skills Tables, list those competences that are in the same skills group, or may otherwise be required of at least one team member |
|---|---|

*What capabilities - and services - would amplify this skill?*

| From the 'related competences' list above, select one that enhances the ability of individuals to be competent, and is supported by one or more of EOSC services | One or more of EOSC service categories |
|---|---|
| From the 'related competences' list above, select a 2nd example that enhances the ability of individuals to be competent, and is supported by one or more of EOSC services | One or more of EOSC service categories |
| From the 'related competences' list above, select a 3rd example that enhances the ability of individuals to be competent, and is supported by one or more of EOSC services | One or more of EOSC service categories |

## B.2.    Role Profile Template

| Role Profile | Name of role | | |
|---|---|---|---|
| **Key skills** | | **Level** | **Who may help?** |
| **Name of key skill**<br>(selected from 10 listed in main report Table 3) | | Basic/ Intermediate / Expert | Selected data advisor and service provider roles |
| Describes knowledge, skill and attitudes or aptitudes listed for the skill in the corresponding Skill Profile, contextualising these to the role.  Approx 100 words | | | |
| **Name of another key skill**<br>(selected from 10 listed in main report Table 3) | | Basic/ Intermediate / Expert | Selected data advisor and service provider roles |
| Describes knowledge, skill and attitudes or aptitudes listed for the skill in the corresponding Skill Profile, contextualising these to the role.  Approx 100 words | | | |
| **Name of another key skill**<br>(selected from 10 listed in main report Table 3) | | Basic/ Intermediate/ Expert | Selected data advisor and service provider roles |
| Describes knowledge, skill and attitudes or aptitudes listed for the skill in the corresponding Skill Profile, contextualising these to the role.  Approx 100 words | | | |
| **Name of another key skill**<br>(selected from 10 listed in main report Table 3) | | Basic/ Intermediate/ Expert | Selected data advisor and service provider roles |
| Describes knowledge, skill and attitudes or aptitudes listed for the skill in the corresponding Skill Profile, contextualising these to the role.  Approx 100 words | | | |
| **Where can you learn?**<br>-    Name of key skill,<br>-    List titles and links to resources relevant to the topic, experise level, role (target audience) | | | |

## B.3.    Skill Profiles Examples

| Skills Profile | PLAN - **Plan the stewardship and sharing of FAIR outputs** |
|---|---|

### *What knowledge, skills and attitudes or aptitudes are needed?*

**K- Basic:** Show awareness of FAIR principles, relevant data policies of funders or journals, and legal and ethical requirements. Understand why funders have these expectations, why scientific organisations and research institutions support them, and what should be covered in the Data Management Plan (DMP).

**S - Basic:** Describe guidelines from relevant funders, journals or other organisations representing your domain, and identify what data, software or other objects will be reused or created, and how in general terms the FAIR principles will be applied to them.

**A - All levels:** Proactively give advice or seek it from appropriate colleagues, data advisors and data service providers.

**K - Intermediate:** Describe how the research, thesis or dissertation objectives will draw on sources of evidence or materials, the methods that will be applied, and how outputs.

**S - Intermediate:** Following guidelines from relevant funders and domain organisations, describe how relevant standards and services will be used to make data or code findable, accessible, interoperable and reusable, and who is responsible for this.

**K - Expert:** Describe appropriate criteria to meet user and stakeholder expectations regarding implementation of FAIR principles, demonstrating in-depth understanding of data-driven research methods in your domain and the data sources, formats, and disciplinary norms applicable in your research field.

**S - Expert -** Determine appropriate criteria to meet stakeholder expectations regarding implementation of FAIR principles. Initiate process for managing data, code, or other outputs across the research team. Develop templates and contribute to local guidance, aligning this with data policies and processes of the organisation and funders. Give support and supervision to team members, and contribute to support services through your peer network.

### *What counts as success?*

Each research study has a Data Management Plan (or similar, e.g. Data Stewardship Plan, Software Management Plan, or Outputs Management Plan) is produced and reviewed by the Principle Investigator, supervisor, or funder. It is referred to throughout the study it relates to, and updated if necessary. Updates include any significant changes in methods, standards, or services that have been or will be used to make outputs FAIR, or in responsibilities for these.

### *Whose roles apply this skill?*                    *(bold= directly involved, \* indicates a key skill for the role)*

Researchers:                  - **R1 \*** - R2 - **R3\*** - R4
Data scientists:             - R1 - **R2 \*** - **R3 \*** - R4
Data advisors 'front-office':  - **Data Steward \*** - **Research Manager** - **User support & Training/Outreach** \*
                                        - **Ethics & Data Protection Advisor** - **Commercialisation Advisor**
Data service provider: - 'back office': - Data Librarian - **Service Manager/ project manager** - Research Software Engineer
                                        - Data Service Architect - Archivist

### *What related competences should the research team have?*

| | |
|---|---|
| Application of policies to comply with legal requirements, ethical, and FAIR principles<br>Costing of data management and preservation<br>Modeling data structures and defining database needs | Specifying metadata and persistent id. standards<br>Analysing requirements for services or software<br>Appraising and selecting repositories for FAIR sharing<br>Training in open methods and services |

### *What capabilities - and services - would amplify this skill?*

| | |
|---|---|
| **Prepare and document data/code to make outputs FAIR :** Supports research groups to determine and fulfil appropriate criteria to reuse, manage, and share FAIR research outputs, and fulfil these expectations according to domain norms and standards. | Data management |

| | |
|---|---|
| **Apply policies to comply with legal requirements , ethical & FAIR principles**: Develops local policies and guidelines for publishing research data and related objects, and for selecting repositories that comply with relevant regulatory and policy frameworks. | Training and Support |
| **Publish FAIR outputs on recommended repositories :** Provides access to repository functions to ensure preserved research objects remain FAIR for as long as required, enabling these functions to interoperate with other systems or processes that provide or extract information to maintain or enhance FAIRness of the research objects held. | Storage Sharing and discovery |

| Skills Profile | **REUSE -  Reuse data from existing sources** |
|---|---|

### *What knowledge, skills and attitudes or aptitudes are needed?*

**K -  Basic:** Understand key characteristics of a trustworthy repository service, show familiarity with directories of these services and other reputable sources of data, code and other research outputs for the domain concerned. Understand basic metadata terms, descriptive information, constraints on reuse due to licensing or privacy conditions, and and the conversion, cleaning or normalisation techniques typically used with secondary data in the domain.

**S -  Basic:**  Search for appropriate sources of data or code, find relevant material, describe techniques to reuse it.
**A -  Basic:** Show willingness to look for existing data or software sources relevant to the problem, rather than creating new data by default, and seek help in doing so from colleagues and professional support services.

**K -  Intermediate:** Identify useful sources of data or code, understand how to search them. Identify good examples of data reuse in the domain, and the metadata or descriptive information used to enable this.  Understand how to apply conversion, cleaning or normalisation techniques. Understand how others, e.g. Data Stewards and Research Software Engineers, may support reuse.
**S -  Intermediate**:  Demonstrate how to find, access and reuse data/ code in your domain, employing appropriate techniques to make these actionable for the current purpose, e.g. by transferring into a local workspace, cleaning and normalising to a standard, re-compiling or re-configuring code to prepare for reuse, possibly through integration with other objects.
**A  -  Intermediate:**  Give helpful advice to students and colleagues on suitable sources and techniques for reuse, and seek advice from professional research support services  where appropriate to broaden reuse opportunities.

**K -  Expert:**  Demonstrate expert knowledge and creativity to find, access, integrate and reuse data for novel purposes.
**S -  Expert:**  Demonstrate the ability to make reused data or code actionable for excellent research. Produce expert advice on reuse constraints due to licensing or privacy conditions.
**A -  Expert:** Show creativity in translating secondary data or code from its original context, to address new questions and problems, or to produce new tools to do so.

### *What counts as success?*

Research data, code, and related outputs are found, accessed, and made to interoperate with those locally available. They are made actionable for reuse within licensing or privacy conditions, and enable excellent research, teaching or other applications.

### *Whose roles apply this skill?*                     *(bold= directly involved, * indicates a key skill for the role)*
Researchers:                    **- R1  - R2\* - R3 - R4**
Data scientists:               **- R1\* -  R2\*  - R3  - R4**
Data advisors 'front-office': - **Data Steward**  - Research Manager  - User support & Training/Outreach
                         - **Ethics & Data Protection Advisor  -  Commercialisation Advisor**
Data service provider: - 'back office': - **Data Librarian**  -  Service Manager/ project manager  - **Research Software Engineer**
                         - Data Service Architect  - **Archivist**

*What related competences should the research team have?*

| | |
|---|---|
| Appraise and select repositories for FAIR sharing | File naming and organisation |
| Manage databases | Data cleaning, processing and software versioning |
| Software prototyping | Creative problem solving, flexibility |
| Set up and document workflows | Data transformation and integration |

*What capabilities - and services - would amplify this skill?*

| | |
|---|---|
| **Apply policies to comply with legal requirements , ethical & FAIR principles**: Develops local policies and guidelines for publishing research data and related objects, and for selecting repositories that comply with relevant regulatory and policy frameworks. | Training and Support |
| **Prepare and document data/code to make outputs FAIR :** Supports research groups to determine and fulfil appropriate criteria to reuse, manage, and share FAIR research outputs, and fulfil these expectations according to domain norms and standards. | Sharing and discovery Data management |
| **Data cleaning, processing and software versioning**: Provides access to relevant tools, services and infrastructure. | Service catalogue Data management Processing and analysis |

| Skills Profile | **SERVICE - Use or develop FAIR research tools, resources, or services** |
|---|---|

*What knowledge, skills and attitudes or aptitudes are needed?*

| |
|---|
| **K - Basic:** Understand the scope of EOSC service catalogue, local catalogues, and other reputable sources of tools, resources, or services. Identify relevant examples that fit research purposes and organisational requirements. Understand service level and user agreement concepts. Appreciate how to identify use cases for needs unmet by available tools, resources or services, and understand basic coding concepts, semantic resources, and scripting languages used to create these. <br> **S - Basic:** Search and navigate service catalogues to find, and access tools (etc), identify how to use these for basic applications in the research context, making them interoperable with local data or tools, within licensing or privacy constraints. Identify scripting languages relevant to the domain, and write scripts to perform simple data management routines. <br> **A - Basic:** Be prepared to offer and seek advice to research colleagues and professional support services. |
| **K - Intermediate:** Understand how FAIR principles apply to software and services. Be familiar with the Research Software Engineer role, and how to articulate functional and non-functional requirements for software in a form that a Research Software Engineer can translate into code that fulfils them. <br> **S - Intermediate**: Demonstrate the ability to apply research tools, resources, or services in virtual environments to fulfil research needs or reuse research data for non-academic applications. Demonstrate that locally developed tools, resources, or services are FAIR, and on what licence or accessibility terms. <br> **A - Intermediate:** Demonstrate willingness to work constructively with others to apply or adapt existing tools, resources or services to solve problems. Be willing to learn new techniques. |
| **K - Expert:** Understand 'rules of participation' for EOSC service providers and how to make services operational in the EOSC. Be familiar with a variety of coding techniques, software architectures, and cloud computing concepts. Be able to identify opportunities to apply these to the research domain. <br> **S - Expert:** Demonstrate the ability to construct novel data management workflows by assembling tools, resources or services that perform data management tasks more efficiently or effectively. Be able to specify service level agreements and terms of use consistent with FAIR principles. <br> **A - Expert:** Demonstrate creativity in adapting existing tools, resources or services to solve problems. Develop novel techniques to perform research data management tasks more efficiently or effectively. |

*What counts as success?*

Researchers are able to apply, adapt, or participate in the development of FAIR research tools, resources, or services.

***Whose roles apply this skill?***                                      *(bold= directly involved, \* indicates a key skill for the role)*

Researchers:                          - **R1  - R2 - R3** - R4
Data scientists:                      - **R1 -  R2 - R3  - R4**
Data advisors 'front-office':  - **Data Steward**  - Research Manager  - **User support & Training/Outreach**
                                     - Ethics & Data Protection Advisor  -  **Commercialisation Advisor**
Data service provider: - 'back office': - Data Librarian  -  **Service Manager/ project manager  - Research Software Engineer**
                                      - **Data Service Architect**  - Archivist

***What related competences should the research team have?***

| | |
|---|---|
| Analyse requirements for services or software<br>Creative problem solving, flexibility<br>Data cleaning, processing and software versioning | Data transformation and integration<br>Data mining, querying, interpretation<br>Perseverance delivering diverse open research projects |

***What capabilities - and services - would amplify this skill?***

| | |
|---|---|
| **Training in open methods, services:** Provides learning resources and advice on applying FAIR principles to tools, resources and services. Resources and advice address domain-specific issues appropriately. | Training and support |
| **Service level management:**  Offers guidance on using service catalogues, understanding service levels and service management frameworks, including EOSC rules of participation for service providers. | Service catalogue<br>Data management |
| **Information security and risk management:**  Assesses data and information security risks of deploying and adapting third-party software and resources, offering practical and proportionate approaches to manage risks. | Security and operations<br>Training and support |

## B.4.    Role Profile Examples

### B.4.1    Researchers from career stages R1 to R4

| Role Profile | First Stage Researcher (R1) (Up to the point of PhD) |
|---|---|

| Key skills | Level | Who may help? |
|---|---|---|
| **Plan for the stewardship and sharing of FAIR outputs** | Basic | Data Steward<br>Established researcher |
| Many funding bodies expect research projects to start with a plan of how the outputs will be made FAIR for future use. Some expectations may also apply to PhD projects. You can show you understand why funders have these expectations, why scientific organisations and research institutions support them, and what should be covered in the Data Management Plan (DMP). The planning should include databases, software code, and physical samples where appropriate. You can articulate how planning benefits yourself as the 'first reuser' of your outputs, and seek help where needed from a research supervisor, and from your institution. | | |
| **Apply policies to comply with legal, ethical and FAIR principles.** | Intermediate | Established researcher<br>Ethics and data protection advisor<br>Data Steward<br>Service / Project Manager |
| From the beginning of your career you need to know the regulations, ethical principles, and community expectations affecting research, and how to apply them. Responsibilities will apply to you, your colleagues, your organisation, and to any service providers used. Making research data and code FAIR helps meet your responsibility to be able to base published claims on evidence, and share that evidence. If the research involves public concerns, risks to individuals, or impacts on society you also have an ethical responsibility to engage with representatives of those affected. You are confident discussing with senior colleagues how you have not only complied with relevant requirements but conducted the research professionally. | | |
| **Recognise, cite and acknowledge contributions** | Intermediate | Data Librarian |
| To apply principles of research integrity and professional conduct you can show that you properly cite any data, code and methods that you reuse. When you publish your thesis or dissertation you also acknowledge your collaborators, technicians or others who have contributed to results, as co-authors where appropriate. You use standards to credit those who helped with collection, management, documentation, publication and archiving of research outputs, so that everyone's expertise is appropriately rewarded. By using standard output identifiers (e.g. DOIs) researcher identifiers (e.g. ORCID) and contributor roles (e.g. the CRediT taxonomy) you also help to make your outputs findable by others. | | |

| **Secure funding for open science /support** | Basic | Established researcher (supervisor) Data Steward Research support advisor Service / Project Manager |
|---|---|---|

You can demonstrate knowledge of research funding sources, including for open science and innovation. You are aware of what funders in your field typically look for in an application. That will include a realistic budget, to include the costs of making outputs FAIR, as open as possible, and as closed as necessary. You know where to find specialist advice e.g. on storage costs, or data and metadata formats. You can also describe how to apply for dedicated funding to promote open science activities, events, training or other professional development activities. You are confident in discussing research questions and the steps needed to turn ideas into a grant application, with help from a research supervisor, and from your organisation's professional advisors.

**Where can you learn?**
Key skills for PhD students:
Introduction to Research Data Management and Data Management Plans (CODATA)
http://indico.ictp.it/event/8329/other-view?view=ictptimetable
Research ethics library (FOSTER Open Science)
https://www.fosteropenscience.eu/content/research-ethics-library
Data citation (ANDS Guide)
https://www.ands.org.au/guides/data-citation-awareness

Further sources on open data science and scholarship for PhD students
Open Science MOOC
https://opensciencemooc.eu/resources/
Utrecht Summerschool - Open science and scholarship
https://docs.google.com/document/d/1wiVoLXz6wsFwpL0zmAoIyGH_wTwHHly1t-ZgZtxm8NE/edit

| Role Profile | Recognised Researcher (R2) (PhD holders or equivalent, not yet fully independent) |
|---|---|

| Key skills | Level | Who may help? |
|---|---|---|
| **Reuse data from existing sources** | Intermediate | Data Steward<br>Data Librarian<br>Archivist |

Understanding the opportunities that existing sources can offer as raw material for your own research, you can demonstrate the skills to find, access, integrate and reuse data from these sources. These may be publicly available trustworthy repositories in your own domain, or other reputable sources, including your network of collaborators. You help students and colleagues to translate secondary data or code from its original context, to address new questions or problems. You seek advice from other professionals where appropriate, to help broaden the opportunities available, or to make data actionable, whether it is from external sources or within your own team.

| | | |
|---|---|---|
| **Use or develop FAIR research tools or services** | Basic | Research Software Engineer<br>Data Service Architect |

You can describe the opportunities that electronic lab notebooks, virtual research environments and other online services should offer your research. If your research analyses depend on specific software code or scripts these are made as open as possible, or as closed as necessary to comply with legal obligations. You.can show you understand how to give and get attribution for any contributions that software authors make to published results.

| | | |
|---|---|---|
| **Prepare and document data/code to make outputs FAIR** | Intermediate | Data Steward<br>Research Software Engineer<br>Data Librarian |

You can identify examples of data and code that offer insights to advance your field, and understand the importance of these being FAIR. You understand also that the value of data for reuse, and as evidence for published research claims, depends on there being a traceable path of documentation from creation to analysis. You can apply provenance concepts in your community, and use standard formats and identifiers for metadata and data. Using these you help yourself and others to find and get further practical value from research data, making it accessible, and recording how data is managed  to ensure it is interoperable and reusable in different contexts.

| | | |
|---|---|---|
| **Recognise, cite and acknowledge contributions** | Intermediate | Data Steward<br>Data Librarian |

Building on your awareness of research integrity principles and codes of professional research conduct you show you are able to properly cite any data, code and methods that you reuse. When you publish results you also acknowledge your collaborators, technicians or others who have contributed to results, as co-authors where appropriate. You apply standards to credit those who helped with collection, management, documentation, publication and archiving of research outputs, so that everyone's expertise is appropriately rewarded. By using standard output identifiers (e.g. DOIs) researcher identifiers (e.g. ORCID) and contributor roles (e.g. the CRediT taxonomy) you also help to make your outputs findable by others.

**Where can you learn?**

**Reuse data from existing sources:**

Basic level:
- Integrating Open Science in Information Literacy education (Foster Open Science: https://www.fosteropenscience.eu/node/2016)
- Github (CODATA)
- Academic Information Seeking (Coursera, https://www.coursera.org/learn/academicinfoseek).
- Discover research data in Europe (CESSDA Training: https://www.cessda.eu/Training/Training-Resources/Data-Discovery-and-Reuse)
- How to search for relevant information on the internet and library resources (UvA, http://gsh.uva.nl/phd-candidates/skills-development-courses/research-skills-library/research-skills-library.html)

**Use or develop FAIR research tools or services:**

Basic level:
- Tutorial: Variable Detection and Linking in Social Sciences Publications (https://www.fosteropenscience.eu/node/2315)
- Introduction to Text and Data Mining (https://www.fosteropenscience.eu/node/2263)
- R (CODATA)
- Github (CODATA)

Intermediate level:
- Self-learn RStudio and Markdown (CODATA)

**Prepare and document data/code to make outputs FAIR:**

Basic level:
- Practicalities of RDM (CODATA)
- Visualization (CODATA)
- R (CODATA)

Intermediate level:
- Advanced R Programming (Coursera, https://www.coursera.org/learn/advanced-r)

**Recognise, cite and acknowledge contributions**

Basic level:
- Academic Information Seeking (Coursera, https://www.coursera.org/learn/academicinfoseek).

| Role Profile | R3 - Established Researcher (developed a level of independence) |
|---|---|

| Key skills | Level | Who may help? |
|---|---|---|
| **Plan for the stewardship and sharing of FAIR outputs** | Expert | Data Steward<br>Data Librarian |

With your in-depth knowledge of funding bodies policy expectations towards research data management, and the application of FAIR principles to other research outputs, you initiate process for managing data, software code, and other outputs across your research team. Taking account of t**he data sources, formats, and research methods your team will use, and disciplinary norms applicable in your research field,** you set the parameters for local guidance, aligning this with data policies and processes of the wider organisation and your funders. You liase with relevant professional services, e.g. data stewards, to develop templates for Data Management Plans and then monitor and support the application of them to research you supervise, including PhD students where appropriate.

| **Recognise, cite and acknowledge contributions** | Expert | Data Steward<br>Data Librarian |
|---|---|---|

From your experience in building collaborative projects and a network of peers you have in-depth understanding of the need for effort to be recognised, cited and acknowledged. Fully aware of research integrity principles and codes of professional research conduct you are able to formulate guidelines for your research group on their practices for citing data, code and methods that they reuse. The students and early career researchers you supervise know how to acknowledge collaborators, technicians or others who have contributed to results, as co-authors where appropriate. You contribute to establishing standards in your domain for crediting those who helped with collection, management, documentation, publication and archiving of research outputs, and can offer guidance on applying standard output identifiers (e.g. DOIs) researcher identifiers (e.g. ORCID) and contributor roles (e.g. the CRediT taxonomy).

| **Secure funding for open science /support** | Intermediate | Data Steward<br>Service Manager<br>Project Manager |
|---|---|---|

From your involvement in obtaining funding for research, and in recruiting staff, you are aware of current moves by funders and institutions towards including criteria for open research in funding and recruitment criteria. You liase with funders regarding priorities for data-intensive research, and understand how to meet and influence their expectations regarding FAIR and open outputs. You are able to budget for costs of research data management and sharing in all grant applications. You seek guidance on costs where appropriate from relevant advisory services.

| **Lead good practice by example** | Intermediate | Data Service Architect<br>R4 Leading Researcher |
|---|---|---|

You draw on your relationship with senior colleagues and encourage them to support you in influencing changes in your research group, including its hiring and promotion decisions. You are able to encourage colleagues and the staff you supervise to adopt practices that will make research outputs as FAIR and open as possible, and to use your influence to ensure they are recognised and rewarded for their efforts in doing so. You contribute to debates in your institution or research field on the application of FAIR principles, research integrity and open science to your context.

**Where can you learn?**

**Recognise, cite and acknowledge contributions**
Basic:
- Academic Information Seeking (Coursera, https://www.coursera.org/learn/academicinfoseek).

**Secure funding for open science /support**
Basic level:
- Entrepeneurship (Coursera, https://www.coursera.org/specializations/wharton-entrepreneurship).
- Framework for Responsible Innovation (Foster Open Science, https://www.fosteropenscience.eu/node/2353).

Intermediate level:
- Open Access Toolkit for RPOs & for Research Funders (Foster Open Science, https://www.fosteropenscience.eu/node/2456).
- Global Financing Solutions (Coursera, https://www.coursera.org/learn/global-financing-solutions).
- Writing Successful Grant Proposals (university Utrecht, https://www.uu.nl/en/events/writing-successful-grant-proposals-online-course-1)


All levels: Research and Innovation research funding courses (University of Plymouth, https://www.plymouth.ac.uk/research/support/funding/r-and-i-research-training)

**Lead good practice by example**
Basic level:
- Strategic Leadership and Management (Coursera, https://www.coursera.org/specializations/strategic-leadership).

Intermediate level:
- Organisational Behaviour: Know your people (Coursera, https://www.coursera.org/learn/organisational-behaviour-know-your-people).

Expert level:
- Relationship Management (Coursera, https://www.coursera.org/learn/relationship-management)

| Role Profile | R4 - Leading Researcher (leading their research area or field) |
|---|---|

| Key skills | Level | Who may help? |
|---|---|---|
| **Apply policies to comply with legal, ethical and FAIR principles** | Expert | Data Steward<br>Service Manager<br>Project Manager |

Being responsible and accountable for your research area's activities and behaviour, you have in-depth understanding of the legal, ethical aspects of research, including principles for making research outputs FAIR. You are able to evaluate and manage the opportunities to demonstrate good practice, as well as the risks to individual and institutional reputations that ethical and/or legal infringement would jeopardise. You demonstrate the knowledge and ability to define the responsibilities of staff or students, and communicate effectively with them and with professional support staff.

| **Recognise, cite and acknowledge contributions** | Intermediate | Data Steward<br>Data Librarian |
|---|---|---|

From your experience in building your organisation's research capabilities you fully understand the need for effort to be recognised, cited and acknowledged. You apply research integrity principles and codes of professional research conduct to formulate guidelines for your research group on their practices for giving recognition.The students and early career researchers you supervise know how to acknowledge collaborators, technicians or others who have contributed to results, as co-authors where appropriate. You are well informed of standards for crediting contributors to collection, management, documentation, publication and archiving of research outputs.

| **Develop open research strategy and vision** | Expert | R3 Established Researcher<br>Service/ project manager<br>Data Librarian<br>Data Service Architect |
|---|---|---|

Based on your insight into how current research policy relates to research practice you are able to formulate your organisation's strategic vision and roadmap for implementing FAIR principles. This includes advising on the degree of openness applicable to research outputs across diverse fields, considering the likely impacts on research excellence, public trust, and the policy and industry networks involved in research. You contribute to working groups dealing with research data strategy in your domain, and engage with the relevant professional support services, including those developing services to support data management.

| **Lead good practice by example** | Expert | R3 Established Researcher<br>Service/ project manager<br>Data Librarian |
|---|---|---|

You draw on your relationship with established researchers in your domain and encourage them to influence changes in in research practice. As a senior member of the research community you demonstrate commitment to recognising open science activities in hiring and promotion decisions. You advocate for changes in practices towards making research outputs as FAIR and open as possible, and as closed as necessary, and influence colleagues to ensure they are recognised and rewarded for their efforts in doing so. You lead debates in your institution and research field on the application of FAIR principles, research integrity and open science.

**Where can you learn?**

**Apply policies to comply with legal, ethical and FAIR principles**
Basic level:
- ORCID and Licensing (CODATA)
- Information security (CODATA)
- Designing Successful Open Access and Open Data Policies: Introductory (Foster Open Science, https://www.fosteropenscience.eu/node/2081).
- Data protection and Ethics (Foster Open Science, https://www.fosteropenscience.eu/node/2330)

Intermediate level:
- Designing Successful Open Access and Open Data Policies: Intermediate (Foster Open Science, https://www.fosteropenscience.eu/node/2075).
- Information security advanced (CODATA)
- Scientific integrity (UvA, http://gsh.uva.nl/phd-candidates/skills-development-courses/scientific-integrity/scientific-integrity.html)

**Recognise, cite and acknowledge contributions**
Basic level:
- Academic Information Seeking (Coursera, https://www.coursera.org/learn/academicinfoseek).

**Lead good practice by example**
Basic level:
- Strategic Leadership and Management (Coursera, https://www.coursera.org/specializations/strategic-leadership).

Intermediate level:
- Organisational Behaviour: Know your people (Coursera, https://www.coursera.org/learn/organisational-behaviour-know-your-people).

Expert level:
- Relationship Management (Coursera, https://www.coursera.org/learn/relationship-management).

## B.4.2  Data Scientists from career stages R1 to R4

| Role Profile | R1 Data Scientist, up to PhD |
|---|---|

| Key skills | Level | Who may help? |
|---|---|---|
| **Reuse data from existing sources** | Intermediate | Data steward<br>R2 Data Scientist |
| Understanding the opportunities that existing sources can offer as raw material for your own research, you can demonstrate the skills to find, access, integrate and reuse data from these sources.  Datasets may come from publicly available repositories in your own domain or from others. They may also come from your network of collaborators. You seek advice from postgraduate data scientists, data steward, data librarian or archivist where appropriate, to broaden the opportunities available to you. You seek their assistance to make data actionable, whether it is sourced externally or from the research team. | | |
| **Use or develop FAIR research tools or services** | Intermediate | Research software engineer |
| You demonstrate the ability to use and develop data analytics applications, algorithms and tools, using machine learning technologies appropriate to the data and domains your research focuses on. You can apply predictive statistical methods relevant to the unfolding nature of the data you derive from these analytic tools.   Aware of the dependencies of your results on specific software code or environments, you ensure results and code are as open as possible, or as closed as necessary to comply with legal obligations. You show you understand how to give and get attribution for the contributions that software authors make to published results. | | |
| **Prepare and document data/code to make outputs FAIR** | Intermediate | Data steward |
| You can identify examples of data and code that offer insights to advance your field, and understand the importance of these being FAIR. You develop effective pipelines for data preparation and pre-processing, You apply provenance standards to ensure a traceable path throughout the data analysis. You use standard formats and identifiers for metadata and data. Using these you demonstrate the application of FAIR principles to gain new research insights and practical application from the integration and reuse of diverse data and computational sources. | | |
| **Recognise, cite and acknowledge contributions** | Intermediate | Data Steward<br>Data Librarian |
| To apply principles of research integrity and professional conduct you properly cite data, code and methods that you reuse. You show that you understand attribution issues affecting text and data mining. You use appropriate methods to cite databases and other forms of dynamic data. When you publish your thesis or dissertation you also acknowledge your collaborators, technicians or others who have contributed to results, as co-authors. You use appropriate identifiers and standards to credit those who helped at all stages of the data lifecycle. | | |

**Where can you learn?**

**Reuse data from existing sources**

Basic level:
- Integrating Open Science in Information Literacy education (Foster Open Science: https://www.fosteropenscience.eu/node/2016)
- Github (CODATA)
- Academic Information Seeking (Coursera, https://www.coursera.org/learn/academicinfoseek).
- Discover research data in Europe (CESSDA Training: https://www.cessda.eu/Training/Training-Resources/Data-Discovery-and-Reuse)
- How to search for relevant information on the internet and library resources (UvA, http://gsh.uva.nl/phd-candidates/skills-development-courses/research-skills-library/research-skills-library.html)

**Use or develop FAIR research tools or services**

Basic level:
- Tutorial: Variable Detection and Linking in Social Sciences Publications (https://www.fosteropenscience.eu/node/2315)
- Introduction to Text and Data Mining (https://www.fosteropenscience.eu/node/2263)
- R (CODATA)
- Github (CODATA)
- Shell (CODATA)

Intermediate level:
- Self-learn RStudio and Markdown (CODATA)
- Other Machine-learning systems (CODATA)
- Artificial Neural Networks (CODATA)
- How to wrap your Java NLP tool into an UIMA component (Foster Open Science, https://www.fosteropenscience.eu/node/2289)

**Prepare and document data/code to make outputs FAIR**

Basic level:
- Practicalities of RDM (CODATA)
- Visualization (CODATA)
- R (CODATA)
- Advanced R Programming (Coursera, https://www.coursera.org/learn/advanced-r)

**Recognise, cite and acknowledge contributions**

Basic level:
- Academic Information Seeking (Coursera, https://www.coursera.org/learn/academicinfoseek).

| Role Profile | R2 Data scientist (PhD holders or equivalent, not yet fully independent) |
|---|---|

| Key skills | Level | Who may help? |
|---|---|---|
| **Plan for the stewardship and sharing of FAIR outputs** | Expert | Data Steward<br>R3 Data Scientist<br>Data Librarian |
| Using your in-depth understanding of data-driven research methods you are able to plan the development of analytic applications, evaluate the range of data management challenges to be expected, and identify solutions that fit the research purpose and level of complexity involved.  Your plan articulates the potential insights and risks of the data intensive research you perform, relating these to ethical and FAIR principles, and funders' policies. In writing your Data Management Plan you seek supervision from your Principal Investigator or other established data scientist. You also liaise with professional services e.g. data stewards, and help peer review plans produced by others. | | |
| **Reuse data from existing sources** | Expert | Data Steward<br>Data Librarian<br>Research Software Engineer |
| You demonstrate expert knowledge and creativity to find, access, integrate and reuse data from novel sources, leading to excellent research, teaching, or non-academic applications.  Datasets may come from publicly available repositories in your own domain or from other reputable sources, including your network of collaborators.  Supporting research students in your team, you also liaise with other professionals where appropriate, to identify new opportunities to assemble data, analytic tools or pipelines from a range of sources. | | |
| **Use or develop FAIR research tools or services** | Expert | Data steward<br>Research Software Engineer |
| You initiate and deliver novel data analytics applications, algorithms and tools, using machine learning technologies appropriate to the data and domains your research focuses on. You can develop predictive statistical methods to exploit novel data types and sources and offer new insights.   Aware of the dependencies of your results on specific software code or environments, you ensure results and code are as open as possible, or as closed as necessary to comply with legal obligations. You contribute to community standards for recognising excellent tools or services and exceed these standards locally. | | |
| **Prepare and document data/code to make outputs FAIR** | Expert | Data steward<br>Research Software Engineer |
| You demonstrate excellence in making the data and code that you use FAIR, and contribute to community guidelines in applying FAIR criteria to these outputs. You develop novel approaches to improving efficiency in data preparation and pre-processing, and to the application of provenance standards in your domains. You contribute to standards in data or metadata formats and apply FAIR principles creatively to the integration and reuse of diverse data and computational sources. | | |

**Where can you learn?**

**Reuse data from existing sources**

Basic level:
- Integrating Open Science in Information Literacy education (Foster Open Science: https://www.fosteropenscience.eu/node/2016)
- Github (CODATA)
- Academic Information Seeking (Coursera, https://www.coursera.org/learn/academicinfoseek).
- Discover research data in Europe (CESSDA Training: https://www.cessda.eu/Training/Training-Resources/Data-Discovery-and-Reuse)
- How to search for relevant information on the internet and library resources (UvA, http://gsh.uva.nl/phd-candidates/skills-development-courses/research-skills-library/research-skills-library.html)

**Use or develop FAIR research tools or services**

Basic level:
- Tutorial: Variable Detection and Linking in Social Sciences Publications (https://www.fosteropenscience.eu/node/2315)
- Introduction to Text and Data Mining (https://www.fosteropenscience.eu/node/2263)
- R (CODATA)
- Github (CODATA)
- Shell (CODATA)

Intermediate level:
- Self-learn RStudio and Markdown (CODATA)
- Other Machine-learning systems (CODATA)
- Artificial Neural Networks (CODATA)
- How to wrap your Java NLP tool into an UIMA component (Foster Open Science, https://www.fosteropenscience.eu/node/2289)

Expert level:
- BabelNet Extractor Tutorial (https://www.fosteropenscience.eu/node/2317).
- Research Computational Infrastructure (CODATA)
- Recommender Systems (CODATA)

**Prepare and document data/code to make outputs FAIR**

Basic level:
- Practicalities of RDM (CODATA)
- Visualization (CODATA)
- R (CODATA)
- Advanced R Programming (Coursera, https://www.coursera.org/learn/advanced-r)

Intermediate level:
- Data Structures and Algorithms Specialization (Coursera, https://www.coursera.org/specializations/data-structures-algorithms).

| Role Profile | R3 Established Data scientist |
|---|---|

| Key skills | Level | Who may help? |
|---|---|---|
| **Plan for the stewardship and sharing of FAIR outputs** | Expert | Data steward |
| Using your in-depth understanding of data-driven research methods you are able to plan the development of analytic applications, evaluate the range of data management challenges to be expected, and identify solutions that fit the research purpose and level of complexity involved.  Your plan articulates the potential insights and risks of data intensive research in your domain, relating these to ethical and FAIR principles, and funders' policies. You give support and supervision in planning data management and stewardship to team members, and contribute to development of support services through your peer network and liaison with senior colleagues in professional services. | | |
| **Apply policies to comply with legal, ethical and FAIR principles** | Expert | Commercialisation advisor Ethics and data protection advisor |
| Being responsible and accountable for your research team's activities and behaviour, you have in-depth understanding of the legal, ethical aspects of research, including principles for making research outputs FAIR. You are able to evaluate and manage the opportunities to demonstrate good practice from data-driven research. You apply knowledge of current data ethics issues and their implications for research integrity in your domain, to manage the risks to individual and institutional reputations that ethical and/or legal infringement would jeopardise. You define the responsibilities of staff or students, and communicate effectively with them and with professional support staff. | | |
| **Publish FAIR outputs on recommended repositories** | Expert | Data librarian |
| Using your in-depth knowledge of repositories for data, code or other outputs (such as samples or protocols) , you liase with data stewards to recommend repositories and other services that are a good fit to the needs of researchers in your area, e.g. the characteristics of their data, and their data management processes.  You provide expert advice on data formats and vocabularies relevant to your research domain, and on the deposition processes of repositories, whether these are locally provided or external. This advice includes appropriate constraints on how FAIR the research outputs will be, depending e.g. on assessment of ethical impacts, data protection, or licensing and IP rights. | | |
| **Secure funding for open science /support** | Expert | Commercialisation advisor |
| From your involvement in obtaining funding for research, and in recruiting staff, you are aware of current moves by funders and institutions towards including criteria for open research in funding and recruitment criteria. You liase with funders regarding priorities for data-intensive research, and understand how to meet and influence their expectations regarding FAIR and open outputs. You are able to budget for costs of research data management and sharing in all grant applications. You seek guidance on costs where appropriate from relevant advisory services. | | |

**Where can you learn?**

**Plan for the stewardship and sharing of FAIR outputs**

Basic level:

- Introduction to Research Data Management and Data Management Plans (CODATA)
- FAIR Sharing (CODATA)
- Managing and sharing research data (https://www.fosteropenscience.eu/node/2328)

Intermediate level:

- How to get your article published (https://www.egsh.eur.nl/doctoral-education/phd-course-guide/how-to-get-your-article-published/)

**Apply policies to comply with legal, ethical and FAIR principles**

Basic level:

- ORCID and Licensing (CODATA)
- Information security (CODATA)
- Designing Successful Open Access and Open Data Policies: Introductory (Foster Open Science, https://www.fosteropenscience.eu/node/2081).
- Data protection and Ethics (Foster Open Science, https://www.fosteropenscience.eu/node/2330)

Intermediate level:

- Information security advanced (CODATA)
- Scientific integrity (UvA, http://gsh.uva.nl/phd-candidates/skills-development-courses/scientific-integrity/scientific-integrity.html)

**Secure funding for open science /support**

Basic level:

- Entrepeneurship (Coursera, https://www.coursera.org/specializations/wharton-entrepreneurship).
- Framework for Responsible Innovation (Foster Open Science, https://www.fosteropenscience.eu/node/2353).

Intermediate level:

- Open Access Toolkit for RPOs & for Research Funders (Foster Open Science, https://www.fosteropenscience.eu/node/2456).
- Writing Successful Grant Proposals (university Utrecht, https://www.uu.nl/en/events/writing-successful-grant-proposals-online-course-1)

| Role Profile | R4 Leading Data Scientist |
|---|---|

| Key skills | Level | Who may help? |
|---|---|---|
| **Apply policies to comply with legal, ethical and FAIR principles** | Expert | Commercialisation advisor<br>Ethics and data protection advisor |
| Being responsible and accountable for your research team's activities and behaviour, you have in-depth understanding of the legal, ethical aspects of research, including principles for making research outputs FAIR. You are able to evaluate and manage the opportunities to demonstrate good practice from data-driven research. You apply knowledge of current data ethics issues and their implications for research integrity in your domain, to manage the risks to individual and institutional reputations that ethical and/or legal infringement would jeopardise. You define the responsibilities of staff or students, and communicate effectively with them and with professional support staff. | | |
| **Develop open research strategy and vision** | Expert | R3 Established Researcher<br>Service/ project manager<br>Data Librarian<br>Data Service Architect |
| Based on your insight into how current research policy relates to practice in data-intensive research you are able to formulate your organisation's strategic vision and roadmap for implementing FAIR principles. This includes advising on the degree of openness applicable to research outputs across diverse fields, considering the likely impacts on research excellence, public trust, and the policy and industry networks involved in research. You contribute to working groups dealing with research data strategy in your domain, and engage with the relevant professional support services, including those developing services to support data management. | | |
| **Secure funding for open science /support** | Expert | Commercialisation advisor |
| From your involvement in obtaining funding for research, and in recruiting staff, you are aware of current moves by funders and institutions towards including criteria for open research in funding and recruitment criteria. You liase with funders regarding priorities for data-intensive research, and understand how to meet and influence their expectations regarding FAIR and open outputs. You are able to budget for costs of research data management and sharing in all grant applications. You seek guidance on costs where appropriate from relevant advisory services. | | |
| **Lead good practice by example** | Expert | R3 Established Researcher<br>Service/ project manager<br>Data Librarian |
| You draw on your relationship with established researchers in your domain and encourage them to influence changes in in research practice. As a senior member of the research community you demonstrate commitment to recognising open science activities in hiring and promotion decisions. You advocate for changes in practices towards making research outputs as FAIR and open as possible, and as closed as necessary, and influence colleagues to ensure they are recognised and rewarded for their efforts in doing so. You lead debates in your institution and research field on the application of FAIR principles, research integrity and open science. | | |

**Where can you learn?**

**Apply policies to comply with legal, ethical and FAIR principles**
Basic level:
- ORCID and Licensing (CODATA)
- Information security (CODATA)
- Designing Successful Open Access and Open Data Policies: Introductory (Foster Open Science, https://www.fosteropenscience.eu/node/2081).
- Data protection and Ethics (Foster Open Science, https://www.fosteropenscience.eu/node/2330)

Intermediate level:
- Designing Successful Open Access and Open Data Policies: Intermediate (Foster Open Science, https://www.fosteropenscience.eu/node/2075).
- Information security advanced (CODATA)
- Scientific integrity (UvA, http://gsh.uva.nl/phd-candidates/skills-development-courses/scientific-integrity/scientific-integrity.html)

**Key skill 7**
Basic level:
- Academic Information Seeking (Coursera, https://www.coursera.org/learn/academicinfoseek).

**Secure funding for open science /support** Basic level:
- Entrepeneurship (Coursera, https://www.coursera.org/specializations/wharton-entrepreneurship).
- Framework for Responsible Innovation (Foster Open Science, https://www.fosteropenscience.eu/node/2353).

Intermediate level:
- Open Access Toolkit for RPOs & for Research Funders (Foster Open Science, https://www.fosteropenscience.eu/node/2456).
- Global Financing Solutions (Coursera, https://www.coursera.org/learn/global-financing-solutions).
- Writing Successful Grant Proposals (university Utrecht, https://www.uu.nl/en/events/writing-successful-grant-proposals-online-course-1)

**All levels**: Research and Innovation research funding courses (University of Plymouth, https://www.plymouth.ac.uk/research/support/funding/r-and-i-research-training).

## B.4.3    Data advisory roles

| Role Profile | Data Librarian - Institutional level |
| --- | --- |

| Key skills | Level | Who may help? |
| --- | --- | --- |
| **Use or develop FAIR research tools or services** | Intermediate | Research Software Engineer<br>Data Service Architect<br>Data Steward |
| You have in-depth knowledge of relevant (inter)national data and e-infrastructure services for making data FAIR, and keeping it FAIR.  You identify and respond to the needs for services and resources, ensuring those provided are also FAIR, and as open as possible to reuse by researchers affiliated with the organisation, and by other users according to organisational policy and legal obligations. You take ownership responsibility for the development, set-up and update of suitable support tools or services, monitor availability and use, and ensure delivery of training on their effective application. Liaising with relevant professional services, e.g. research software engineers and data service architects, you ensure the organisation's service management processes are followed. | | |
| **Apply policies to comply with legal, ethical and FAIR principles** | Expert | Service / Project Manager<br>Ethics and data protection advisor<br>Data Steward |
| You have broad and deep understanding of your organisation's research strategy and how FAIR data and software outputs contribute to its implementation. You have in-depth knowledge of legal and ethical frameworks applicable to research communities you support, and of the role of FAIR data in underpinning research integrity. Liaising with senior stakeholders and your organisation's professional services, you formulate RDM policy to align with relevant funder policies, including for Data Management Plans. You establish systems and processes, including advice and training, to ensure data management is aligned with processes for research ethics, data protection and academic conduct. | | |
| **Prepare and document data/code to make outputs FAIR** | Expert | Data Steward<br>Ethics and data protection advisor<br>Legal compliance, commercialisation advisor |
| Your have expert knowledge of relevant (inter)national data and e-infrastructure services and provide advice on their alignment with current standards and resources available from the organisation to assist researchers in making their outputs FAIR.  You ensure guidance and instruction is available on making data findable, and that standard solutions are available for cleaning data to make it actionable and interoperable for reuse.  With in-depth knowledge of ethical and commercial constraints on data access, you monitor liaison with relevant professional services and systems across the organisation.  You ensure expert advice is available on appraising and selecting data of value for reuse, and on good practice in managing software code, including versioning and documentation. | | |

| Develop open research strategy and vision | Intermediate | Service Manager / Project Manager, Data Service Architect |
| --- | --- | --- |

Based on your insight into how current and emerging infrastructure and tools may support implementation of RDM policies in the research areas you support, you formulate a strategy or roadmap for implementing FAIR principles through participation in relevant committees and working groups. Your advise on the development of standard solutions for recurring RDM issues and monitor the availability of support. With extensive knowledge of relevant skills and the resources available to develop skills, you initiate and support training and recruitment of colleagues in professional services to support data management.

**Where can you learn?**

**Use or develop FAIR research tools or services**
Basic level:
- Tutorial: Variable Detection and Linking in Social Sciences Publications (https://www.fosteropenscience.eu/node/2315)
- Introduction to Text and Data Mining (https://www.fosteropenscience.eu/node/2263)
- R (CODATA)
- Github (CODATA)
- Shell (CODATA)

Intermediate level:
- Self-learn RStudio and Markdown (CODATA)
- Other Machine-learning systems (CODATA)
- Artificial Neural Networks (CODATA)

Expert level
- BabelNet Extractor Tutorial (https://www.fosteropenscience.eu/node/2317).
- Research Computational Infrastructure (CODATA)
- Recommender Systems (CODATA)

**Prepare and document data/code to make outputs FAIR**
Basic level:
- Practicalities of RDM (CODATA)
- Visualization (CODATA)
- R (CODATA)

Intermediate level:
- Advanced R Programming (Coursera, https://www.coursera.org/learn/advanced-r)
- Data Structures and Algorithms Specialization (Coursera, https://www.coursera.org/specializations/data-structures-algorithms).

| Role Profile | Data Steward  - Operational level e.g. research group or department/ faculty |
|---|---|

| Key skills | Level | Who may help? |
|---|---|---|
| **Plan for the stewardship and sharing of FAIR outputs** | Expert | Data Librarian |

You have in-depth understanding of your organisation's RDM policy, and of the policies of relevant research funders in relation to planning for FAIR outputs, including Data Management Plans. You have in-depth knowledge of  the legal and ethical frameworks applicable to the research communities you support. You are also familiar with these communities' research methods, tools, standards and services used locally.   You are able to translate this knowledge and understanding into Data Management Plan (DMP) templates for the researchers you support, providing expert advice appropriate to  all career stages.  You can articulate how planning benefits the researcher as the 'first reuser' of their outputs, and seek help where needed from other professional services**.**

| Key skills | Level | Who may help? |
|---|---|---|
| **Use or develop  FAIR research tools or services** | Intermediate | Data Librarian<br>Research Software Engineer<br>Data Service Architect |

Aware of relevant (inter)national data and e-infrastructure services you identify and respond to the needs of researchers and stakeholders in the areas you support for FAIR tools, services or resources. You provide guidance and instruction on discovery, acquisition and (re-)use of data, including local collection and those in the public domain. You ensure these are as open possible to reuse, according to organisational policy and legal obligations. Liaising with relevant professional services, e.g. research software engineers and data service architects, you develop and implement standard solutions for recurring issues and deliver training on their effective application, also ensuring that the organisation's service management processes are followed.

| Key skills | Level | Who may help? |
|---|---|---|
| **Prepare and document data/code to make outputs FAIR** | Expert | Data Librarian<br>Research Software Engineer |

You provide researchers in fields you support with examples of FAIR and open data or software that advances research, and describe the relevance of FAIR principles to the norms for validation or reproducibility in their domains . You provide expert advice on making data findable, and on cleaning data to make it actionable and interoperable for reuse. Knowledgeable about ethical and commercial constraints on data access, you liaise with relevant professional services and systems.  You provide expert advice on appraising and selecting data of value for reuse, and on good practice in managing software code, including versioning and documentation.

| Publish FAIR outputs on recommended repositories | Expert | Data Librarian Project/ Service Manager |
|---|---|---|

Using your in-depth knowledge of repositories for data, code or other outputs (such as samples or protocols) , you direct researchers in your area to these repositories and other services that are a good fit to their domain, the characteristics of their data, and their data management processes.  You provide expert advice to researchers in your area on data formats and vocabularies relevant to their research domains, and on the deposition processes of repositories, whether these are locally provided or external. You liaise with relevant professional services to ensure appropriate constraints on how FAIR the research outputs will be, depending e.g. on assessment of ethical impacts, data protection, or licensing and IP rights.

**Where can you learn?**
**Plan for the stewardship and sharing of FAIR outputs**
Basic level:
- Introduction to Research Data Management and Data Management Plans (CODATA)
- FAIR Sharing (CODATA)
- Managing and sharing research data (https://www.fosteropenscience.eu/node/2328)

Intermediate level:
- How to get your article published (https://www.egsh.eur.nl/doctoral-education/phd-course-guide/how-to-get-your-article-published/)

**Use or develop  FAIR research tools or services**
Basic level:
- Tutorial: Variable Detection and Linking in Social Sciences Publications (https://www.fosteropenscience.eu/node/2315)
- Introduction to Text and Data Mining (https://www.fosteropenscience.eu/node/2263)
- R (CODATA)
- Github (CODATA)

**Prepare and document data/code to make outputs FAIR** Basic level:
- Practicalities of RDM (CODATA)
- Visualization (CODATA)
- R (CODATA)

**Publish FAIR outputs on recommended repositories**
Basic level:
- FAIR sharing (CODATA)
- Practicalities of RDM (CODATA)
- Publishing to GitHub and Rstudio (CODATA)
- Be Persuasive: Write a Convincing Position Paper or Policy Advice (Project-Centered Course from Coursera, https://www.coursera.org/learn/persuasive-writing)

## B.5. Skills Tables

Legend: ○ = basic, ◖ = intermediate, ● = expert, - = none

**Table 11: FAIR4S skills table – researchers & data scientists**

| FAIR4S competences and capabilities | | Recommended expertise by professional group and service role | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Researchers | | | | Data Scientists | | | |
| | | R1 | R2 | R3 | R4 | R1 | R2 | R3 | R4 |
| **Plan and design** | **Plan stewardship and sharing of FAIR outputs** | ◖ | ◖ | ○ | ● | ● | ● | ○ | ● |
| | Model data structures and define database needs | ○ | ● | ● | ○ | ◖ | ◖ | ◖ | ● |
| | Specifying metadata and persistent id. standards | ○ | ◖ | ○ | ◖ | ○ | ○ | ◖ | ◖ |
| | Analyse requirements for services or software | ○ | ● | ● | ◖ | ◖ | ● | ◖ | ● |
| | Appraise and select repositories for FAIR sharing | ○ | ◖ | ○ | ● | ◖ | ○ | ◖ | ● |
| **Capture and process** | **Reuse data from existing sources** | ◖ | ◖ | ○ | ● | ○ | ○ | ○ | ○ |
| | Manage databases | - | ◖ | ● | ○ | ○ | ◖ | ● | ◖ |
| | Software prototyping | - | ◖ | ● | ○ | ◖ | ● | ● | ◖ |
| | Set up and document workflows | ◖ | ● | ○ | ◖ | ● | ○ | ○ | ○ |
| | File naming and organisation | ◖ | ◖ | ◖ | ● | ◖ | ◖ | ◖ | ◖ |
| | Data cleaning, processing and software versioning | ◖ | ● | ● | ● | ◖ | ● | ● | ◖ |
| **Integrate and analyse** | **Use or develop FAIR research tools/services** | ◖ | ● | ◖ | ○ | ◖ | ○ | ○ | ○ |
| | Critical thinking and theory building | ◖ | ● | ◖ | ○ | ◖ | ○ | ○ | ○ |
| | Creative problem solving, flexibility | ◖ | ● | ● | ◖ | ◖ | ● | ○ | ○ |
| | Math and statistical knowledge application | ○ | ● | ● | ◖ | ◖ | ● | ● | ● |
| | Data transformation and integration | ○ | ● | ● | ◖ | ◖ | ● | ○ | ○ |
| | Data mining, querying, interpretation | ○ | ● | ○ | ○ | ○ | ● | ○ | ○ |
| | Predictive modelling and analytics | ○ | ● | ○ | ○ | ○ | ○ | ○ | ○ |
| | Machine learning methods | ○ | ● | ○ | ○ | ○ | ○ | ○ | ○ |

| FAIR4S competences and capabilities | | Recommended expertise by professional group and service role | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Researchers | | | | Data Scientists | | | |
| | | R1 | R2 | R3 | R4 | R1 | R2 | R3 | R4 |
| **Appraise and preserve** | **Prepare and document data/code to make outputs FAIR** | ◖ | ◖ | ○ | ● | ● | ● | ○ | ◖ |
| | Data quality assurance using open standards | ◖ | ◖ | ○ | ● | ◖ | ◖ | ◖ | ◖ |
| | Assess FAIRness and check policy compliance | ◖ | ◖ | ◖ | ● | ◖ | ◖ | ◖ | ◖ |
| | Data transfer and long-term storage | ◖ | ◖ | ● | ◖ | ○ | ◖ | ◖ | ◖ |
| | Format and media migration | ◖ | ○ | ● | ● | ○ | ○ | ◖ | ◖ |
| | Software review and preservation | ○ | ◖ | ● | ○ | ○ | ◖ | ◖ | ◖ |
| **Publish and release** | **Publish FAIR outputs on recommended repositories** | ○ | ◖ | ● | ○ | ● | ◖ | ◖ | ◖ |
| | Access control and management | ○ | ● | ● | ◖ | ○ | ● | ◖ | ◖ |
| | Ethical application of patents, licences | ◖ | ◖ | ◖ | ◖ | ◖ | ◖ | ◖ | ◖ |
| | Open access publishing and self-archiving | ◖ | ◖ | ○ | ● | ○ | ○ | ○ | ○ |
| | Documentation for public use, lay understanding | ◖ | ◖ | ◖ | ● | ○ | ○ | ◖ | ◖ |
| | Engaging in open innovation beyond academia | ◖ | ◖ | ◖ | ○ | ◖ | ◖ | ○ | ○ |
| **Expose and discover** | **Recognise, cite and acknowledge contributions** | ◖ | ● | ◖ | ○ | ○ | ◖ | ◖ | ○ |
| | Metadata and persistent id. exposure | ◖ | ● | ◖ | ● | ○ | ● | ○ | ○ |
| | Visualisation and presentation of results | ◖ | ● | ◖ | ○ | ○ | ○ | ○ | ○ |
| | Evaluation of repository and publishing platforms | ◖ | ◖ | ○ | ● | ◖ | ○ | ○ | ○ |
| | Searching repositories and scientific databases | ◖ | ◖ | ● | ● | ○ | ◖ | ○ | ○ |
| | Vocabulary/ ontology application | ◖ | ◖ | ◖ | ● | ○ | ○ | ○ | ○ |
| **Govern and assess** | **Develop open research strategy and vision** | ◖ | ● | ◖ | ● | ◖ | ○ | ○ | ○ |
| | **Apply policies to comply legal, ethical & FAIR principles** | ◖ | ◖ | ◖ | ● | ◖ | ○ | ○ | ○ |
| | Research integrity, attribution, impact awareness | ◖ | ◖ | ○ | ● | ◖ | ○ | ○ | ○ |
| | Information security and risk management | ◖ | ◖ | ◖ | ● | ◖ | ◖ | ◖ | ◖ |
| | Data governance, handling third-party data | ◖ | ◖ | ◖ | ◖ | ◖ | ◖ | ◖ | ◖ |
| | Storage security management | - | ◖ | ● | ◖ | ◖ | ◖ | ◖ | ◖ |
| | Contributing to quality assessment or peer review | ◖ | ◖ | ◖ | ○ | ○ | ◖ | ○ | ○ |

| FAIR4S competences and capabilities | | Recommended expertise by professional group and service role | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | Researchers | | | | Data Scientists | | | |
| | | R1 | R2 | R3 | R4 | R1 | R2 | R3 | R4 |
| **Scope and resource** | **Secure funding for open science /support** | ◐ | ◐ | ◐ | ◐ | ◐ | ○ | ○ | ○ |
| | Perseverance delivering diverse open research projects | ◐ | ◐ | ◐ | ◐ | ◐ | ○ | ◐ | ● |
| | Service level management | - | ○ | ○ | ○ | ◐ | ○ | ◐ | ● |
| | Change management | ○ | ○ | ○ | ○ | ○ | ○ | ◐ | ● |
| | Workflow set-up and provenance information mgmt | - | ● | ● | ◐ | ○ | ● | ◐ | ● |
| | Cloud environment and storage management | ○ | ◐ | ● | ◐ | ○ | ● | ● | ● |
| | Authentication and authorisation (AAI) management | ○ | ○ | ● | ◐ | ○ | ○ | ○ | ● |
| | Costing of data management and preservation | ◐ | ◐ | ○ | ● | ○ | ○ | ○ | ● |
| **Advise and enable** | **Lead good practice by example** | ◐ | ◐ | ◐ | ○ | ● | ◐ | ◐ | ◐ |
| | Engaging with research users and stakeholders | ○ | ◐ | ◐ | ● | ◐ | ◐ | ◐ | ◐ |
| | Developing a profile of open research | ◐ | ● | ● | ● | ◐ | ○ | ○ | |
| | Training in open methods, services | ◐ | ◐ | ◐ | ◐ | ◐ | ○ | ○ | ◐ |
| | Contributing to education, professional development | ◐ | ◐ | ◐ | ◐ | ○ | ○ | ○ | ○ |
| | Contributing to open research, networks, standards bodies | ◐ | ● | ● | ● | ○ | ◐ | ○ | ○ |
| | Building open inter-disciplinary collaborations | ○ | ○ | ○ | ○ | ◐ | ◐ | ◐ | ◐ |

Table 12: FAIR4S skills table – data advisors and data service providers

| FAIR4S competences and capabilities | | Recommended expertise by professional group and service role | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | Data Advisors | | | | Data Service providers | | | | |
| | | Data Steward | Research Mgr | Ethics & DP | Comm. & IP | Data Librarian | Research Swr Eng | Service Mgr | Service Architect | Archivist |
| **Plan and design** | **Plan stewardship and sharing of FAIR outputs** | ● | ◖ | ◖ | ◖ | ◖ | ◖ | ○ | ○ | ○ |
| | Model data structures and define database needs | ◖ | ○ | ○ | ○ | ◖ | ● | ○ | ● | ◖ |
| | Specifying metadata and persistent id. standards | ◖ | ○ | ○ | ○ | ○ | ◖ | ○ | ◖ | ● |
| | Analyse requirements for services or software | ○ | ○ | ○ | ○ | ◖ | ● | ○ | ● | ◖ |
| | Appraise and select repositories for FAIR sharing | ● | ○ | ○ | ○ | ● | ○ | ○ | ○ | ● |
| **Capture and process** | **Reuse data from existing sources** | ◖ | ○ | ○ | ○ | ● | ◖ | ○ | ○ | ● |
| | Manage databases | ◖ | ○ | ○ | ○ | ◖ | ● | ○ | ● | ◖ |
| | Software prototyping | ◖ | ○ | ○ | ○ | ○ | ● | ○ | ● | ○ |
| | Set up and document workflows | ◖ | ○ | ○ | ○ | ◖ | ◖ | ○ | ◖ | ◖ |
| | File naming and organisation | ◖ | ○ | ○ | ○ | ● | ◖ | ○ | ◖ | ● |
| | Data cleaning, processing and software versioning | ◖ | ○ | ○ | ○ | ◖ | ● | ○ | ◖ | ● |
| **Integrate and analyse** | **Use or develop FAIR research tools/services** | ◖ | ◖ | ◖ | ○ | ◖ | ● | ◖ | ◖ | ◖ |
| | Critical thinking and theory building | ○ | ○ | ○ | ○ | ○ | ◖ | ○ | ● | ○ |
| | Creative problem solving, flexibility | ◖ | ◖ | ○ | ○ | ◖ | ◖ | ○ | ● | ◖ |
| | Math and statistical knowledge application | ○ | ○ | ○ | ○ | ○ | ◖ | ○ | ○ | ○ |
| | Data transformation and integration | ◖ | ○ | ○ | ○ | ● | ● | ○ | ◖ | ● |
| | Data mining, querying, interpretation | ◖ | ○ | ○ | ○ | ◖ | ◖ | ○ | ◖ | ◖ |
| | Predictive modelling and analytics | ◖ | ○ | ○ | ○ | ○ | ◖ | ○ | ○ | ○ |
| | Machine learning methods | ○ | ○ | ○ | ○ | ○ | ◖ | ○ | ○ | ○ |

| FAIR4S competences and capabilities | | Recommended expertise by professional group and service role | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Data Advisors | | | | Data Service providers | | | | |
| | | Data Steward | Research Mgr | Ethics & DP | Comm. & IP | Data Librarian | Research Swr Eng | Service Mgr | Service Architect | Archivist |
| **Appraise and preserve** | **Prepare,document data/code to make outputs FAIR** | ● | ○ | ◖ | ○ | ◖ | ◖ | ○ | ○ | ● |
| | Data quality assurance using open standards | ◖ | ○ | ○ | ○ | ● | ◖ | ○ | ◖ | ● |
| | Assess FAIRness and check policy compliance | ◖ | ◖ | ● | ◖ | ◖ | ◖ | ◖ | ○ | ◖ |
| | Data transfer and long-term storage | ◖ | ○ | ◖ | ○ | ◖ | ◖ | ○ | ○ | ● |
| | Format and media migration | ◖ | ○ | ○ | ○ | ● | ◖ | ○ | ○ | ● |
| | Software review and preservation | ◖ | ○ | ○ | ◖ | ◖ | ◖ | ○ | ○ | ● |
| **Publish and release** | **Publish FAIR outputs on recommended repositories** | ● | ◖ | ◖ | ◖ | ● | ◖ | ○ | ○ | ● |
| | Access control and management | ◖ | ◖ | ◖ | ◖ | ◖ | ◖ | ○ | ◖ | ◖ |
| | Ethical application of patents, licences | ○ | ◖ | ◖ | ● | ◖ | ◖ | ◖ | ◖ | ◖ |
| | Open access publishing and self-archiving | ◖ | ○ | ◖ | ○ | ● | ◖ | ○ | ○ | ◖ |
| | Documentation for public use, lay understanding | ○ | ○ | ○ | ○ | ○ | ◖ | ○ | ○ | ◖ |
| | Engaging in open innovation beyond academia | ○ | ○ | ○ | ● | ◖ | ◖ | ◖ | ○ | ◖ |
| **Expose and discover** | **Recognise, cite and acknowledge contributions** | ◖ | ◖ | ○ | ◖ | ● | ◖ | ○ | ○ | ● |
| | Metadata and persistent id. exposure | ◖ | ○ | ○ | ○ | ● | ◖ | ○ | ○ | ● |
| | Visualisation and presentation of results | ◖ | ○ | ○ | ○ | ○ | ◖ | ○ | ○ | ◖ |
| | Evaluation of repository and publishing platforms | ● | ○ | ◖ | ○ | ● | ○ | ○ | ○ | ◖ |
| | Searching repositories and scientific databases | ◖ | ○ | ○ | ○ | ● | ◖ | ○ | ○ | ◖ |
| | Vocabulary/ ontology application | ○ | ○ | ○ | ○ | ◖ | ◖ | ○ | ○ | ◖ |

| FAIR4S competences and capabilities | | Recommended expertise by professional group and service role | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Data Advisors | | | | Data Service providers | | | | |
| | | Data Steward | Research Mgr | Ethics & DP | Comm. & IP | Data Librarian | Research Swr Eng | Service Mgr | Service Architect | Archivist |
| **Scope and resource** | **Secure funding for open science /support** | ○ | ● | ○ | ○ | ● | ◖ | ◖ | ○ | ○ |
| | Perseverance delivering diverse open research projects | ◖ | ◖ | ○ | ◖ | ◖ | ◖ | ◖ | ◖ | ◖ |
| | Service level management | ◖ | ◖ | ○ | ○ | ◖ | ◖ | ● | ◖ | ◖ |
| | Change management | ○ | ◖ | ○ | ○ | ◖ | ◖ | ● | ◖ | ◖ |
| | Workflow set-up and provenance information mgmt | ◖ | ○ | ○ | ○ | ◖ | ◖ | ● | ◖ | ◖ |
| | Cloud environment and storage management | ○ | ○ | ◖ | ○ | ◖ | ◖ | ◖ | ◖ | ● |
| | Authentication and authorisation (AAI) management | ○ | ○ | ○ | ○ | ◖ | ◖ | ◖ | ◖ | ◖ |
| | Costing of data management and preservation | ◖ | ◖ | ○ | ◖ | ● | ◖ | ◖ | ○ | ◖ |
| **Advise and enable** | **Lead good practice by example** | ◖ | ◖ | ○ | ○ | ● | ○ | ◖ | ○ | ◖ |
| | Engaging with research users and stakeholders | ● | ◖ | ◖ | ◖ | ◖ | ◖ | ◖ | ○ | ◖ |
| | Developing a profile of open research | ◖ | ◖ | ○ | ○ | ◖ | ◖ | ○ | ○ | ◖ |
| | Training in open methods, services | ● | ◖ | ◖ | ○ | ○ | ◖ | ○ | ○ | ◖ |
| | Contributing to education, professional development | ● | ◖ | ◖ | ○ | ◖ | ◖ | ◖ | ○ | ◖ |
| | Contributing to open research, networks, standards bodies | ◖ | ◖ | ◖ | ◖ | ◖ | ◖ | ◖ | ◖ | ◖ |
| | Building open inter-disciplinary collaborations | ◖ | ◖ | ○ | ◖ | ◖ | ◖ | ◖ | ○ | ◖ |