D7.3: Skills and Capability Framework

Author(s)	Angus Whyte, Jerry de Vries, Rahul Thorat, Eileen Kuehn, Gergely Sipos, Valentino Cavalli, Vasso Kalaitzi, Kevin Ashley
Status	Final
Version	V1.0
Date	2018-07-04

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:

The report offers a Framework to help organisations to plan the professional development of their staff, as EOSC service operators or users, and for any individual to identify competences and learning materials that match the capabilities they need. It provides a set of core competences for data stewardship, relating topics to recommended expertise levels for researchers and the professional groups that support them. The Framework also offers examples of capability and competence statements, focusing these on skills areas that we have identified as gaps for stakeholders. The Framework provides an approach to describing similar competence and capability statements that will be of continued use as EOSC services evolve. It is of use to service operators and others with an interest in skills development. Conclusions are provided about the need to consult further on Research Infrastructure and other stakeholder expectations about the role of the Skills Framework in scoping and supporting the generic skills needed to enable the EOSC. This consultation is planned with EOSCpilot partners and Research Infrastructure training coordinators to inform the final recommendations of WP7.

The European Open Science Cloud for Research pilot project (EOSCpilot) is funded by the European Commission, DG Research & Innovation under contract no. 739563

Document identifier: EOSCpilot -WP7-D7.3	
Deliverable lead	**UEDIN-DCC**
Related work package	**WP7**
Author(s)	Angus Whyte, Jerry de Vries, Ellen Leenarts, Rahul Thorat, Eileen Kuehn, Gergely Sipos, Valentino Cavalli, Vasso Kalaitzi, Kevin Ashley
Contributor(s)	
Due date	**30/06/2018**
Actual submission date	**06/07/2018**
Reviewed by	**Mark Thorley; Celia van Gelder**
Approved by	**Brian Matthews**
Start date of Project	**01/01/2017**
Duration	**24 months**

Versioning and contribution history

Version	Date	Authors	Notes
0.1	8/05/2018	Angus Whyte (UEDIN)	ToC Description
0.2	14/05/2018	Angus Whyte (UEDIN), Jerry de Vries (DANS), Eileen Kuehn (KIT)	Revised competences, mapping to OS-CAM, initial competence and capability statements
0.3	01/06/2018	Angus Whyte, Jerry de Vries, Rahul Thorat, Eileen Kuehn, Gergely Sipos, Valentino Cavalli, Vasso Kalaitzi	Skills user stories, further competence and capability statements, initial drafts of narrative sections
0.4	14/06/2018	Angus Whyte, Jerry de Vries, Ellen Leenarts, Rahul Thorat, Eileen Kuehn, Gergely Sipos, Valentino Cavalli, Vasso Kalaitzi	Further iterations of above and responses to initial comments.
0.5	28/06/2018	Angus Whyte, Jerry De Vries, Eileen Kuehn, Valentino Cavalli, Vasso Kalaitzi, Kevin Ashley, Gergely Sipos	Response to internal reviewer comments; completion of introduction, conclusions, rewrite of exec summary
1.0	4/07/2018	Angus Whyte; Kevin Ashley	Final reformatting and reworking of figures, annexes, abstract and conclusions; application of standard template.

TABLE OF CONTENT

EXECUTIVE SUMMARY **6**

1. INTRODUCTION **7**

1.1. Overall objectives in the project context 7
1.2. Purpose of this document 7
1.3. Relationship to other project outputs 7
1.4. Structure of this document 8

2. THE SKILLS FRAMEWORK IN CONTEXT **9**

2.1. The Need for an EOSC Skills and Capability Framework 9
2.2. Overview of the Framework 11
2.2.1. Audience and target group 11
2.2.2. Structure 12
2.3. Linking EOSC Capabilities to Skills and Competences 15
2.4. Conceptual model 19

3. USE CASES AND APPLICATION OF THE FRAMEWORK **21**

3.1. A framework for service operators to specify skills requirements 21
3.1.1. Finding relevant capabilities 21
3.1.2. Expanding the capabilities 22
3.1.3. Identifying the competence dimensions 22
3.1.4. Specifying the skills 22
3.2. A framework for EOSC users to identify learning resources 23
3.3. A framework for research leads and human resource professionals to plan skills development 25
3.3.1. Describing service-specific roles 25
3.3.2. Using job descriptions 26

4. DEVELOPMENT OF THE MODEL **29**

4.1. Principles used in drafting the framework 29
4.1.1. Recognise FAIRness spans multiple data lifecycles 29
4.1.2. Recognise the policy and organizational context 29
4.1.3. Recognise disciplinary norms in stewardship responsibility 29
4.2. Competences used in the draft framework 29
4.2.1. EDISON 30
4.2.2. Competency Matrix for Data Management Skills 30
4.3. Service-related capability models 30
4.4. Synthesis and revisions to the Framework 31
4.5. Skills gaps from science demonstrator outcomes 32
4.6. Key workshop outcomes and responses in the framework 33

5. CONCLUSIONS **34**

ANNEX A. EOSCPILOT SKILLS AND CAPABILITY FRAMEWORK **37**

ANNEX B. SKILLS AND CAPABILITY EXAMPLES FOR EOSC SERVICES **43**

ANNEX C. MAPPINGS TO OTHER FRAMEWORKS **56**

ANNEX D. GLOSSARY **63**

LIST OF FIGURES

Figure 1 - Professional Groups....12
Figure 2 - Core Skills Groups & Dimensions....13
Figure 3 - Skills Framework Dimensions....13
Figure 4 - Skills Groups....14
Figure 5 – Service Classes; Connections; Access Channels....17
Figure 6 - Conceptual model for EOSCpilot skills framework....20
Figure 7 - Overview of three step procedure to identify relevant learning resources....24
Figure 8 - Mapping vocabulary from user stories to EOSCpilot Framework to training metadata....25
Figure 9 - Job description....26
Figure 10 - Responsibility levels in the EOSCpilot Skills Framework....27

LIST OF TABLES

Table 1 – Scope Of the Skills Groups....15
Table 2 - Service Functions....18
Table 3 – Service roles, activities and knowledge....19
Table 4 – Skills specification entry for a repository service in an EOSC service catalogue....24
Table 5 - Job description analysis....28

EXECUTIVE SUMMARY

The importance of skills development amongst researchers and EOSC service providers has been highlighted by a number of expert reports and European Commission statements in recent years. These calls for action are summarised in section 2 of this report. The EOSCpilot project plan recognised the importance of these issues, establishing a work package on skills and capabilities alongside those dealing with matters such as policy, service definitions, interoperability and science demonstrators. One of the primary goals of this work package was the development of a Skills and Capability Framework which could be of use to many actors in the EOSC - researchers, service providers, the organisations that employ them and those who fund them, amongst others. This document provides the first full version of this Framework in Annex A, and builds on earlier work within and outside of the EOSCpilot project.

The Skills and Capability Framework presented in this report aims to help research infrastructures and user organisations to recognise the skills demanded of researchers and support professionals, and to relate them to the organizational capabilities needed to provide EOSC services and ensure they are effective tools in skilled hands. The report synthesizes existing competence frameworks for data science and data management, for the EOSC environment of services to support open science. The focus is on the skills needed to perform open science, and also to provide the service infrastructures to support open science. There is a special focus on ensuring the stewardship of research outputs according to FAIR principles.

The Framework describes competences, which are properties of an individual, and capabilities, which are the properties of a group or organisation wishing to use or provide services. In this document, we define skills as the application of competences and capabilities in a specific context, such as the deployment of a service. We show how the Framework can be used to answer questions such as what skills a researcher needs to use a service, what expectations a service has of the groups which use it, and how skills should be presented in a job description for a researcher or EOSC service provider.

The report offers an approach to describing these skills in a standard form based on existing methodologies, some of which are common in industry. The Framework offers the basis for a vocabulary of competences and capabilities relevant to EOSC. It presents use cases for the Framework; for service providers to define the skills that users will need, for human resource and training professionals to determine what skills their organisation should improve based on services it will adopt, and for EOSC users to find learning resources that match their individual requirements.

For service-specific skills, the Framework offers a ‘skills user story’ format. Using this, a service provider can define a service capability, and relate it to a competence and competence level from the Framework, according to the professional group the user belongs to. For the more service-independent skills, such as writing a data management plan, the Framework offers competence and capability statements at various levels of expertise, scoped to match individual, research team, or organisation needs. The report illustrates these for selected topics that represent current skills gaps.

Finally, the report draws conclusions about the assumptions made about the sustainability of the Skills Framework. These assumptions will be tested through consultation with EOSCpilot partners and Research Infrastructure training coordinators to inform the final recommendations of WP7.

1. INTRODUCTION

1.1. Overall objectives in the project context

The overall EOSCpilot workplan recognises the importance of an integrated approach to skills for the development of EOSC and therefore has a dedicated work package for skills, WP7. The work in WP7 looks at skills for individuals ('competences', as they are described in this framework) and at the related concept of 'capabilities', which apply to groups and organisations. In the EOSC context individuals and organisations want to be able to find out how to acquire skills, how to demonstrate to others that they have them, and also want to know what mix of skills – both competences and capabilities – will be needed to use a particular set of EOSC services to accomplish a data-related task. EOSCpilot's WP7 addresses each of these issues to some extent. In its workplan it recognises that success for all of them depends on having a common framework or language in which skills can be defined, expressed and compared. The workplan describes incremental development of a competence framework, tested against experience from those parts of EOSCpilot concerned with service development and science demonstrators in particular. The resulting output is the EOSCpilot Skills and Capability Framework (abbreviated to 'the Framework' throughout this document.)

1.2. Purpose of this document

This document provides a version of the Framework. Service providers can use it to describe what skills are needed to use their services effectively. It can be used by training providers to describe what skills can be acquired from a particular course. Individuals can use it to assess their own skills and to search for courses or other mechanisms that will allow them to acquire or enhance particular skills that they need to use certain services. By mapping training provision to the framework, education, training and skills providers can identify gaps in their provision - skills which are needed but which no existing provision helps people to acquire. (National and international bodies such as funders can also make use of the framework in this way, to identify areas where funding for skills acquisition is best targeted.) Organisations can use the skills framework to construct job descriptions and to determine the overall skill sets needed by groups to accomplish particular tasks. The framework builds on work carried out by others (notably the EDISON project), on relevant, recognised service skills frameworks (notably FitSM) and on work carried out earlier by EOSCpilot itself.

1.3. Relationship to other project outputs

The first deliverable of WP7 in EOSCpilot was D7.1 Skills landscape analysis and competence model, which identified and summarised earlier work in this area and presented an initial competence model for individuals which has since been subject to consultation and refinement through presentations, workshops, surveys and dialogue with interested parties. This document (D7.3) updates the competences presented in D7.1 in light of feedback gathered, and maps them to an Open Science Career Assessment Matrix published by the Open Science Policy Platform working group on Rewards.

This document also builds on EOSCpilot's D7.2 *Interim report and catalogue of EOSC skills training and educational materials*. The D7.2 report identifies a range of existing training and learning resources, a set of descriptors for those resources, and gaps in their provision. This successor report D7.3 updates the identified gaps. D7.2 also set out an initial conceptual model for the current framework, extending the notion of competences for individuals to capabilities for groups. D7.3 draws on more recent outputs of the EOSCpilot Science Demonstrators, revises the conceptual model, and clarifies the minimal skill set that the Skills and Capability Framework assumes as a starting point for researchers and others to work within the EOSC.

Later outputs from this work package will draw together the lessons learned from this work and identify priorities for future action by EOSC and its supporting projects.

1.4. Structure of this document

Following the introduction and summary material of section 1, section 2 provides a broad overview of the Framework and the reason for its production. Section 2.1 gives context on the issues that others have identified with skills requirements in relation to EOSC. It identifies calls to action from sources such as High Level Expert Group reports and European Commission statements such as the EOSC Declaration and relates them to the work presented here. Section 2.2 gives a high-level overview of the Framework and section 2.3 describes the model in more detail and the formal terms it uses. It also demonstrates the relationship of this work with the earlier version presented in EOSCpilot D7.2. Section 2.4 looks more closely on how the model presented here relates to the evolving EOSC services landscape and indicates how the model is applicable to EOSC service providers as opposed to EOSC service users. (We recognise that some people and groups will act in both these roles, provider and user, in some cases.)

Section 3 describes the intended users of this Framework and provides use cases to illustrate how the Framework can be employed by them. Three use cases described are:

- Service providers wishing to indicate what skills and capabilities are required of individuals and organisations wishing to use their services
- EOSC users who wish to identify learning resources that can help them acquire necessary skills
- Research leaders and human resource (HR) professionals who can use the Framework to plan skills development, job descriptions and recruitment

Section 4 provides information on the background and development of the model and the sources we drew on, whilst section 5 presents some conclusions and pointers to further work.

The full Framework is presented in table form in Annex A. Annex B shows examples of applying the Framework in a variety of contexts through the means of user stories. Annex C maps the Framework to two of the source models, representing a minimum level of the Framework. These are:

- the Pilot Competence Matrix for Data Management Skills,
- the OS-CAM (the Open Science Career Assessment Matrix) produced by the European Commission's Working Group on Rewards under Open Science in July 2017.

Finally, Annex D contains a glossary of terms used.

2. THE SKILLS FRAMEWORK IN CONTEXT

2.1. The Need for an EOSC Skills and Capability Framework

The need for an EOSC Skills and Capability Framework is clear from a number of European Commission (EC) statements released since the EOSCpilot project started at the beginning of 2017. This section offers a brief review of these statements, which include several dealing with EOSC specifically. These include the first EOSC High Level Expert Group report,[1] EOSC Declaration,[2] and more recently the EOSC Implementation roadmap.[3] Additionally, and with a broader concern for education and training strategy more generally, the EC has set out its vision for a European Education Area, including a Digital Education Action Plan. This plan is intended to "help EU citizens, educational institutions and education systems to make the most of the opportunities opened up by rapid digital change, but also to deal with the challenges associated with this transformation" and to realise three aims it proposes:[4]

- to make better use of digital technology for teaching and learning,
- to develop digital competences and skills needed for living and working in an age of digital transformation, and
- to improve education through better data analysis and foresight.

These aims are consistent with the broader societal impact that the Commission seeks for EOSC. Both it and the Digital Education Action Plan are concerned with development of the appropriate skills to fully engage with open science. This entails expectations of access to and exploitation of digital information. Those expectations, and the ambition of the Digital Single Market, are the backdrop for the Commission's recently revised Recommendation on access to and preservation of scientific information[5]. The recommendations identify a number of key expectations in the area of skills and competences, are shown in Box 1.

The recommendations identify a range of closely related skills areas; "open access, data research management, data stewardship, data preservation, data curation and open science". It is also notable that the recommendations emphasise that education and training should be provided for "all career stages" of science and also extend to industry.

Considering the EOSC more specifically, *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.)."

[1] 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

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

[3] European Commission (2018) *Implementation Roadmap for the European Science Cloud* (Staff Working Document SWD(2018) 83)

[4] European Commission (2018) *Digital Education Action Plan* COM(2018) 22 final. https://ec.europa.eu/education/sites/education/files/digital-education-action-plan.pdf

[5] European Commission (2018) *Recommendation on access to and preservation of scientific information*. C(2018) 2375 final. http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51636

The emphasis here is on a similar range of overlapping skills areas, with slightly more emphasis on data stewardship, and on the development of skills for compliance with the FAIR principles, i.e. that research data should be findable, accessible, interoperable, and reusable.

> *8. Member States should set and implement clear policies (as detailed in national action plans) for the necessary skills and competences of researchers and personnel of academic institutions regarding scientific information. Those policies and action plans should provide for:*
> *- concrete objectives and indicators to measure progress;*
> *- implementation plans, including the allocation of responsibilities;*
> *- associated financial planning.*
>
> *Member States should ensure that, as a result of those policies or action plans:*
>
> *- the necessary training and education are provided about open access, data research management, data stewardship, data preservation, data curation and open science, as part of the higher education and training system, at all career stages, and they reach on-the-job best practice in the industry;*
>
> *- the promotion or implementation, or both, of advanced-degree programmes of new professional profiles in the area of data handling technologies are provided;*
>
> *- the development and training of data-intensive computational science experts are supported, including for data specialists, technicians and data managers.*
>
> Commission Recommendation of 24.4.2018 on access to and preservation of scientific information, p.8

Box 1. Extract on 'Skills and Competences' from C(2018) 2375 final

That emphasis on stewardship is also to be found in the first High Level Expert Group (HLEG) report on EOSC. The HLEG report recommends "a concerted effort to develop core data expertise in Europe" and to "urgently develop adequate data stewardship capacity in European Member States". They refer to "a distinct and largely novel class of research professional... embedded data specialists that are able to support domain specific researcher throughout the entire knowledge discovery cycle." [6]

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 further their research goals most effectively. It is for these reasons, amongst others, that EOSCpilot aimed to deliver a common Framework that could be of use in the widest variety of settings.

There was prior work which we knew we could build on. The Implementation Roadmap for the European Science Cloud [7] identifies a range of prior and current work in the skills area that EOSCpilot WP7 draws on, particularly the EDISON Data Science Framework[8], and the extensive set of open science training materials

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

[7] European Commission (2018) *Implementation Roadmap for the European Science Cloud* (Staff Working Document SWD(2018) 83)

[8] EDISON *Data Science Framework*: http://edison-project.eu/edison/edison-data-science-framework-edsf

collected and produced by FOSTER and FOSTERplus[9]. Our Framework builds on this prior work. It offers 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. Stewardship refers to research output management in an open science and data science context. For the purposes of EOSCpilot we propose a broad definition of stewardship shown in Box 2.

> 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.

Box 2. Definition of stewardship for open data science

This definition reflects the pivotal role of source code, workflows and other outputs in addition to research data and scholarly publications. The term 'data steward' is currently common but the profession is fast evolving and is likely to include responsibilities for outputs other than data.

2.2. Overview of the Framework

2.2.1. Audience and target group

The 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. However, there are at least two 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 developing job descriptions and their assessment for pay and reward and also in recruitment against these job descriptions. 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 small and medium-sized enterprises (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.

The professional groups we consider relevant for this Framework are based on selected competence areas and professional profiles that were defined in the EDISON project. They include data stewards and more established professional groups that support research, particularly data managers, data service engineers, and data scientists/analysts.

The Framework, like that of EDISON and other sources used to define competences, aims to underpin the education of professional data stewards as a distinct role. It should also aim to underpin the development and reward of stewardship skills for current researchers and support staff, who will share responsibility for stewardship. So the Framework treats stewardship as an emergent role whose competences overlap between these established professional support roles, as shown in Figure 1.

[9] FOSTERplus: https://www.fosteropenscience.eu/

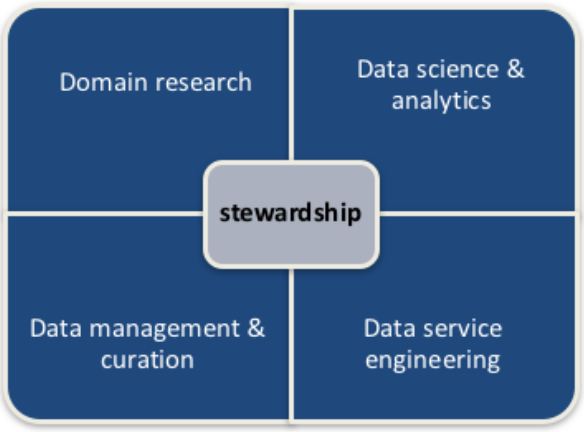

Figure 1 - Professional Groups

2.2.2. Structure

The Framework utilises a set of skills groups which are presented in tabular form in Table 1 later in this section. The groups (Plan and Design; Capture and Process; Integrate and Analyze; Appraise and Preserve; Publish and Release; Expose and Discover; Govern and Assess; Scope and Resource; Advise and Enable) include generic and project-specific elements.

This Framework draws on pre-existing competence frameworks developed for data management, open science, and data science. The starting point – what the Framework assumes as the basic level knowledge needed to work in the EOSC environment – is a core group of eleven data handling competences identified for undergraduate-level information professionals, and collectively termed 'digital information literacy' skills (Sapp Nelson, 2017[10]). These core competences have been mapped in Annex C to the six 'skills groups' shown in Figure 2.

The Framework envisages a stepwise enhancement of skills across three dimensions shown in Figure 2: competence, responsibility, and organisation. Competence levels range from basic comprehension, through to the application of the competence, and the acquisition of expertise in that competence (the ability to evaluate the application of the competence, and synthesise or create something new in that area of competence). These competence levels are derived from Bloom's taxonomy[11].

The other dimensions of the Framework are increasing levels of responsibility, and the organizational level that competences are applied to. The responsibility levels are informed by the SFIA model (Skills Framework for the Information Age[12]).

[10] Sapp Nelson, Megan R.. 2017. "A Pilot Competency Matrix for Data Management Skills: A Step toward the Development of Systematic Data Information Literacy Programs." *Journal of eScience Librarianship* 6(1): e1096. https://doi.org/10.7191/jeslib.2017.1096

[11] 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

[12] https://www.sfia-online.org/

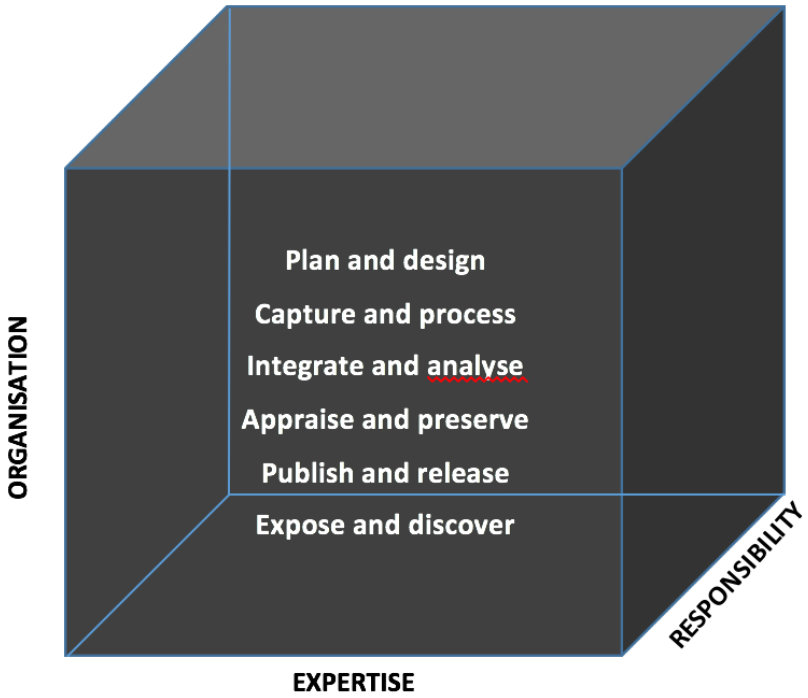

Figure 2 - Core Skills Groups & Dimensions

The organisational dimension is borrowed from the digital information literacy model (Sapp Nelson, 2017). Figure 3 illustrates these.

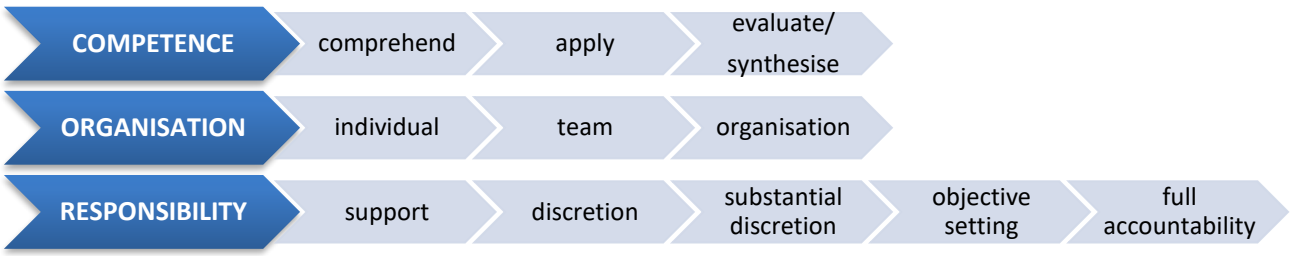

Figure 3 - Skills Framework Dimensions

For individuals new to data stewardship, and for educators or trainers planning a curriculum for data stewardship, the minimum level is the lower level of each of the dimensions shown in Figure 2, i.e. to be able to comprehend the topics identified in Annex A as the recommended minimum, and be able to support others in applying their competences in these areas, for example by using services or supporting others to do so.

The Framework uses the term *capability* to refer to the collective competence level and responsibility level needed by a group to perform a service user or provider role. 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 – the skills groups in Figure 4, which was initially presented in

deliverable D7.2 of this project. The groups themselves are presented in more detail in Table 1.

Figure 4 - Skills Groups

As Figure 4 shows, certain skills groups 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. So the Framework 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 1 describes the scope of these and the core skills groups.

The Framework allows for flexibility in how responsibilities for the skills and capabilities will be distributed between organizational levels. Managers in organisations and research teams will need to take into account their disciplinary and organizational context to match competences with the appropriate level of responsibility.

Table 1 – Scope Of the Skills Groups

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 the requirements of research output users, the organisation and research funders, then establishing effective approaches to meet their requirements, and then testing and validating the initial outputs.
Capture and process	Capturing and processing of data or related materials to enable research evidence to be prepared for analysis; provisioning of secure managed access to networked storage, scalable to meet demands, plus resources, tools, standards and workflows for collaboration between research team members, and relevant third parties
Integrate and analyse	Developing and applying appropriate methods to enable lines of enquiry to be formulated and pursued towards the research objectives, by assembling and integrating selected data, software, systems, or other resources, and enabling relevant knowledge and techniques to be applied in their analysis and transformation into research outputs.
Appraise and preserve	Developing and applying appropriate methods to appraise research outputs for their compliance with ethical, FAIR and research integrity principles, their value to the organisation and to research reproducibility, and their potential to serve new purposes or communities; planning and taking action to mitigate risks to long-term access for further appraisal.
Publish and release	Describing research products and their inter-relationships, providing access to meet the needs of their providers, users, and other stakeholders, in order to maintain or enhance their value and comply with ethical, FAIR and research integrity principles and policies.
Expose and discover	Ensure that processes and mechanisms for providing access to research products and their inter-relationships follow technology developments, community standards, and good practices for compliance with ethical, FAIR and research integrity principles.
Govern and assess	Developing and maintaining the organisation's strategies, policies, and processes on FAIR/open research outputs, and associated documents and processes that enable these to be implemented, and relevant laws or regulations to be complied with. Continually reviewing these strategies, policies and processes through stakeholder consultation, communication, and impact monitoring.
Scope and resource	Identifying the scope of research data services and stewardship activities and securing the resources to sustain these. Continually reviewing the business case considering the service value propositions, processes, and relevant costs and benefits, taking into account governance processes and timelines, and the need for cost recovery mechanisms to comply with funder requirements.
Advise and enable	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.

2.3. Linking EOSC Capabilities to Skills and Competences

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.

An initial service architecture was defined for EOSC by EOSCpilot WP5 in December 2017 (D5.1). Certain aspects have been updated and clarified in D5.2 during spring 2018. The architecture is currently under further-development and will be documented in one of the final project deliverables (D5.4 due at December 2018). Meanwhile, for the purpose of this report, we use the broad service classes identified in the EOSC implementation roadmap[13] to illustrate how to link competences to the capabilities associated with these service classes, for skills development purposes.

Within the 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 soon-to-start H2020 projects. According to this model EOSC "should be a federation of existing and planned research data infrastructures, adding a soft overlay to connect them and making them operate as one seamless European research data infrastructure." In terms of architecture, the EOSC would essentially comprise a federating core and a variety of federated research infrastructures committed to providing services as part of the EOSC. Figure 6 below is our attempt to present the main service classes, their connections and the user access channels:

- 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/are required to use to become part of the EOSC federation. (Primarily a Catalogue/Marketplace; Security and identity management; Persistent identifier service). 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*[14] *and thematic*.
    - 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. Typical generic services are, for example Infrastructure as a Service clouds; Service Monitoring; Resource Usage Accounting service; Helpdesk; Scalable storage.

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

[14] 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.

- 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 start in 2019 and are expected to federate further services and data from disciplinary areas covered by the ESFRI roadmap.
-

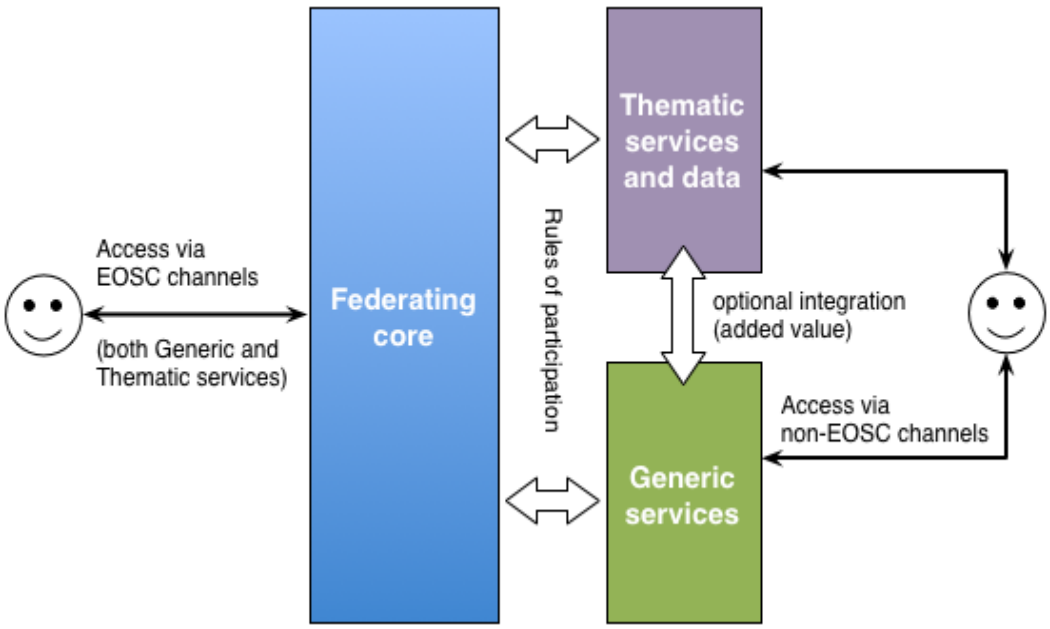

Figure 5 – Service Classes; Connections; Access Channels

The EOSC implementation roadmap also includes a classification scheme for the services envisaged in EOSC. Based on the classification we can define initial service functions that EOSC will expose towards its future users. All of this is still subject to change, but for the purpose of demonstrating the applicability of the Framework this is not a concern.

Table 2 - Service Functions

Service class [15]	Suggested service function(s)	Target user(s)
A unique identification and authentication service and an access point and routing system towards the resources of the EOSC.	● Generate PIDs for federated services and data ● Authenticate users ● Make federated items searchable, locatable, accessible	● EOSC service users ● Those who want to federate new services into EOSC
A protected and personalised work environment/space (e.g. logbook, settings, compliance record and pending issues).	Persistent space to transfer and integrate data or other objects across multiple services	● EOSC service users ● Service operators (both core, generic and thematic)
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).	● Guides and good practice documentations ● Service status updates (e.g. real-time feeds)	● EOSC service users ● Service operators (both core, generic and thematic)
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).	The biggest part of EOSC. The 'real content'. The diverse types of datasets[16] and analysis tools to process, store, share.	● EOSC end users (scientific users) ● Providers of thematic services (they may need Generic services to implement value-added features)
Services to make their own data FAIR, to store them and ensure long-term preservation.	Repository and PID and preservation services	● Providers of thematic services (they need these services to implement preservation/FAIR-ification services for domain users) ● EOSC end users (scientific users not having Thematic services for preservation, FAIR-ification, etc.)

Based on the architecture and service classification in Table 2 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:

[15] The service classes in Table 2 are as described in the EC Staff Working Document *Implementation Roadmap for the EOSC* (http://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf). The other columns are the authors' own work based on current thinking in EOSC-Hub and related projects.

[16] The EC document does not consider data to be part of 'services'. The authors of this document look at data as services that are provisioned for access/download/transfer under certain conditions and SLAs.

Table 3 – Service roles, activities and knowledge

Service role	Main activities performed	Required knowledge
Owners and operators of the federating core	• Fit service into the core • Monitor proper operation of the core as well as the federated services • Respond to errors/malfunctioning directly or via peers • Support Generic and Thematic providers federate into EOSC	• Architectural understanding of the EOSC core • Expert-level knowledge of EOSC core services • Deep technical skills in the operation of core services in scope for the operator • User/customer support skills
Owners and operators of Generic services	• Federate the service into EOSC via the federating core • Maintain operational continuity of the service • Upgrade the service as required (based on new releases from developer team) • Recognise and respond to service errors and malfunctioning • Support and train users of the service	• Deep technical skills in the deployment, operation and monitoring the Generic service in scope • Application of the EOSC 'Rules of Participation' to the Generic service • User-level knowledge of EOSC operational support tools • User/customer support skills
Owners and operators of Thematic services (incl. Data as a service)	• Federate the service into EOSC via the federating core • Maintain operational continuity of the service • Upgrade the service as required (based on new releases from developer team) • Recognise and respond to service errors and malfunctioning • Support and train users of the service	• Deep technical skills in the deployment, operation and monitoring the Thematic service in scope • Application of the EOSC 'Rules of Participation' to the Thematic service • User-level knowledge of EOSC operational support tools • User/customer support skills
Users of Generic and/or Thematic services and data (incl. scientific 'end users' as well as local institutional supporters for scientists)	• Develop user-level knowledge of the relevant services of EOSC (via documentation, training, consultancy) • Apply the services and data to generate new scientific findings • Share the tangible outputs in a way that makes them visible in EOSC (data, workflows, new algorithms, papers)	• Navigating the EOSC catalogue and identification of relevant services • User-level knowledge of relevant services of EOSC • Generic data science skills (DMP, stewardship, statistics, etc.) • Data science skills depending on the discipline (data formats, validated algorithms, ethics, etc.)

The 'required knowledge' in Table 3 above is a high level description of the competences and capabilities described in more detail in the Skills Framework in Annex A.

2.4. Conceptual model

The EOSCpilot Skills and Capability Framework defines terms such as 'competence', 'skill', and 'capability', as detailed in the Glossary at the end of this report. The relationships between the main concepts are shown in Figure 5. This is an update of work initially presented in EOSCpilot D7.2, which was missing a direct link from capabilities to skills. The relationships are now also labelled for clarity.

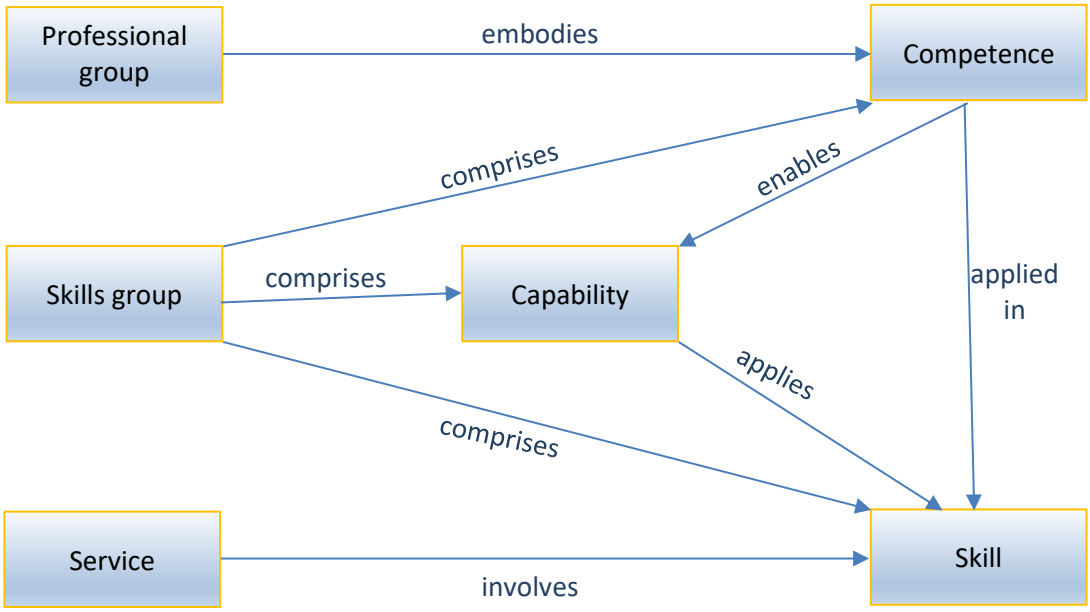

Figure 6 - Conceptual model for EOSCpilot skills framework

The Framework proposes a user story format for service operators to specify skills. The format combines some of the terms shown in Figure 5 into a single sentence, as follows:

> As a [*professional group*] needing to [*deliver a capability*] (and as a [*service - role*]) I should have [*expertise level*] in [*competence*]"

The [service - role] element is optional, as not all skills relate to specific services (e.g. data management planning is a competence names the service and identifies whether the story refers to a user of the service or an operator of the service. For example:

> "As a [data manager] needing to [support colleagues in selecting repository services] and as a [repository operator] I should [be able to comprehend] [repository/database evaluation and options appraisal]."

The next section expands on this idea and Annex B gives further example of skills user stories.

3. USE CASES AND APPLICATION OF THE FRAMEWORK

The Framework aims to help plan skills development at both an individual and organisational level. It offers service operators a simple method to specify the skills involved in using their service, to identify these in a Service Catalogue. For individual researchers, and research or training managers, it offers a tool for identifying the capabilities they need to deliver for their team/organisation, based on EOSC Service Catalogue descriptions, and also find the relevant competences and levels of expertise needed for the researcher and professional support groups that are target users of the service.

The worked examples in this section describe three applications of the Framework: firstly for service operators to specify the skills and, secondly, for service users to find relevant materials. The third application is for the broader context – the research manager or human resources professional who needs to define the broad set of stewardship skills needed to deliver new capabilities for their team or organisation to operate the EOSC context.

3.1. A framework for service operators to specify skills requirements

EOSC service operators provide entries to the EOSC Service Portfolio, a subset of which will be available to service users in the EOSC Service Catalogue. The development of these is informed by the eInfra Central project, which is aiming to create a harmonised service catalogue view across a number of European eInfrastructures. The service description template offers an example of the attributes to be described, which include the option to specify training information[17].

The Skills Framework can help service operators to specify skills requirements to include in a service description, as it provides a 'user story' format, to help articulate the dimensions of the Framework that specify those skills. Using this format, Operators can specify the competences their service will require and/or improve, and the appropriate expertise level for the service target users (researchers or professional groups).

To offer an example, one of the classes of service planned for the EOSC is described in the Implementation Roadmap (see Table 2 in the previous section) as "*services to make data FAIR, to store them and ensure long-term preservation"*. This description is close to a 'trustworthy digital repository' as defined by international standards such as ISO16383 and, at a basic level, by the CoreTrustSeal (see section 4.3) standard.

As the Core-Trust-Seal defines common requirements for a repository service it has been used as a source of some of the capabilities described in the Skills Framework. Service operators can translate these high-level capabilities into more specific functions identified with their users, and also to specify the skills needed to use their repository service effectively. Training coordinators can also use the high-level capabilities to specify generic skills to use any repository service effectively.

3.1.1. Finding relevant capabilities

The starting point is the top level of the Skills Framework. This has several 'skills groups' that can readily be mapped to the Core-Trust-Seal requirements, with more specific competence and capability topics at the next level. For example:

Govern and assess
- Information security and risk management
- Storage security management

Appraise and preserve
- Data preparation and documentation for reproducibility
- Data quality assurance using open standards

[17] The current eInfra Central service description does not define how training information will be specified but offers the potential for this to be an actionable link to a training catalogue, see: http://einfracentral.eu/basic-page/common-service-catalogue

- Data transfer and long-term storage
- Format and media migration

Publish and release

- Access control and management
- Sharing via open repositories/ platforms

Expose and discover

- Metadata and persistent id. exposure

3.1.2. Expanding the capabilities

The next step is to describe how each capability is delivered by an organisation and if appropriate, at the research team level. Take 'Sharing via open repositories/ platform', for example, a relevant organisation level capability would be:

> "Provides repository functions where necessary to ensure preserved research objects remain accessible for as long as required, enabling these functions to interoperate with other systems or processes that provide or extract information to enhance the research objects' compliance with FAIR principles.".

Annex A gives further examples relevant to this and other service-related and generic skills areas.

3.1.3. Identifying the competence dimensions

Competences define a level of expertise needed for a service role (user or operator) in a professional group. They should be described in the form of behaviours that can (in principle) be assessed by others.

The dimensions that need specified in order to describe how the competences are applied are:

- service role: user or operator
- professional group: data managers, domain researchers, data service engineers, or data scientist/analyst
- expertise level - comprehend, apply, or synthesise and evaluate (or 'be expert in')

The Framework identifies examples of these relevant here, e.g.

- service role: operator
- professional group: data manager
- expertise level - synthesise and evaluate (or 'be expert in')
- competence: "develop, maintain and review repository processes for ingesting research objects, handling the associated bitstreams, descriptive information and metadata for preservation, and ensuring the appropriate level of access for the objects' designated communities"

3.1.4. Specifying the skills

A user story format may help to write text that combines all the dimensions in a single statement, e.g. to show to service users as an expression of what they need to be able to know about or do, and revise based on their feedback. For example, using the format already described:

> "As a data manager needing to provide an appropriate level of curation - and as an operator of a repository service - I need to be able to develop, maintain and review repository processes for ingesting research objects, handling the associated bitstreams, descriptive information and metadata for preservation, and ensuring the appropriate level of access for the objects' designated communities."

The format does not guarantee to produce sentences that are grammatical or easy to read, but making them so, and getting feedback on the results, should help the service operator to define skills in a form that their target groups can relate to their learning needs.

3.2. A framework for EOSC users to identify learning resources

This process is currently a manual one. To make a more automated process sustainable, we foresee the EOSC service developers and operators volunteering (or being required) to provide specifications of the skills involved in operating and using their service. This could involve describing relevant user stories in the EOSC Service Portfolio, and using these stories to specify the skills requirements in the Service Catalogue.

On that assumption, the Framework helps Identify learning resources that are relevant to using or operating an EOSC service in three steps:

1. A service operator specifies in the EOSC Service Catalogue the skills involved in applying the service's main functions. The specification gives the targeted professional groups, the competences, and the expertise levels. Optionally, links to training materials or to catalogues of these could be provided.
2. Reading the specification, the EOSC service user works out whether they need training, by comparing the specified skills (competences and expertise) against those they already have. This might draw on a skills development plan, identifying the individual's current learning needs, based on their role and the EOSC services that they or their organisation plans to make use of.
3. The user refers to an EOSC training catalogue to identify materials about the competences they need, targeted to their professional group, and at a level of expertise suitable for them and the service they plan to use.

As the terms used in the skills specification (target group, competence, expertise level) are very generic it should be possible for service users to use any catalogue or portal offering training and learning resources for their domain, and to search on these for learning resources relevant to delivering the services capabilities.

It should be possible, however, for the EOSC to offer a greater level of automation and precision to the researcher looking for relevant learning materials. The Training-as-a-Service concept, proposed in the 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 training e-Support Service (TeSS) system.

As discussed in D7.2, an EOSC Training-as-a-Service would include a cross-domain training registry, offering a minimal set of discriminators to describe learning materials. This D7.3 report builds on that idea, with the proposition that services link to relevant learning materials via a skills specification in the Service Catalogue.

D7.2 also foresees the automatic derivation of EOSCpilot training metadata, 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.

Figure 7 shows how EOSC service users would find relevant learning resources in that long-term scenario.

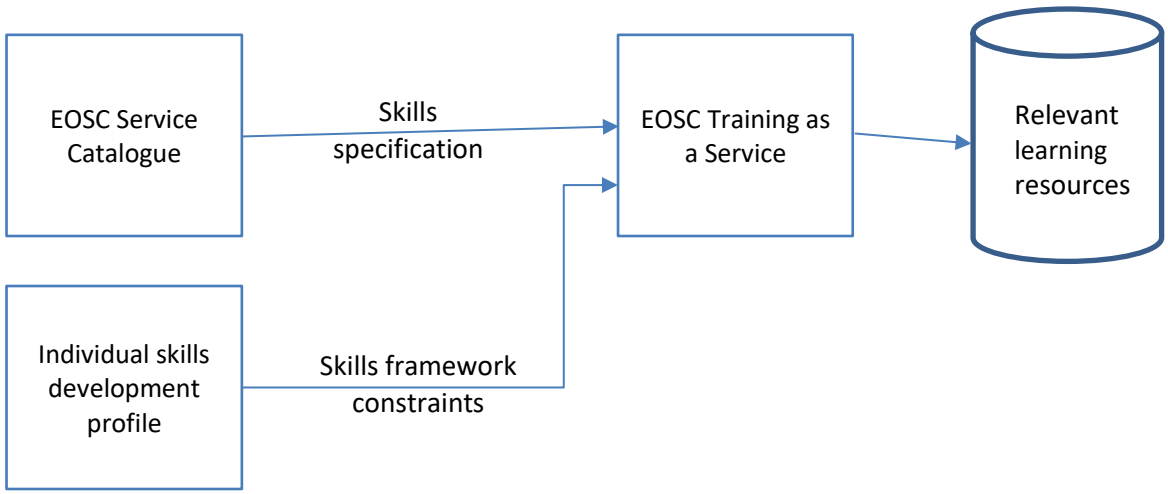

Figure 7 - Overview of three step procedure to identify relevant learning resources

As an example, consider a data manager in a faculty of life science. The data manager has been asked to consider whether the EOSC repository service would be an effective solution for researchers in the faculty to store any research data that does not fit better to an existing life science domain repository. Consulting the EOSC Service Catalogue entry for the repository service class they find a table in the 'training' field. This table specifies the relevant skills for users of a repository service, including '*repository/ database options appraisal and selection*', which is listed alongside the relevant target groups and levels of expertise as shown in Table 4.

Table 4 – Skills specification entry for a repository service in an EOSC service catalogue

Competences	Professional groups and expertise level
Repository/ database options appraisal and selection	• Data manager: expert • Data service engineer: comprehension • Data scientist/analyst: application • Domain researcher: comprehension
Sharing via open repositories/ platforms	• Data manager: expert • Data service engineer: application • Data scientist/analyst: application • Domain researcher: comprehension

The Service Catalogue training entry may also identify sources of training materials that the EOSC repository service operator has agreed with interested Research Infrastructures. These could include, for example, the ELIXIR TeSS. Based on the information in the table, the data manager may then follow a link to TeSS and searches for relevant material.

The EOSCpilot Skills Framework training metadata discriminators specifically support the finding of materials in terms of the *competence*, the *target professional group*, and the *expertise level*. The Repository Service operator can therefore more easily support service users by specifying these, using the user stories to identify terms from the Skills Framework he underlying mapping of user stories to the discriminators of training materials is shown in Figure 8.

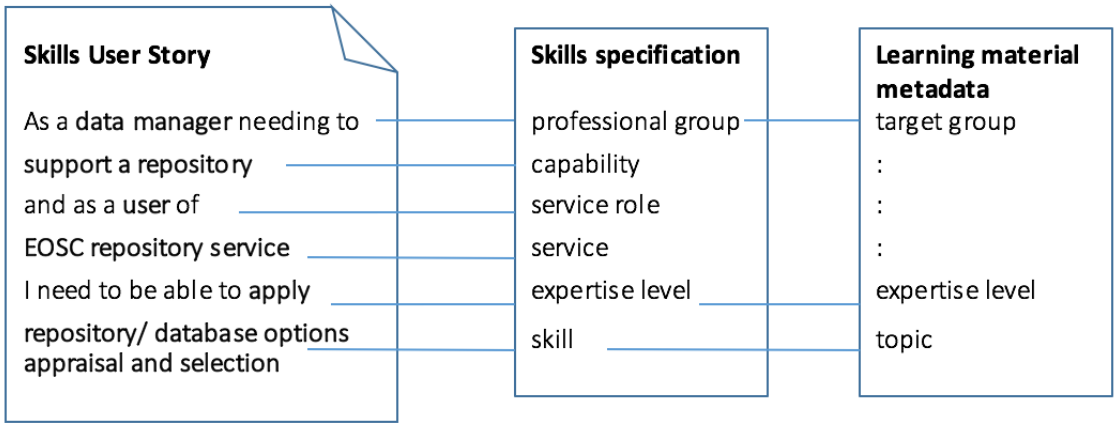

Figure 8 - Mapping vocabulary from user stories to EOSCpilot Framework to training metadata

At the current state of development, the described three step procedure is a conceptual model that needs evaluation by EOSC project partners and RI training coordinators.

3.3. A framework for research leads and human resource professionals to plan skills development

The third use case is the research manager, principal investigator or human resource professional who is interested in EOSC services, and wants to understand the skills implications of these services and, if necessary, define new roles in their group or organisation.

3.3.1. Describing service-specific roles

The Framework can support managers to establish the capabilities they need from a role, by first narrowing down the class of services they need, and then following links to descriptions of capabilities for that class of service, and/or from profiles of the professional groups that most closely match that service class. EDISON profiles of data science roles have informed the Framework, which links the professional groups to the relevant EDISON professional profiles. As it also draws on other sources (outlined earlier in Section 2) the competences used are not a precise match to those used in the EDISON professional profiles, however they can be used with the Framework as follows.

1) Find relevant skills group

From the top level of the model scan the scope descriptions for each skills group, listed in Table 1, identify those likely to be relevant, considering whether the cross-project or project-specific skills groups role are likely to be more relevant. Factors to consider here are whether the project is mainly managerial, and whether it is broad-based to serve multiple research teams. In either case the cross-project groups will be more relevant. Project-specific skills groups will be relevant to hands-on roles.

2) Find relevant professional groups

Scan the Framework listed in Annex A, and identify the professional groups that are the closest match, considering the service classes likely to be relevant, and whether the role will mainly involve operating services or using them.

3) Adapt capability statements and professional profiles

Identify the professional groups that are of interest, follow links to the corresponding EDISON profiles, and adapt wording from these and the capability statements for the service classes, considering the level of responsibility identified in the framework.
Annex B includes examples of capability statements to choose from, based on the services and skills needed, with matching competence statements.

3.3.2. Using job descriptions

It can be helpful to work in a more bottom-up direction from existing job descriptions. In the example below we look at a job description from a research institution, TU Delft in the Netherlands, which identified a need for Research Data Stewards, and analyse the wording to match it to competence statements in the Framework.

Job description (extract)
Data Stewards:

- Have experience with research methodologies of the faculties.
- Have (or are studying for) a PhD in a relevant subject area, or have significant experience with how research data is managed.
- Knowledge of how research software is managed would be a plus.
- Have a broad understanding of how research operates and how data and software underpins high-quality research.
- Are excellent communicators, able to speak not only with researchers but also with other support staff,
- Have understanding of the faculty specific needs
- Are sensitive to organisation-specific culture and practices.

The Data Steward will take the **lead** in engaging researchers from both faculties in better data management practices. Your main task is to **advise** researchers on data management throughout the research lifecycle. You will:

- **Act** as spokesperson for your faculty, **create** awareness and **explain** to researchers the added value of good data management.
- **Lead** the development and implementation of the faculty's data management policy, with understanding of faculty-specific needs.
- **Assist** in planning the collection, management, and publication of data in research projects.
- **Explore** and **analyse** trends in research data management specific to your faculty.
- **Advise** which short- and long-term actions to take to advance research data management across the university.
- Regularly **liaise** with Faculty Secretaries and the Data Stewardship Coordinator.
- **Develop** and **run** training events tailored to researchers' needs and **inspire** researchers to participate.
- **Assess** and report on the progress of the project.

Figure 9 - Job description

The background to the appointment was a one-year project to appoint discipline-specific Data Stewards to each of the eight TU Delft faculties. The long-term objective of the project is "to ensure that researchers across all the disciplines supported at TU Delft adhere to good research data management practice in their day-to-day work."[18]

[18] T.U. Delft (2017) Data Stewardship – addressing disciplinary data management needs Open Working (Aug 29, 2017) available at: https://openworking.wordpress.com/2017/08/29/data-stewardship-addressing-disciplinary-data-management-needs

The role is broad advising and enabling, aiming to attract young researchers to bring their domain expertise to the role and develop their data management skills. This implies data management competences may need to be enhanced more than research competences, with a focus on those under the 'advise and enabling' heading, and on the planning and publication ends of the project lifecycle.
The extract of the job advertisement for the Data Steward roles shown in Figure 9 highlights the verbs used to describe the job function. These are useful to identify the personal attributes and objectives of the job, and from there the relevant competences and levels of expertise and responsibility.

As an example we will analyse the first sentence of the job description:

"The data steward will take the lead in engaging researchers from both faculties in better data management practice"

This sentence can be divided into the job attribute, the objective of the task, and the verb used to conduct the specific task:

Attribute: *"Have experience with research methodologies of the two faculties"*
Objective: *"Engaging researchers from both faculties in better data management practices"*
Verb used: *"lead"*

The verb "lead" indicates a high responsibility level, although as this is not a senior management role the indication is that "objective setting" rather than 'full accountability" would be the appropriate level on the five-point scale used in the Framework, shown in Figure 10.

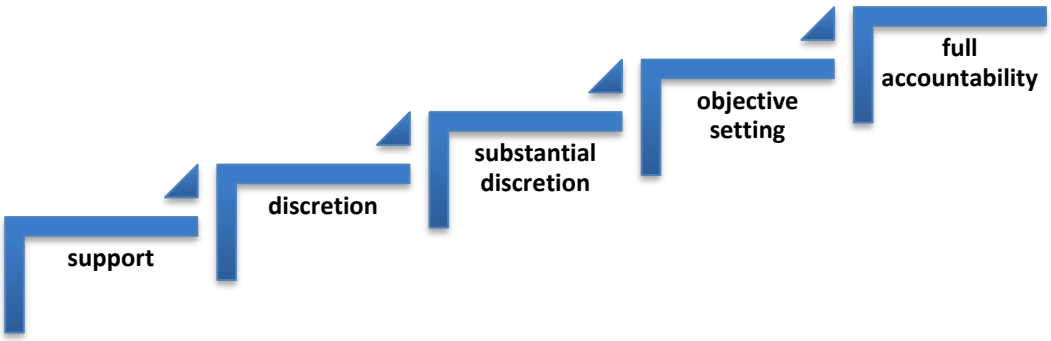

Figure 10 - Responsibility levels in the EOSCpilot Skills Framework

Following this example we can analyse the complete job description. The results are listed in Table 5.

Table 5 - Job description analysis

Job profile description			Skills Framework model		
Function	**Verb**	**Objective**	**Responsibility**	**Expertise**	**Competence**
Have experience with research methodologies of the faculties.	lead	in engaging researchers from faculties in better data management practices.	substantial discretion	evaluate and synthesize	Planning data management and sharing (DMP)
Have (or are studying for) a PhD in a relevant subject area, or have significant experience with how research data is managed.	advise	researchers on data management throughout the research lifecycle	substantial discretion	evaluate and synthesize	Planning data management and sharing (DMP)
Knowledge of how research software is managed would be a plus.	act	as spokesperson for your faculty	support	apply	Open source software / service development
Have a broad understanding of how research operates and how data and software underpins high-quality research.	create	awareness to researchers the added value of good data management.	support	evaluate and synthesize	Advocacy of FAIR principles and Open Access policy
Are excellent communicators, able to speak not only with researchers but also with other support staff,	explain	to researchers the added value of good data management.	discretion	apply	Supervision and mentoring
Have understanding of the faculty specific needs	lead	the development and implementation of the faculty's data management policy	substantial discretion	apply	Engaging with research users and stakeholders
Are sensitive to organisation-specific culture and practices.	assist	in planning the collection, management, and publication of data in research projects.	substantial discretion	comprehend	Research design
	advise	which short- and long-term actions to take to advance research data management across the university.	discretion	evaluate and synthesize	Costing of data management and preservation
	explore	trends in research data management specific to your faculty.	discretion	evaluate and synthesize	Research strategy/ open research potential
	liaise	with Faculty Secretaries and the Data Stewardship Coordinator.	support	apply	Building collaboration (cross-sector)
	develop	training events tailored to researchers' needs	substantial discretion	evaluate and synthesize	User support and training
	run	training events tailored to researchers' needs	objective setting	evaluate and synthesize	User support and training
	inspire	researchers to participate.	support	apply	Personnel development
	assess	the progress of the project.	full accountability	evaluate and synthesize	Project management

The responsibility levels could be further used to match the job description to a career grading scheme, and the matched competences can be used to find training resources to meet any individual needs appropriate. For writing future job profiles, the competences can also be used to identify the matching capability statements, as a source of text on relevant job functions and objectives.

4. DEVELOPMENT OF THE MODEL

4.1. Principles used in drafting the framework

The following principles were set out in D7.1 as a basis for designing a framework flexible enough to use in varied professional development contexts, and alongside existing frameworks.

4.1.1. Recognise FAIRness spans multiple data lifecycles

As the FAIR principles articulate, stewardship implies long-term and sustainable care across multiple lifecycles. This makes stewardship a collective endeavour, involving at least the individual researcher, colleagues in the study ('the team'), their host organisation (or and others providing services) and the research domain(s) or communities that care about the data.

4.1.2. Recognise the policy and organizational context

The framework should enable users to construct role profiles, e.g. to define professional development goals, taking into account their organisation's professional development and data policy context.

The framework should recognize that organisation and research cultures will influence the levels of responsibility that different roles have across the data lifecycle. There will also be variation in how funder and institutional policies (or legislation) define roles and responsibilities, e.g. by placing differing expectations on the research team, host institution and third-party organisation (such as data centres).

The framework should be flexible enough to allow competences to be identified with different levels of the organisation, reflecting the relationships in place between individual researchers and staff with stewardship responsibilities in their team, centralised research data management services, and other providers in their research community, including EOSC service providers. The extent of outsourcing or shared responsibility for stewardship will also depend on organizational characteristics e.g. research strategy and capability.

4.1.3. Recognise disciplinary norms in stewardship responsibility

The framework should help users deal with disciplinary differences across the data lifecycle in the level of responsibility for data stewardship that they can realistically assign to 'embedded' individuals – referring to those with a mandate to manage data at research team level (e.g. for one or more Principal Investigators).

In some domains it may be preferable to expect individual researchers to be primarily responsible for stewardship, supported by data librarians operating across higher level organizational groups (departments etc.). This variation is likely because research domains vary in their propensity to work in teams, availability of domain-level standards e.g. for formats and metadata, well-defined reuse cases, and availability of third-party repositories.

4.2. Competences used in the draft framework

The first draft of the Skills Framework focused on competences and was reported in D7.1. That report described sources used for the competences as follows:

- Demchenko, Y. Belloum,A. and Witkorski, T, (2016.) EDISON Data Science Competence Framework v.0.7
- Sapp Nelson, M. (2017). A Pilot Competency Matrix for Data Management Skills: A Step toward the Development of Systematic Data Information Literacy Programs.
- Molloy, L. Demchenko, Y, Jung, C. et al. (2016) 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
- Hodson, S. et al (2017) Objectives, Scope and Activities of a Possible GO-TRAIN Implementation Network, Unpublished discussion document
- Lyon, L. & Brenner, A. (2015): Bridging the Data Talent Gap: Positioning the iSchool as an Agent For

Change ; International Journal of Digital Curation 10(1) http://dx.doi.org/10.2218/ijdc.v10i1.349

The D7.1 report describes in more detail how these were synthesized, and we provide a brief summary here focusing on the main sources. These were the first and last in the above list: the EDISON framework, and Sapp Nelson's Competency Matrix for Data Management Skills.

4.2.1. EDISON

The EU-funded project EDISON established mechanisms intending to increase the number of competent and qualified Data Science professionals across Europe and beyond. The core output of the project is the EDISON Data Science Framework, a collection of documents comprising a Data Science Framework, Body of Knowledge, Model Curriculum and Professional Profiles. The EDISON Data Science Competence Framework [19] is the main source for D7.1.

EDISON defines four role profiles associated with "Data Management": Data Steward, Digital Curator, Digital Librarian and Data Archivist. Data management itself is one of the six "skills areas" along with Data Analytics, Data Science Engineering, Business Process Management, Scientific Methods and Domain Knowledge. We have drawn the Skills Framework competences from these, and rephrased the skills areas they relate to as the 'professional groups' shown on Figure 1, as follows:

- Data Service Engineering - broadening the scope of 'data science engineering' to include the competences to develop, operate and manage research software and data services, beyond those specific to data science.
- Data Science & Analytics - combining data analytics competences with those for data science in the academic research context, e.g. set up and management of scientific workflows for research projects.
- Data Management & Curation - highlighting the curation and stewardship competences to enable reuse of research data, as well as those to make the data actionable in the research project it is first used in.
- Domain Research - combining domain knowledge with the competences to apply scientific methods.

EDISON provided twenty-four data science oriented competences, drawing on six described for each of the four groups above.

4.2.2. Competency Matrix for Data Management Skills

The matrix described in Sapp Nelson (2017), and in more detail in related spreadsheets, is aimed at supporting the design of undergraduate curricula for information professionals. It distinguishes between three organisational levels of competence; personal, team and research enterprise. The personal level is associated with basic "data information literacy', while data stewardship is associated with a mastery level, a set of skills that leaners encounter "… as they try to organize the data management of many individuals working on a common research endeavour (p. 3).

The competences are contextualized to each of the three organizational levels (personal, team, and research enterprise), and in each case divided into thirty-six statements, each with three competence statements which are then broken down into knowledge, skills and attitudes. The statements are in twelve groups described in Annex C.

4.3. Service-related capability models

In D7.1, the first draft of the Skills Framework set out first steps towards the aim expressed in the EOSCpilot proposal (p.11), that the WP would relate the skills demanded of researchers and support professionals to "the capabilities needed to provide EOSC services and ensure they are effective tools in

[19] Demchenko, Y. Belloum,A. and Witkorski, T, (2016.) EDISON Data Science Competence Framework v.0.7 EDISON Project, http://edison-project.eu/data-science-competence-framework-cf-ds

skilled hands". The first steps were to identify standards that could act as a proxy for the EOSCpilot Service Architecture (before its release) by describing capabilities needed to provide EOSC services, and to use them.

Two standards were identified to support this effort, both on the grounds that WP7 partners were able to contribute expertise on them, having been directly involved in their development. These were the CoreTrustSeal standard for data repositories[20], and the FitSM standard for service management.[21] These were also of interest as the standards are backed by certification processes.

The input to D7.1 was limited to ensuring that major topics the standards address have high-level representation in the competence framework (e.g. 'change management' in the case of FitSM, and 'metadata and PID exposure' in the case of CoreTrustSeal).

D7.2 expanded on the conceptual link between competences and service capabilities, by identifying a user story format as a basis for articulating the skills needed to deliver on the high-level requirements that services fulfil.

4.4. Synthesis and revisions to the Framework

The competences synthesized in the D7.1 report are intended to represent 'open data science stewardship'. They comprised 82 topics drawn from EDISON, the Sapp Nelson matrix, and the Education & Training IG wiki. This synthesis represented two overlapping sets of competences; those to apply data science methods in research, and those to ensure that the resulting knowledge and evidence are FAIR (findable, accessible, interoperable and reusable).

This synthesis was also indirectly influenced by the GO-TRAIN discussion paper and the thesaurus of open science topics produced in the FOSTER and FOSTERplus projects. These projects have collated and produced training materials on open science skills, which were included in the landscape analysis in D7.1 and training materials catalogued in D7.2. The other influence was the SFIA (Skills Framework for the Information Age), a broader competence model for IT skills, to define levels of responsibility for exercising skills in the Framework.

D7.1 adopted the three organizational levels from the Sapp Nelson competence matrix, and proposed that capabilities should be treated as competences whose delivery is counted at levels beyond the individual, i.e. at the levels of the research team or organisation (and their respective services).

The Framework differs in format from that matrix, which articulates 'knowledge, skill and attitude' statements at all three organisational levels. Instead we use a similar typology at the individual level, to identify the expertise level for individuals, but at the research team and organisation level we identify capabilities by single statements that express a high-level requirement (whether potentially delivered by a single service or otherwise).

The aim was to derive capability statements from the EOSCpilot service architecture. This proved unrealistic as the WP5 work towards that has dealt with more abstract definitions of the architecture components and activities. The service classes provided by the Implementation Roadmap and the EOSC-hub have more directly informed sections 2 and 3 of this report, and the approach to specifying capabilities. The high-level scope of the Framework also draws on the capability model RISE (Research Infrastructure Self Evaluation)[22]. RISE describes research data management service capabilities from the perspective of institutions that would use the services EOSC offers.

[20] Core Trust Seal https://www.coretrustseal.org/about/
[21] FitSM http://fitsm.itemo.org/
[22] RISE: http://www.dcc.ac.uk/resources/how-guides/RISE

4.5. Skills gaps from science demonstrator outcomes

In D7.2 we reported further work to distil the first draft and identify priorities. This was based on feedback gained from a breakout session at the at the EOSCpilot-OpenAIRE joint workshop in Berlin (24 October 2017) attended by representatives of academic libraries, institutions, and e-infrastructures. This resulted in sixteen changes, four additions and an indication from the participants that they saw the main competence gaps as those representing 'soft' skills and policy-oriented areas of knowledge. The main priority 'gaps to fill' were nine competence topics that were identified in D7.2.

The Science Demonstrator projects, which pilot the application of EOSCpilot project partners' services and towards scientific goals, were a further source of gaps. Through analysis of their reports, and liaison with their 'shepherds' (project technical representatives), D7.2 derived skills gaps from the challenges experienced in these short projects.

Since the D7.2 report (M12), the Framework has been revised based on further analysis of Science Demonstrator reports, and further workshops to consult the professional groups the Framework addresses, with representation from research institutions and infrastructures.

Analysis of Science Demonstrator reports in Feb 2018 provided an update of the competence gaps reported in D7.2, shown in Table 6 below.

Table 6 Updated gap analysis

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

Further discussion of these, and the plans of the EOSC-hub and OpenAIRE plus projects to address some of them, led to agreement to focus efforts in WP7 on developing specific examples of competence and capability statements to reflect the nine gaps described in Annex B.

4.6. Key workshop outcomes and responses in the framework

The Framework has been presented and discussed in four events since D7.2 was written, as follows:

- 'Building Data Stewardship Expertise in Europe: How can we fill the gaps?' EOSCpilot WP7 Session. Stakeholders Forum Brussels. 28 Nov. 2017
- EDISON workshop, DI4R 2017 conference, 1 Dec. 2017
- 'The EOSC as a skills commons providing FAIR training for FAIR data stewardship'- a co-located half-day WP7 workshop at EUDAT conference, Porto, 25 Jan. 2018
- 'The EOSC as a skills commons' for developing research data stewardship skills at scale' half-day WP7 workshop at IDCC 18 conference, Barcelona, 19 Feb. 2018

The outcomes of these workshops (a full account of which will be given in D7.4) were more relevant to the delivery of skills than the detail of the Framework, but there were five key implications for the Framework:

1) *A distinction should be made between competences that may be applied independently of the domain (such as service management) and those applied through domain-dependent skills; and the Research Infrastructure communities should be listened to on the latter.*

   The Framework is necessarily cross-domain, but offers a format for specifying competences and capabilities that is intended to be easily extensible and applicable in domain-specific contexts.

   Consultation with Research Infrastructures has been planned for the final 6 months of the project, and will gather feedback from training coordinators through a survey, through direct contact with a sample of them and through several further workshops.

2) *Some competences are applied in skills that are independent of services (e.g. FAIR and OA policy) while others are more service specific (e.g. containers).*

   The framework offers training coordinators a 'skills user story' format to specify how capabilities may be delivered by applying competences in a service-specific or a non-specific way.

3) *The Framework should make more explicit links to EDISON professional profiles.*

   This was addressed in Section 3 by suggesting that EDISON profiles be consulted when selecting competences from the Framework to use in defining a job description.

4) *Outputs of the OSPP Skills and Rewards working groups should be taken into account, especially the OS-CAM (Open Science-Careers Assessment Matrix) resulting from the Rewards report.*

5) *The Framework should distinguish between core topics and competence levels that would comprise a minimum standard or set of core topics.*

   The framework in Annex A addresses 4) and 5) by making more explicit the links to the Sapp Nelson competence matrix, which starts at an undergraduate level, and by drawing on the OS-CAM matrix as mentioned above. The OS-CAM is not a competence framework as such, but as a list of OSPP recommended behaviours and indicators of open science activities researchers should undertake, it offers a very useful basis to sharpen the focus of the Framework on 'core' skills that support these activities.

   A workshop planned for September 2018 will address whether the Framework can usefully be mapped to researcher career stages, and potentially be used to identify career stages for data stewards.

5. CONCLUSIONS

This report offers a Framework to help organisations to plan the professional development of their staff, as EOSC service operators or users, and for any individual to identify competences and learning materials that match the capabilities they need. It provides a set of core competences for data stewardship, relating topics to recommended expertise levels for researchers, and the professional groups that support them. In doing so it complements the work of EDISON, FOSTERplus, GO-FAIR and related initiatives to establish educational curricula for data science and data stewardship.

The Framework also offers examples of capability and competence statements, focusing these on skills areas that the work-package has identified as gaps for stakeholders. Repositories are highlighted as an example of the services EOSC will enhance capabilities in. The EOSC Implementation Roadmap has offered guidance on other broad classes of services in EOSC, enabling a high-level analysis of the expertise required. The nature of these services will evolve, as will the capabilities they enhance by fulfilling the needs of research organisations. The Framework will provide operators of these services and others with an interest in skills development with an approach to describing similar competence and capability statements.

The Framework does not and cannot at this stage provide a resource that provides a comprehensive description of the capabilities for the EOSC. Rather than representing a static set of services and infrastructure the EOSC will continue to evolve. The capabilities offered will change with these services, and emerge from combinations of services as well as individual service considered in isolation.

Alongside the prospect of a complex service ecosystem it is important to recognise that EOSC skill development should consist of two parts: (1) Skill development strongly linked with specific services (i.e. how to use service X) and (2) Cross-cutting skills across multiple services (e.g. Research Data Management; Good practices in Open Science; Implementation strategies for FAIR data). Service-specific skill development (training, webinars, etc.) are expected to be the responsibility of the service operators or independent training providers (or, in other words, the skill development is part of the service itself). Cross-cutting skills development must be provided by dedicated initiatives/institutes or again by independent training providers; examples of the former have been collated already in EOSCpilot D7.1 and D7.2.

The Framework could and should be elaborated further to define competence and capability statements for further cross-cutting topics than the nine examples given in Annex B. The value of doing so would be enhanced by getting further feedback on the approach already presented and, if necessary, refining it in the Final Report D7.5.

The fit between the conceptual model used in this report and the service architectures of both projects has been established - in the sense that each recognises a distinction between ‘core’, ‘generic’ and ‘thematic’ services. The Skills Framework contains competences relating to each of these. It also assumes parallel relationships between:

- *Core* training services – a training registry to aggregate metadata from existing training catalogues and portals.
- *Generic* competences – those applied in project-independent ways and easily transferable across domains, applied in skills to govern and assess, to scope and resource, and to advise and enable ‘FAIRified’ research.
- *Thematic* competences – those applied in domain or project-specific ways, and served by domain-specific training services and learning materials.

Whether the EOSC offers a new channel, ‘one stop shop’ or marketplace of these, the skills and competences relevant to FAIR science and its stewardship are relevant beyond EOSC. This begs the question; what *additional* skills and competences are required for the EOSC?

Our expectation is that the additional skills needed specifically for EOSC will mainly be in the mix of

technical and 'soft' generic skills to enable governance, support, and federation of high-quality services. These skills and competences will ensure the data shared through the services, and generated by them, remains FAIR, with exceptions that are acceptable to the communities concerned.

The message we have so far heard from those communities is, broadly, that the 'thematic' skills and competences – those needed to apply FAIR principles in domain and project-specific ways – are best addressed within those communities. At present, this is an interim conclusion based on liaison with Science Demonstrators, and has to be tested through broader consultation with Research Infrastructures.

The work-package has yet to establish how skills and training services will be accommodated in the more general service management and governance frameworks, service portfolio and catalogue (all currently being defined in parallel). This will be reported on in D7.5 after further liaison with project partners. Training community expectations about these relationships are also the subject of consultation during Q3 of 2018.

The necessity to further validate the Framework and our assumptions about how it will be used means that of the conclusions previously stated in D7.2 the first remains very relevant, as follows:

> "The EOSCpilot Skills Framework requires further development … including validation of its conceptual model against the EOSCpilot and EOSC-hub service architectures, and consultation with the target users of the framework in Research Infrastructures and institutions."

Concrete steps have been made to plan workshops and a consultation survey in the latter half of 2018, and these will be reported on further in D7.4, with conclusions and recommendations in D7.5 Final Report.

Annexes

ANNEX A. EOSCPILOT SKILLS AND CAPABILITY FRAMEWORK

A.1. Competences and Capabilities

General notes

1) The table lists the topics of 59 competences relevant to professional groups involved in open science data stewardship, according to their role in service application as a user or operator.
2) The Framework assumes that individuals in all professional groups can play different roles for different services, e.g. researchers may be involved in both operating and using a service to support data management planning.
3) Competences and expertise levels for service operators should be read as those needed to operate services generally in the EOSC environment, rather those needed to enhance the same competence for service users.
4) The Framework does not take into account the additional competences or levels of expertise that may be needed to use or operate different classes of service, or specific services that enhance particular skills. Our assumption is that service providers will specify the capabilities their service offers organisations, and the competences and expertise levels involved. The model offers a framework and method to do that. Examples are given in Annex B, and represent gaps that were identified in the first 14 months of the project.
5) The MIN column indicates the 44 recommended competences that represent a basic undergraduate-level knowledge of data stewardship in the open science context. These are marked ○ and each maps to statements from one of two sources for the Framework:
    - Undergraduate-level 'data information literacy' skills in the Scaffolding for Data Management Skills (Sapp Nelson, 2017), a recommended starting point for stewardship skills development.
    - Open Science Career Assessment Matrix proposed by the working group on Rewards under the Open Science Policy Platform.

EOSCpilot Skills and capability framework Version 1.0 for consultation

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

Open science stewardship competences and capabilities		Recommended expertise by professional group and service role								
			Service users				Service operators			
		MIN.	Researcher	Data Scientist/ analyst	Data service engineer	Data manager /curator	Researcher	Data Scientist/ analyst	Data service engineer	Data manager /curator
Plan and design	Planning data management and sharing (DMP)	○	◖	◖	○	●	●	●	○	●
	Open data model and database design	○	○	●	●	○	◖	◖	◖	●
	Metadata, persistent id. specification	○	○	◖	○	◖	○	○	◖	◖
	Open source software / service requirements	○	○	●	●	◖	◖	●	◖	●
	Repository and data management platform appraisal	○	○	◖	○	●	◖	○	◖	●
Capture and process	Workflow set up and documentation	○	◖	◖	○	●	○	○	○	○
	Database management	-	-	◖	●	○	○	◖	●	◖
	Software prototyping	-	-	◖	●	○	◖	●	●	◖
	Data collection and reuse of open data	○	◖	●	○	◖	●	○	○	○
	File naming and organisation	○	◖	◖	◖	●	◖	◖	◖	◖
	Data provenance and software versioning	○	◖	●	●	●	◖	●	●	◖
Integrate and analyse	Math and statistical knowledge application	○	◖	●	◖	○	◖	○	○	○
	Critical thinking and theory building	○	◖	●	◖	○	◖	○	○	○
	Creative problem solving, flexibility	○	◖	●	●	◖	◖	●	○	○
	Open source software / service development	○	○	●	●	◖	◖	●	●	●
	Data transformation and integration	-	○	●	●	◖	◖	●	○	○
	Data mining, querying, interpretation	○	○	●	○	○	○	●	○	○
	Predictive modelling and analytics	-	○	●	○	○	○	○	○	○
	Machine learning methods	-	○	●	○	○	○	○	○	○

Open science stewardship competences and capabilities		Recommended expertise by professional group and service role								
			Service users				Service operators			
		MIN.	Researcher	Data Scientist/ analyst	Data service engineer	Data manager /curator	Researcher	Data Scientist/ analyst	Data service engineer	Data manager /curator
Appraise and preserve	Data preparation, documentation for reproducibility	○	◖	◖	○	●	●	●	○	◖
	Data quality assurance using open standards	○	◖	◖	○	●	◖	◖	◖	◖
	Ethical, legal and data policy compliance	○	◖	◖	◖	●	◖	◖	◖	◖
	Data transfer and long-term storage	○	◖	◖	●	◖	○	◖	◖	◖
	Format and media migration	○	◖	○	●	●	○	○	◖	◖
	Software review and preservation	-	○	◖	●	○	○	◖	◖	◖
Publish and release	Documentation for public use, lay understanding	-	○	◖	●	○	●	◖	◖	◖
	Access control and management	-	○	●	●	◖	○	●	◖	◖
	Ethical application of patents, licenses	○	◖	◖	◖	◖	◖	◖	◖	◖
	Open access publishing and self-archiving	○	◖	◖	○	●	○	○	○	○
	Sharing via open repositories/ platforms	○	◖	◖	◖	●	○	○	◖	◖
	Engaging in open innovation beyond academia	○	◖	◖	◖	○	◖	◖	○	○
Expose and discover	Vocabulary/ ontology application	○	◖	●	◖	○	○	◖	◖	○
	Metadata and persistent id. exposure	○	◖	●	◖	●	○	●	○	○
	Visualisation and presentation of results	○	◖	●	◖	○	○	○	○	○
	Evaluation of repository and publishing platforms	○	◖	◖	○	●	◖	○	○	○
	Repository/ database search	○	◖	◖	●	●	○	◖	○	○
	Citation of research outputs	○	◖	◖	◖	●	○	○	○	○
Govern and assess	Research strategy/ open research vision	○	◖	●	◖	●	◖	○	○	○
	Advocacy of FAIR principles and Open Access policy	○	◖	◖	◖	●	◖	○	○	○
	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	○	◖	◖	◖	○	○	◖	○	○

Open science stewardship competences and capabilities		Recommended expertise by professional group and service role								
			Service users				Service operators			
		MIN.	Researcher	Data Scientist/ analyst	Data service engineer	Data manager /curator	Researcher	Data Scientist/ analyst	Data service engineer	Data manager/ curator
Scope and resource	Securing funding for open science /support	○	◖	◖	◖	◖	◖	○	○	○
	Perseverance delivering diverse open research	○	◖	◖	◖	◖	◖	○	◖	●
	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	Building open inter-disciplinary collaborations	○	◖	◖	◖	○	●	◖	◖	◖
	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 res, networks, standards bodies	○	◖	●	●	●	○	◖	○	○
	Supervision and mentoring	-	○	○	○	○	◖	◖	◖	◖

This page intentionally left blank

This page intentionally left blank

ANNEX B. SKILLS AND CAPABILITY EXAMPLES FOR EOSC SERVICES

B.1. Repository and data management platform appraisal and selection

Scope

Shortlisting and selection of repositories or similar platforms (e.g. scientific databases) suitable for sharing and preserving research data or related outputs of research (e.g. source code). Consideration of policies (institution, journal, funder), repository collection policies, and other relevant criteria and standards, including the various certification schemes for Trusted Digital Repositories and for information security.

Individual competences

Comprehend: understand criteria for selecting repositories or similar platforms (e.g. scientific databases) including relevant trusted digital repository certification standards. Be able to contact the operators of repositories/platform in question and ask relevant questions regarding the deposit and ingestion workflows, commitment to long-term data storage etc.

Apply: Be able to select a repository or data management platform on identified criteria to ensure the specific needs of a project or dataset will meet user and stakeholder expectations regarding implementation of FAIR principles.

Evaluate and synthesize: Be able to assess and determine appropriate criteria to meet user and stakeholder expectations regarding implementation of FAIR principles.

Capabilities

Team: Supports team members to choose repository or data management platforms on identified criteria appropriate to the needs and expectations of project users and stakeholders

Organisation: Supports research groups to assess and determine the appropriate criteria to meet user and stakeholder expectations regarding implementation of FAIR principles.

Skills user stories

Comprehend: As a researcher, and repository service user, I need to be able to understand criteria for selecting repositories or similar platforms (e.g. scientific databases) including relevant trusted digital repository certification standards. Be able to contact the operators of repositories/platform in question and ask relevant questions regarding the deposit and ingestion workflows, commitment to long-term data storage etc.

Apply: As a data manager, and as a user of a repository service, I need to be able to assess and determine appropriate criteria to meet user and stakeholder expectations regarding implementation of FAIR principles.

Evaluate and synthesize: As a data manager, and as an operator of a repository service, I need to be able to assess and determine appropriate criteria to meet user and stakeholder expectations regarding implementation of FAIR principles.

Relevant learning resources

- Websites where repositories are registered along with their certification status include the Registry of research data repositories Re3data.org, the directory of open access repositories opendoar.org, and registry of open access repositories roar.eprints.org.
- Guidance on the European Trusted Digital Repository certification scheme and on each of its three layers: CoreTrustSeal, DIN 31644/nestorSeal, and ISO 16363. http://www.trusteddigitalrepository.eu/Trusted%20Digital%20Repository.html
- FORCE11 guide for researchers to choose a FAIR data repository in the form of FAIR Data Decision Tree https://docs.google.com/document/d/100Yv2hd9yfrlOscrMiGLqwqW_VWLruxRGoMr68YstNg/edit

B.2. Data preparation and documentation for reproducibility

Scope

Preparation of research objects and related documentation tasks, including data collection, data cleaning, filtering, selection and consolidation of data, packaging of files with narrative description (e.g. a 'readme' file) of what the data comprises, its size, sources, what processes and transformations have been applied to it, by whom and when,, plus structured information about the data (i.e. 'metadata'). This might e.g. include a codebook describing variables, their labels, and validity ranges. It may also include details specific to the research domain (see RDA Metadata Standards Catalogue).

Individual competences

Comprehend: The individual understands the importance of reproducibility and knows about necessary steps in conducting reproducible research.

Apply: apply relevant tools and domain standards to document steps taken in research, e.g. to apply methods to process data, and record the methods applied sufficiently to enable others to reproduce these steps.

Synthesize and evaluate: assess and determine relevant tools and domain standards to prepare and document the steps taken to process data sufficiently to enable others to reproduce these steps.

Capabilities

Team: Support the adaptation of workflows and policies to prepare and document data for reproducibility.

Organisation: Provide policies and infrastructure to enable research teams to assess and determine the relevant tools and domain standards they need to ensure data is reproducible from publicly funded research.

Skills user stories

Comprehend: As a researcher needing to integrate a data set from another research group and as a user of an EOSC workflow management service, I need to understand the importance of reproducibility and know about necessary steps in conducting reproducible research.

Apply:

- As a researcher needing to publish raw data that was recorded during an experiment and as a user of the EOSC repository service I need to be able to apply relevant tools and domain standards to document steps taken in research, e.g. to apply methods to process data, and record the methods applied sufficiently to enable others to reproduce these steps.
- As a data scientist needing to share intermediate data with interested research group and as a user of EOSC repository service I need to be able to apply relevant tools and domain standards to document steps taken in research, e.g. to apply methods to process data, and record the methods applied sufficiently to enable others to reproduce these steps.

Synthesize and evaluate:

- As a researcher, and operator of an EOSC recommender service, I need to be able to assess and determine relevant tools and domain standards to prepare and document the steps taken to process data sufficiently to enable others to reproduce these steps.
- As a data service engineer wanting to operate a repository for research data I need to be able to be able to assess and determine relevant tools and domain standards to prepare and document the steps taken to process data sufficiently to enable others to reproduce these steps.

Relevant learning resources

- Research Data Management: https://zenodo.org/record/154433#.Wxrfqq2B3x4
- Understanding research data, Sharing data, Archiving data: https://www.coursera.org/learn/data-management#creators
- The reproducibility problem, Open materials, data and code: https://osf.io/wxhq6/

B.3. Ethical, legal and data policy compliance

Scope

Engage with stakeholders to build awareness of the regulatory and policy environment as well as the ethical implications for open research and data science, identify what they require in order to respond effectively, and develop the strategies, policies, and processes that enable their needs to be met.

Individual competences

Comprehend: Understand the diversity of data and their management needs across the research data lifecycle, be able to identify the components of good data management plans; be familiar with best practices for working with data, and the ethical implications of data collection and management.

Apply: Implement/adapt policies and protocols based on funder requirements, existing policies on data, legal frameworks, ethical principles and regulations. Ensure the ethical application of patents, licenses and other Intellectual Property Rights, consistently with FAIR principles.

Synthesise and evaluate: Review plans and protocols to ensure they reflect changes in relevant stakeholder requirements.

Capabilities

Team: Enables the team to adapt their research practices to the relevant regulatory regime and to the appropriate ethical policy environment in domain-relevant ways.

Organisation:

- Provides researchers with appropriate knowledge and support to understand regulatory, ethical and policy requirements affecting their research.
- Enables researchers (offers them the tools and instruments/architecture) to adapt their specific domain (data/software) policy towards implementing national/European laws, regulations and guidelines.
- Coordinates the development and implementation of institutional policies based on funder requirements and existing data/legal/ethical policies.

Skills user stories

Comprehend:

- As a Domain Researcher wanting to make my research data accessible and reusable by other researchers and as a user for Repository Services I need to be able to implement/adapt policies and protocols based on funder requirements, existing policies on data, legal frameworks, ethical principles and regulations.

Apply:

- As a Domain Researcher wanting to provide research integrity, attribution, impact awareness, and as system manager for Repository Services, I should be able to understand the diversity of data and their management needs across the research data lifecycle, be able to identify the components of good data management plans; be familiar with best practices for working with data, and the ethical implications of data collection and management.
- As a Domain Researcher wanting to comply with my organisation's strategies, policies, and as user of the Open Science Monitor, I should be able to ensure the ethical application of patents, licenses and other Intellectual Property Rights, consistently with FAIR principles.

Synthesise and evaluate:

- As a data manager/ librarian wanting to maintain and validate my organisation's strategies, policies, and processes on FAIR research outputs, and as user of the Open Science Monitor, I need to be able to review plans and protocols to ensure they reflect changes in relevant stakeholder requirements.

Relevant learning resources

- Research Data Management at CODATA-RDA Summer School in Research Data Science Aug 1-12 2016: https://zenodo.org/record/154433#.Wxrfqq2B3x4
- MANTRA Research Data Management Training: https://mantra.edina.ac.uk/
- Understanding research data, Sharing data, Archiving data: https://www.coursera.org/learn/data-management#creators
- FOSTER project E-Learning Portal: https://www.fosteropenscience.eu/resources
- LEARN RDM Project: http://learn-rdm.eu
- S. Battaglia,S. Canham, V. Cavalli, F. Molnar-Gabor, C. Ohmann, E. Papadopoulou, D. Robertson, P. Rouse, A. Schenk, R. Thorat, P. Tsiavos, M. Thorley "EOSCpilot D3.3 Draft Policy Recommendations" especially the supporting White Paper on Ethics, which lists extensive literature on Ethical policies.

B.4. Sharing via open repositories / platforms

Scope

Provide access to research objects and outputs and ensure they continue to offer potential value to others, by applying repository policies and digital object management functions and related services, including any integration with other tools or services needed to ensure the preserved objects remain FAIR for as long as required.

Individual competences

Comprehend: know about repository processes for ingesting research objects, what descriptive information and metadata they need to plan for preservation, and to provide the appropriate level of access for the objects' designated communities.

Apply:
Follow repository processes in order to deposit research objects, provide descriptive information and metadata to enable the repository to plan for preservation, and to provide the appropriate level of access for the objects' designated communities.

Apply repository processes for ingesting research objects, handle the associated bitstreams, provide descriptive information and metadata for preservation, and specify the appropriate level of access for the objects' designated communities.

Synthesise and evaluate: develop, maintain and review repository processes for ingesting research objects, handling the associated bitstreams, descriptive information and metadata for preservation, and ensuring the appropriate level of access for the objects' designated communities.

Capabilities

Research team:

- Develops processes and guidelines for depositing research data and related objects in repositories relevant to the research projects, domains, and institutions involved, also considering relevant journals, museums, archives, or national and international governmental repositories.
- Provides sufficient information for others to assess the scientific and scholarly quality of the research objects, their compliance with FAIR principles, disciplinary and ethical norms, and their value for reuse.

Organisation:

- Develops cross-domain policies and guidelines for publishing research data and related objects, and for selecting repositories that comply with relevant regulatory and policy frameworks.
- Provides repository functions where necessary to ensure preserved research objects remain accessible for as long as required, enabling these functions to interoperate with other systems or processes that provide or extract information to enhance the research objects' compliance with FAIR principles.
- Provides a level of curation appropriate to the value of preserved research objects, from oversight of their compliance with policies and guidelines, through to enhancements of their findability, accessibility, interoperability and reusability, to meet specific user requirements.

Skills user stories

- As a domain researcher needing to share via open repositories/ platforms, and a repository service user, I need to follow repository processes in order to deposit research objects, provide descriptive information and metadata to enable the repository to plan for preservation, and to provide the appropriate level of access for the objects' designated communities.

Relevant learning resources

- Ulrich, Robert – Making research data repositories visible and discoverable. https://www.fosteropenscience.eu/node/592
- CoreTrustSeal: https://www.coretrustseal.org/about/

- COAR (2017) Next Generation Repositories: https://www.coar-repositories.org/activities/advocacy-leadership/working-group-next-generation-repositories

B.5. Metadata and persistent id. exposure

Scope

To enable recommendation, search and browse, workflow, citation, attribution and reward services within EOSC in terms of findability, accessibility and reusability metadata and persistent identifiers for research artefacts need to be exposed.

Individual competences

Comprehend: Understand the importance of exposing metadata and persistent identifiers to enable FAIR research artefacts and know about tools and standards to manage and expose metadata as well as persistent identifiers.

Apply: Use standards and tools to publish metadata and assign persistent identifiers to research artefacts.

Synthesize and evaluate: Review of standards, schemas and tools to expose metadata and persistent identifiers.

Capabilities

Team: Supports and enables the process of exposing metadata and persistent identifiers by providing relevant information on the workflow, policies and tools that may be used.

Organisation: Provides access to or provision of the required infrastructure and tools to expose metadata and persistent identifiers.

Skills user stories

Comprehend:

- As a domain researcher wanting to publish research artefacts of my research and as a user of the Recommender service I should know about the need to expose persistent identifiers and their advantages, regarding the FAIRness of data, to efficiently make the data findable and accessible for the public.
- As a data science engineer needing to enable the FAIR exposure of research artefacts and their versioning in repositories and as a service developer of Repository service I should know about the relevance of persistent identifiers and about the relevant versioning concepts.

Apply:

- As a data science engineer needing to integrate reliable bit storage for archiving with integrated DOI generation and as a user of PID minting I should be able to apply tools and techniques to expose metadata and persistent identifiers.
- As a domain researcher needing to publish my research to make it available to the public, and as a service user of Citation, Attribution and Reward service, I should be able to assign metadata and create persistent identifiers to properly take care to promote my work to maximise citation.

Synthesize and evaluate:

- As a data science engineer needing to create a service to support the association of persistent identifiers to research artefacts for the EOSC, and as a service developer of PID minting, I must be expert in exposing metadata and persistent identifiers to be able to evaluate existing tools and techniques for utilisation.
- As a domain researcher needing to maximise my research impact and the visibility of my research and as a service user of Repository service I should be able to evaluate policies and functionality of different services to expose metadata and generate persistent identifiers.

Relevant learning resources

- Research Data Management: https://zenodo.org/record/154433#.Wxrfqq2B3x4
- Open Science: https://zenodo.org/record/154433#.Wxrfqq2B3x4
- Research Data Management and Sharing: https://www.coursera.org/learn/data-management
- Open Science and Reproducible Research: https://osf.io/fa9rg/

B.6. Advocating FAIR principles and Open Access policy

Scope

The application of FAIR and open access principles is advocated and supported across the organisation, by identifying how these principles relate to research goals and practices, and to the organisation's local policies and processes.

Individual competences

Comprehend: has knowledge about FAIR sub principles, knows various open access policies, has awareness of relevant legal and ethical aspect of data collection, analysis, storage and reuse.

Apply: contextualises the FAIR and open access principles of specific research domains and journals relevant to the organisation, relating them to local research practices and forms of research output, considering the needs of academic and non-academic groups and citizens.

Synthesise and evaluate: identifies strengths and weaknesses in how principles are applied, including relevant exceptions and needs for change in policy or practice, to reflect methodological or technological changes in relevant research domains.

Capabilities

Team: motivates FAIR and open access practices among researcher colleagues, and supports them in ensuring that relevant data or other research outputs are made FAIR/Open in accordance with domain norms and stakeholder expectations.

Organisation: proposes and encourages FAIR and open access policies within the organisation, ensuring that FAIR and open access policy is implemented, monitored, and integrated into the organisation's processes for handling research outputs.

EOSC services that support these capabilities and enable individuals to develop the relevant competences include the Open Science Monitor. This is described[23] as a set of services for supporting Research Performing Organisations (RPOs), Research Funding Organisations (RFOs) and Government Bodies to measure:

- levels of compliance with European Union's laws, regulations and policies regarding research and research results dissemination;
- Open Science Resources' (i.e. research artefacts, educational resources, research collaboration, citizen science) levels of openness, trustworthiness and FAIRness that cover each stage of the research lifecycle;
- excellence of science, which includes quantitative and qualitative metrics of different levels (bibliometrics, webometrics, scientometrics etc);
- impact of science on society and economy.

Skills User Stories

Skills User stories that express what individuals need from training and learning resources include:

- As a domain researcher I need to apply FAIR principles and various open access policies when govern and assess the research output to use open science monitor service;
- As a data manager I need to advise on FAIR and open access principles when researchers want to make their data and relevant research output FAIR and open to use open science monitor services;
- As a data manager I need to be an expert in training the organizational research teams in FAIR and open-access policy skills to integrate the open science monitor services.

[23] Eoscpilot D5.1 P. 32

Relevant learning resources

- Briefing paper Research Data Management (OpenAIRE), https://www.openaire.eu/briefpaper-rdm-infonoads, April 2017.
- Blended-learning course Essentials 4 data supporters (RDNL), http://datasupport.researchdata.nl.
- Online guide that can be used for local training workshops in DM: CESSDA Training Working Group (2017). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC. Retrieved from https://www.cessda.eu/DMGuide.

Webinars:

- How to manage your data to make them Open and FAIR, OpenAIRE-EOSC-hub webinar, 15 May 2018, https://www.openaire.eu/how-to-manage-your-data-to-make-them-open-and-fair.
- Open Research data in Horizon 2020, 9 May 2018, https://www.openaire.eu/open-access-to-publications-in-horizon-2020.
- Open Access to Publications in Horizon 2020, 8 May 2018, https://www.openaire.eu/2018-05-09-09-06-13.
- Results Survey on Horizon 2020 DMP template, 15 January 2018, https://www.openaire.eu/openaire-webinar-results-survey-on-h2020-dmp-template.
- FAIR Data in Trustworthy Data Repositories, DANS-EUDAT-OpenAIRE webinar, 12-13 December 2016, https://www.openaire.eu/fair-data-in-trustworthy-data-repositories-webinar-dans-eudat-openaire-webinar-dec-2016.

B.7. Cloud environment and storage management

Scope

Delivering facilities by the as-a-Service provision mode, e.g. 'Infrastructure as a Service' (IaaS) and 'Platform as a Service' (PaaS) including the management of virtual machines, virtual disks, and containers at IaaS and PaaS providers.

Individual competences

Comprehend: Understand the 'as a Service' cloud computing model, and know the most frequently used compute and storage services of at least one IaaS/PaaS provider.

Apply: Ability to design, deploy and operate applications and application services within IaaS/PaaS environments.

Synthesize and evaluate: Monitoring and optimization of applications and application services within IaaS/PaaS clouds.

Capabilities

Team: Define and integrate applications with EOSC; Turn applications into services that are federated into EOSC.

Organisation: Define policies and rules in choosing and using IaaS/PaaS services from EOSC and federating services into EOSC.

Skills user stories

- As a research-model developer I need to be able to access IaaS cloud resources and manage Linux distributions and my own code within Virtual Machines and Virtual Storage instances.
- As service developer working for a research project I need to be able to deploy, configure and operate software services within IaaS and PaaS cloud sites provided via EOSC.

Relevant learning resources

- 'Introduction to Cloud Computing' courses from EGI members.
- OpenStack training.
- https://acloud.guru online course.
- Introduction to Linux administration.

B.8. Service selection and deployment

Scope

EOSC is expected to federate a large number of services from diverse providers during the next few years. Users and future providers building on EOSC hub services must possess basic skills on navigating the service catalogue and choosing services and service groups that fit their specific purpose. While in-depth knowledge of specific services is necessary for domain scientists (and operators), skills to orient among service groups/classes within the catalogue is a must for all users.

Individual competences

Comprehend: Be familiar with the EOSC service catalogue and the service areas within the catalogue.

Apply: Ability to navigate the EOSC service catalogue and identify relevant services and service groups for a given purpose.

Synthesize and evaluate: Ability to find technical details, and to compare capabilities of services from the EOSC catalogue.

Capabilities

Team: Critical assessment of services and service groups in EOSC to decide whether they fit the purposes/use cases of the institute or research group.

Organisation: Define policies and rules in choosing and using services from EOSC.

Skills user stories

- I am a researcher working on biodiversity modelling, using both third-party applications and my own models. I need to find relevant services for my field from EOSC, and be able to assess and compare them to decide which one fits to my research purposes.

Relevant learning resources

- EOSC service catalogue overview course (does not yet exist).

B.9. Training and Support

Scope

High quality user support will be crucial to enlarge the support of EOSC within user and provider communities. Federated providers as well as members of the EOSC hub will need to be skilled in understanding about good practices, and tools adopted for user support and training.

Individual competences

Comprehend: Familiarity with the user support processes, best practices and tools used in EOSC support and training.

Apply: Ability to act as user support personnel in EOSC, particularly in the negotiation of Service Level Agreements, responding to support requests, handling of complaints, performing customer satisfaction interviews.

Synthesize and evaluate: Ability to identify gaps and improvement-opportunities in the EOSC user support processes, and ability to fill these gaps and implement the improvements.

Capabilities

Team: Supports the adaptation of local user-support processes and tools into the broader EOSC context.

Organisation: Define policies and rules in supporting users coming through the EOSC channel to organizational services.

Skills user stories

- I recently joined a research team that develops new algorithms in molecular simulations and operates online services for RIs and EOSC communities in this domain. I am responsible for user support so need to become familiar with the requirements, tools and practices of customer/user relationship management.

Relevant learning resources

- FitSM trainings (Foundation level; Advanced level in Service Planning and Delivery; Expert level): https://www.egi.eu/services/fitsm-training/
- EOSC user supporter training (does not yet exist).

ANNEX C. MAPPINGS TO OTHER FRAMEWORKS

C.1. Data Information Literacy

Source: Sapp Nelson (2017) Scaffolding Spreadsheet v3.0 Personal Domain

Data Information literacy - knowledge required of learners	*EOSCpilot WP7 Skills and Capability Framework*
Databases, Data Sets, and Data Formats	
1. Develop search strategies using Boolean logic, faceted search and other techniques as appropriate to the discipline to identify and access data sets relevant for a specific purpose.	Repository/ database search
2. Learners identify standard or preferred data formats for their academic disciplines, and select appropriate data formats for a given task.	Planning data management and sharing (DMP)
3. Learners identify proprietary data formats relevant to their academic disciplines, articulate potential problems with the use of proprietary data formats, and select alternative formats when appropriate.	-- as above –
4. Learners comprehend that data sets are collected into groups by topic or by commonality such as place of origin, that the technology housing these groups are called repositories, and that they can be used to access data sets relevant to a research topic.	Repository and data management platform appraisal
Discovery and Acquisition of Data	
5. Consider authority of data set, quality of data set as indicated by existing metadata and documentation, and other critical factors as determined by research question and goals when selecting data for reuse.	Data collection and reuse of open data
6. Download tabular (.csv, .tsv, .tab) data sets and critically consider the contents of the table with a goal of recombining relevant data into a new data table.	Open data model and database design
7. Identify sources (organizations, agencies) that may produce data relevant for a specific purpose.	Planning data management and sharing (DMP)
Data Management and Organization	
8. Explain the scholarly communication life cycle and that data are produced during the earliest phases of the scholarly communication lifecycle.	Advocacy of FAIR principles and Open Access policy
9. Explain the concept of a data management plan, its constituent parts and its use as an organizational document of research endeavours.	Planning data management and sharing (DMP)
10. Use a rubric to identify the credibility of databases that they find online in order to verify currency, relevance, authority, accuracy, and purpose.	Data governance, handling third-party data

Quality Assurance 11. State that data errors can be prevented by stating measurement units, data fields to be captured and formats for data captured prior to starting data capture. 12. State that data corruption can be caused by any number of interactions with data, including human error, computer file creation or transformation problems.	Data quality assurance using open standards Data quality assurance using open standards
Data Conversion and Interoperability 13. Change files from one format to another as part of the workflow. 14. Save files in standard file formats. 15. Awareness of the difference between proprietary and non-proprietary formats and understand why using non-proprietary formats for files that they want to access in 5-10 years is best.	Workflow set up and documentation File naming and organisation Data preparation and documentation for reproducibility
Metadata 16. Understand that metadata describes the context in which a data set was created and answers the questions of Who, What, When, Where, and Why about that data set. 17. Define the concept of a standard and recognize that standards govern physical objects, processes and methodologies and that metadata is a methodology governed by standards. 18. Recognise basic fields that frequently appear in metadata records. 19. Discuss the concept of ontologies applied to disciplinary applications such as physics, earth sciences, curriculum development, etc.	Data provenance and software versioning Data preparation and documentation for reproducibility Metadata, persistent id. Specification Vocabulary/ ontology application
Data Curation and Re-Use 20. Recognise that scientific reproducibility requires reuse of scientific data. 21. Understand the cost associated with the production and access of all information. 22. Recognise data management plans (DMPs) are artefacts that describe research projects and practices in their disciplines, and look to data management plans to learn about their disciplinary culture.	Data preparation and documentation for reproducibility Costing of data management and preservation Planning data management and sharing (DMP)
Cultures of Practice 23. Recognise data sets as a part of the domain knowledge, and learn to find and use data sets along with scholarly journal articles and monographs as appropriate for their discipline. 24. Define the concept of a standard and recognize that standards govern physical objects, processes and methodologies and that metadata is a methodology governed by standards.	Critical thinking and theory building Contributing to quality assessment or peer-review/ Metadata and persistent id. exposure

Data Preservation 25. Understand that there is a cost associated with the preservation of all information. 26. Integrate regular backups and Lots of Copies Keeps Stuff Safe (LOCKSS) to preserve personal files. 27. Recognize that some information has more value than other information and will need more preservation efforts.	Costing of data management and preservation Data transfer and long-term storage Data preparation and documentation for reproducibility
Data Analysis 28. Employ analytical methods during applied research opportunities such as undergraduate research fellowships, internships, or undergraduate research experiences. 29. Employ workflow management tools during applied research opportunities such as undergraduate research fellowships, internships, or undergraduate research experiences.	Creative problem solving, flexibility/ Math and statistical knowledge application/ Data mining, querying, interpretation Workflow set up and documentation
Data Visualization 30. Employ visualization tools during applied research opportunities such as undergraduate research fellowships, internships, or undergraduate research experiences. 31. List a variety of types of visualizations, note those that are frequently used within their discipline, and recognize them by format or appearance 32. Recognize ambiguity and misleading presentation of data in visualizations authored by others and critique how the ambiguity is engendered.	Visualisation and presentation of results -- as above – Critical thinking and theory building
Ethics, including citation of data 33. Identify data as the intellectual property of a person or entity similar to articles, books, or music. 34. Cite data as well as articles, books, or any other resource used in compiling a new scholarly work. 35. State that data including information about individual persons are subject to greater care and scrutiny for research and data management. 36. Locate the ethical statements for disciplinary bodies and investigate them for the presence of information about data management or handling.	Ethical application of patents, licenses Citation of research outputs Research integrity, attribution, impact awareness Ethical, legal and data policy compliance

C.2. Open Science Career Assessment Matrix (OS-CAM)

Open Science Career Assessment Matrix (OS-CAM)		*EOSCpilot WP7 Skills and Capability Framework*
Open Science activities	*Possible evaluation criteria*	*Competences/ Capabilities*
RESEARCH OUTPUT		
Research activity	1. Pushing forward the boundaries of open science as a research topic	1. Research strategy and open research vision
Publications	2. Publishing in open access journals 3. Self-archiving in open access repositories	2. Open access publishing and self-archiving 3. (as 2 above)
Datasets and research results	4. Using the FAIR data principles 5. Adopting quality standards in open data management and open datasets 6. Making use of open data from other researchers	4. Advocacy of FAIR principles and Open Access policy 5. Data quality assurance using open standards 6. Data collection and reuse of open data
Open source	7. Using open source software and other open tools 8. Developing new software and tools that are open to other users	7. Open source software / service requirements 8. Open source software / service development
Funding	9. Securing funding for open science activities	9. Securing funding for open science /support
RESEARCH PROCESS		
Stakeholder engagement / citizen science	10. Actively engaging society and research users in the research process 11. Sharing provisional research results with stakeholders through open platforms (e.g. Arxiv, Figshare) 12. Involving stakeholders in peer review processes	10. Engaging with research users and stakeholders 11. Sharing via open platforms 12. Contributing to quality assessment or peer review
Collaboration and Interdisciplinarity	13. Widening participation in research through open collaborative projects 14. Engaging in team science through diverse cross-disciplinary teams	13. Building open inter-disciplinary collaborations 14. (as 13 above)

Open Science Career Assessment Matrix (OS-CAM)		*EOSCpilot WP7 Skills and Capability Framework*
Open Science activities	*Possible evaluation criteria*	*Competences/ Capabilities*
Research integrity	15. Being aware of the ethical and legal issues relating to data sharing, confidentiality, attribution and environmental impact of open science activities 16. Fully recognizing the contribution of others in research projects, including collaborators, co-authors, citizens, open data providers	15. Ethical, legal and data policy compliance 16. Providing research integrity, attribution, impact monitoring
Risk management	17. Taking account of the risks involved in open science	17. Information security and risk management
SERVICE AND LEADERSHIP		
Leadership	18. Developing a vision and strategy on how to integrate OS practices in the normal practice of doing research 19. Driving policy and practice in open science 20. Being a role model in practicing open science	18. Research strategy and open research vision (as 1 above) 19. Driving application of FAIR principles and Open Access policy (as 4 above) 20. (as 4,19 above)
Academic standing	21. Developing an international or national profile for open science activities 22. Contributing as editor or advisor for open science journals or bodies	21. Developing a profile of open research 22. Contributing to open research, networks or standards bodies
Peer review	23. Contributing to open peer review processes 24. Examining or assessing open research	23. Contributing to quality assessment or peer review 24. (as 23 above)
Networking	25. Participating in national and international networks relating to open science	25. Contributing to open research, networks or standards bodies (as 22 above)
RESEARCH IMPACT		
Communication and Dissemination	26. Participating in public engagement activities 27. Sharing research results through non-academic dissemination channels 28. Translating research into a language suitable for public understanding	26. Engaging with research users and stakeholders 27. Sharing via open platforms 28. Documentation for public use, lay understanding
IP (patents, licenses)	29. Being knowledgeable on the legal and ethical issues relating to IPR 30. Transferring IP to the wider economy	29. Ethical application of patents, licenses 30. (as 29 above)

Open Science Career Assessment Matrix (OS-CAM)		*EOSCpilot WP7 Skills and Capability Framework*
Open Science activities	*Possible evaluation criteria*	*Competences/ Capabilities*
Societal impact	31. Evidence of use of research by societal groups 32. Recognition from societal groups or for societal activities	31. Providing research integrity, attribution, impact monitoring (as 16 above) 32. (as 16, 31 above)
Knowledge exchange	33. Engaging in open innovation with partners beyond academia	33. Engaging in open innovation beyond academia
TEACHING AND SUPERVISION		
Teaching	34. Training other researchers in open science principles and methods 35. Developing curricula and programs in open science methods, including open science data management 36. Raising awareness and understanding in open science in undergraduate and masters' programs	34. Training in open methods, services 35. Contributing to education and professional development 36. (as 35 above)
Mentoring	37. Mentoring and encouraging others in developing their open science capabilities	37. Training in open methods, services (as 34 above)
Supervision	38. Supporting early stage researchers to adopt an open science approach	38. Driving application of FAIR principles and Open Access policy (as 4 above)
PROFESSIONAL EXPERIENCE		
Continuing professional development	39. Investing in own professional development to build open science capabilities	39. Contributing to education and professional development (as 35, 36 above)
Project management	40. Successfully delivering open science projects involving diverse research teams	40. Perseverance delivering diverse open research projects
Personal qualities	41. Demonstrating the personal qualities to engage society and research users with open science 42. Showing the flexibility and perseverance to respond to the challenges of conducting open science	41. Advocacy of FAIR principles and Open Access policy (as 4, 38 above) 40. Perseverance delivering diverse open research projects (as above) 42. Creative problem solving, flexibility

Open Science Career Assessment Matrix (OS-CAM)		*EOSCpilot WP7 Skills and Capability Framework*
Open Science activities	*Possible evaluation criteria*	*Competences/ Capabilities*
TEACHING AND SUPERVISION		
Teaching	34. Training other researchers in open science principles and methods 35. Developing curricula and programs in open science methods, including open science data management 36. Raising awareness and understanding in open science in undergraduate and masters' programs	34. Training in open methods, services 35. Contributing to education and professional development 36. (as 35 above)
Mentoring	37. Mentoring and encouraging others in developing their open science capabilities	37. Training in open methods, services (as 34 above)
Supervision	38. Supporting early stage researchers to adopt an open science approach	38. Driving application of FAIR principles and Open Access policy (as 4 above)
PROFESSIONAL EXPERIENCE		
Continuing professional development	39. Investing in own professional development to build open science capabilities	39. Contributing to education and professional development (as 35, 36 above)
Project management	40. Successfully delivering open science projects involving diverse research teams	40. Successfully delivering diverse open research projects
Personal qualities	41. Demonstrating the personal qualities to engage society and research users with open science 42. Showing the flexibility and perseverance to respond to the challenges of conducting open science	41. Driving application of FAIR principles and Open Access policy (as 4, 38 above) 42. Creative problem solving, flexibility and perseverance

ANNEX D. GLOSSARY

Term	Explanation
Capability	*Competence* applied at a research team or organisational level, with a defined level of expertise and responsibility, to perform a service role or work in the EOSC environment.
Competence	An element (topic) of theory or practice e.g. 'workflow set-up and management', combined with an *expertise level* to indicate whether someone has an awareness of the area, or an ability to do it, or expert knowledge of it.
Expertise level	A description of level of *competence*; e.g. comprehension, ability to apply, or expert knowledge of the given competence; i.e. be able to evaluate its application or synthesise new ways of applying the relevant knowledge.
Professional group	A person's domain of responsibility defined by a set of *competences*; 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.
Service role	Role in the application of a *service*; i.e. service operator, or user.
Skills group	A group of *skills*, *competences*, or *capabilities*. Skills groups are of two types: those applied according to parameters that are specific to each individual (research) project e.g. 'capture and process', and those according to parameters applied consistently across (research) projects, e.g. 'govern and assess'.
Skill	The application of a *competence* or *capability* to a specific context, e.g. deployment of a service. 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: As a [*professional group*] needing to [*deliver a capability*] and as a [*service - role*] I should have [*expertise level*] in [*competence*]" 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: "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] and as a [repository operator] I should [comprehend] [repository/database evaluation and options appraisal]".