



Digital**Preservation**Coalition

# FAIR Forever?

## Long Term Data Preservation Roles and Responsibilities, Final Report

**February 2021 (V.7)**

Dr Amy Currie and Dr William Kilbride



EOSC FAIR Forever has received funding from the European Union under the EOSC Secretariat project. EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-05-2018-2019, grant Agreement number 831644. Europeana is an initiative of the European Union, financed by the European Union's Connecting Europe Facility and European Union Member States (<https://pro.europeana.eu/> and <https://www.europeana.eu>)

## Contents

Executive Summary .....	4
Recommendations .....	6
1. Introduction .....	8
1.1. Overview of the FAIR Forever Study .....	8
1.2. Main Activities.....	9
1.3. Outline of Document .....	9
2. Design and Methods .....	10
2.1. Desk-Based Assessment.....	10
2.2. Semi-Structured Interviews .....	11
2.3. Use Cases and Candidate Models .....	12
Focus Group Participants .....	14
User Groups and Persona Types .....	17
2.4. Other Inputs.....	19
2.5. Project Execution and Constraints.....	19
3. Findings .....	20
3.1. Preservation and FAIR Principles .....	20
3.2. The Current State of Preservation in EOSC.....	22
Digital Preservation is not explicit in the EOSC Vision .....	22
Preservation roles, responsibilities, and accountabilities are opaque .....	23
Risks to data, reputation, and sustainability.....	25
3.3. Articulation and Assessment of Use Cases .....	28
Use Case Group Activity.....	28
Articulation and Assessment of Use Cases for Candidate Services in EOSC.....	34
4. Conclusion, Recommendations, and Roadmap .....	39
4.1. Long-term Preservation Roles and Responsibilities.....	39
Foundations for a framework of preservation roles and responsibilities .....	41
4.2. Strategies, Activities and Enhancements to the Preservation of Research Data .....	45
Recommendations for technical challenges and prerequisites.....	45
Recommendations for social, financial, legal, educational, cultural challenges and prerequisites .....	47
4.3. Summary Recommendations and Roadmap .....	48
Strategies, activities and possible enhancements to the preservation of research data.....	48
Emergent roles and responsibilities for the preservation of digital research data .....	49
Implication and intersection of FAIR principles on long term preservation of research data and data management planning.....	49
Costing parameters and funding models for the preservation of research data .....	49
Training and dissemination needs to promote and sustain a culture of data preservation .....	49

Potential benefits and risks of EOSC data preservation services.....	49
Appendices.....	52
Appendix A: Interview Guide .....	52
Appendix B: Interview Participants.....	54
Appendix C: Focus Group Participants.....	54
Appendix D: Focus Group Survey Responses.....	55

## Executive Summary

The Digital Preservation Coalition (DPC) was commissioned by the EOSCsecretariat.eu project at the request of the Sustainability Working Group (SWG) of the European Open Science Cloud (EOSC) to assess and make recommendations concerning digital preservation capacity within the EOSC community, and to make recommendations about the role of digital preservation within the emerging EOSC vision. This report (Deliverable 2) is a summary of the key findings from this research conducted for the FAIR Forever study with recommendations.

The study has allowed the DPC to make a series of statements about digital preservation in the context of EOSC through desk-based assessment, stakeholder interviews, and interactions with the broader digital preservation community through focus groups. An interim statement of findings was supplied to the EOSC in September 2020, and an expanded, updated interim statement was supplied in October. The SWG has been involved and informed throughout the research process through a series of presentations and meetings, and this final report was previewed by the SWG as a presentation before submission.

The study has found the EOSC's emerging vision, articulated most fully in the drafts of the Strategic Research and Innovation Agenda (SRIA), lacks clarity about digital preservation.<sup>1</sup> Additionally, while the SRIA provides an encompassing definition of data to include all digital outputs of research, the study found a strong (perhaps rhetorical) tendency to focus on the preservation of data, with the term often used inconsistently among EOSC stakeholders.<sup>2</sup> All EOSC stakeholders should be aware of and recognize the width of the preservation challenge implied by a broad, maximal definition of data; data sets, publications, correspondence, software, applications, libraries, code, micro-service dependencies, execution environments and operating systems will all need to be preserved or recreated depending on scientific use cases. Software in particular is a significant digital preservation challenge as the certification of code repositories and the validation of emulation or virtualization services are still immature.

Undoubted strengths within the vision, including a commitment to persistent identifiers, data management planning, robust data storage and repository certification, provide a necessary but insufficient basis to secure digital assets in the long term. Roles, responsibilities and accountabilities are opaque and the path-dependency of digital preservation is not fully understood. Risks to reputation and data for EOSC arise from the technical complexity and uncertain accountabilities in the EOSC vision. Furthermore, there remain additional challenges tied to existing and available resources such as the lack of clear funding and costing models for digital preservation and specific skills and training for the various actors in and across preservation activities. This overall reading of the EOSC vision against the classification scheme offered in the DPC Global List of Digitally Endangered Species would suggest that data in the EOSC ecosystem is 'critically endangered'.<sup>3</sup>

Key findings from the study's interviews, interactions, and focus groups reinforce the need for elucidation of roles and responsibilities, and recommended solutions to mitigate the risks. In particular, participants in this research have emphasized the need to clarify accountabilities that are

---

<sup>1</sup> Based on the study's reviewed drafts of the EOSC SRIA, which included EOSC SRIA Version 0.8, October 2020: <https://www.eoscsecretariat.eu/sites/default/files/eosc-sria-v08.pdf>; and EOSC SRIA Version 0.9, November 2020 (provided by the SWG). Since the publication of this report, there has been the release of the final SRIA of the EOSC Version 0.9 16 November 2020: <https://www.eoscsecretariat.eu/sites/default/files/eosc-sria-v08.pdf>.

<sup>2</sup> In EOSC Version 0.9 16 November 2020, data is defined in the Glossary as "An encompassing term used in the EOSC context for all digital outputs of research including datasets, metadata, publications and software code.": <https://www.eoscsecretariat.eu/sites/default/files/eosc-sria-v08.pdf>.

<sup>3</sup> DPC The BitList 2020: The Global List of Digitally Endangered Species, November 2020: <http://doi.org/10.7207/DPCBitList20-01>.

implicit but never activated within data management plans (DMPs). Based on the cumulative findings, the study makes nineteen recommendations for action tabulated with respect to owners of the recommendation to offer particular responsibilities and accountabilities corresponding to each, and delineates five candidate model services, with respective strengths and weaknesses, which would satisfy key use cases identified in the course of the research.

EOSC is an emergent entity currently in transition. Consequently, the recommendations and findings offered here will not be current for long. Relatively new services and tools, notably those being developed through the Archiving and Preservation for Research Environments (ARCHIVER) project, have made progress in the lifetime of this research and hold promise even if they have not yet reached maturity.<sup>4</sup>

December 2020

---

<sup>4</sup> ARCHIVER project: <https://www.archiver-project.eu>.

## Recommendations

The initial study design, approved for funding in August 2020, proposed seven areas for recommendations. These are reported here and repeated later in the report where they are tabulated with respect to owners of the recommendation. Each is assigned one of three priority levels based on its importance and urgency:

- Urgent - of highest importance and urgency (for immediate action)
- High - of high importance and urgency (as soon as possible)
- Medium - of importance and some urgency (when feasible)

### Strategies, activities and possible enhancements to the preservation of research data

1. For the EOSC Secretariat: of urgent priority, establish a working party or task group, reporting directly to the EOSC Association Board with respect to digital preservation.
2. For the EOSC Secretariat: of high priority, formalize terms of reference and host an initial meeting of a digital preservation task group to establish an iterative work plan.
3. For the EOSC Secretariat: of medium priority, establish an operational basis for partnership to deliver the candidate model services proposed in this report:
  - A legacy code or software preservation service
  - A mechanism to ensure accountability and implementation of preservation in data management plans
  - A business case factory or service for preservation cost modelling
  - A programme to support researchers with preservation at the point of creation
  - A mechanism for digital preservation policy across institutions within EOSC
4. For Research Repositories: of urgent priority, adapt workplans to include quality improvement mechanisms where these do not already exist, including DPC Rapid Assessment Model, establishing thereby a strategic framework to achieve baseline certification for primary preservation services, or identifying preservation pathways for data.

### Emergent roles and responsibilities for the preservation of digital research data

5. For the EOSC Association Board: of urgent priority, designate a Senior Digital Preservation Rapporteur on behalf of the Board to directly communicate and liaison with a Digital Preservation Task Group, to monitor and oversee EOSC's responses to digital preservation risks.
6. For Funders: of urgent priority, articulate to all grant holders the clear view that adherence to FAIR principles requires data to be monitored and preservation issues to be addressed over the entire life of a project not simply at the point of project completion
7. For Funders: of high priority, audit preservation pathways for all research outputs to identify critically endangered content.
8. For Funders: of high priority, initiate a process to establish accountabilities and obligations with respect to implementation of data management plans.
9. For Funders: of medium priority, establish mechanisms to engage expert communities of practice in the validation of data management plans.
10. For Research Repositories: of medium priority, provide strategic framework for audit of data management plans.

### Implication and intersection of FAIR principles on long term preservation of research data and data management planning

11. For EOSC Secretariat: of medium priority, establish mechanism to align EOSC implementation and interpretation of FAIR with the path-dependent and continuous quality improvement cycles of digital preservation.
12. For the Digital Preservation Community: of urgent priority, provide a place for EOSC to share lessons and articulate emerging requirements outwith the research data 'bubble'.

### **Costing parameters and funding models for the preservation of research data**

13. For EOSC Secretariat: of medium priority, establish and verify business models for preservation services.
14. For Research Repositories: of medium priority, identify costs of action versus inaction with respect to high value, critically endangered content.
15. For Funders: of medium priority, establish and identify costs of action versus inaction with respect to high value, critically endangered content.

### **Training and dissemination needs to promote and sustain a culture of data preservation**

16. For EOSC Secretariat: of high priority, establish an ongoing basis for partnership in the digital preservation community, expanding beyond the research data community.
17. For Funders: of medium priority, establish more sustained digital preservation training for researchers and repository managers.

### **Potential benefits and risks of EOSC data preservation services**

18. For EOSC Association Board: of high priority, obtain strategic control of digital preservation risks to EOSC.
19. For EOSC Association Board: of medium priority, establish a strategic trajectory for management of digital preservation risks, embedding these within reviews and enhancements.

# 1. Introduction

## 1.1. Overview of the FAIR Forever Study

Digital preservation has emerged in recent years as a fast-moving and growing community of practice that is of ubiquitous relevance, but in which capability is unevenly distributed. Capacity erodes quickly, establishing a need for ongoing reconnaissance to ensure skills, technology, and policy remain fit for changing purpose.

Research disciplines have made a substantial and early contribution to the field, not least through investments of the European Commission. Digital preservation in the research community has a close alignment to the FAIR principles and is delivered, albeit unevenly, through a complex specialist infrastructure comprising not simply technology but also capacity of staff and 'know why' of policy.

The European Open Science Cloud (EOSC) initiative has extensively worked to promote and enable access to Open Science data with the stated aim of ensuring that researchers can maximize the value of their research processes and shared large-scale Research Infrastructures (RIs). EOSC has a compelling and ambitious prospectus: to bring scientists and their audiences together; to federate existing infrastructures; to augment these infrastructures with new added-value services; and to revolutionize research and how scientific knowledge is created in all disciplines, in all geographies. The envisioned EOSC system will offer a Web of FAIR Digital Objects and Related Services for Science through three layers:

1. A federating core (EOSC-Core)
2. The federation of existing and planned research data infrastructures; and
3. A service layer comprising common services and thematic services (EOSC-Exchange).

There is the need for ongoing reconnaissance and assessment of digital preservation within EOSC because digital preservation involves the continuous interaction of policy, technology and capacity. New standards develop while older standards become obsolete, and policy objectives become fossilized and redundant. In addition to technological advances, there are technological challenges with managing and preserving large quantities of open data coupled with high-performance storage and computing resources. While the need for ongoing assessment and renewal of technical infrastructure is apparent, the need to assess and renew social and organizational infrastructure can frequently be overlooked.

The FAIR Forever study, which has received funding from the European Union under the EOSC Secretariat project, was developed by the Digital Preservation Coalition (DPC) in response to an invitation by the EOSC Sustainability Working Group (SWG) to assess current strengths, weaknesses, opportunities and threats to the preservation of research data across EOSC, and determine the feasibility of establishing shared approaches, workflows and services that would benefit EOSC stakeholders.

The DPC, established in 2002, is a not-for-profit membership organization with 111 members around the world that aims to ensure a secure digital legacy. The DPC enables its members to deliver resilient long-term access to digital content and services, helping them to derive enduring value from digital assets and raising awareness of the strategic, cultural and technological challenges they face. The DPC is vendor and technology neutral, offering clear and impartial advice on the maturity and suitability of different digital preservation tools, products and techniques. The DPC's large network of like-minded organizations is supported by established social and organizational infrastructure. With existing and effective communication channels in place, as well as a strong web and social media presence, the organization draws from the expertise of its diverse membership and in the ongoing custodianship and promotion of study outcomes.



## 1.2. Main Activities

The main activities of the FAIR Forever study are listed below:

- Background research and data collection, undertaking a desk-based assessment of the current state of digital preservation in EOSC governance documentation and EOSC projects' outputs and plans
- Interviews and interactions with representatives of the EOSC related European Strategy Forum on Research Infrastructures (ESFRI) Cluster and Regional projects to discuss and assess their current requirements and capabilities in digital preservation
- Interactions with DPC membership to establish comparative capabilities and provisional roadmap of service delivery and emerging requirements
- Presentation of an interim report describing progress, forward plans, with an interim statement of emerging findings for consideration by the Sustainability Working Group
- Focus groups comprised of a purposive sample of representatives from DPC Member organizations whose work involves the digital preservation and management of research data at Research Performing Organizations (RPOs) to identify possible use case scenarios
- Articulation and assessment of use cases to propose candidate model services for the preservation of research data in the context of EOSC including SWOT analysis
- Development and review of recommendations based on analysis of cumulative findings and stakeholder feedback
- Presentation of a draft of the final report to the Sustainability Working Group for review and feedback
- Presentation and dissemination of the final report to stakeholders

## 1.3. Outline of Document

The next section of this report, Section 2, provides an overview of the study's design and methods with comments on the study execution and constraints. Section 3 details key findings from the study relating to preservation and the FAIR Principles (3.1), the current state of preservation in EOSC (3.2), and the articulation and assessment of use cases for candidate services for the preservation of research data in EOSC (3.3). The fourth and final section provides a conclusion with recommendations. It offers recommendations regarding long-term data preservation roles and responsibilities (4.1) and for strategies, activities and possible enhancements to the preservation of research data (4.2), concluding with a summary of recommendations and roadmap outlining the nineteen recommendations for action under seven headings (4.3).

## 2. Design and Methods

### 2.1. Desk-Based Assessment

The first stage of research in this study involved preliminary research and a desk-based assessment of the current state of digital preservation from EOSC governance documentation and EOSC projects' provided by the Sustainability Working Group. These documents are listed in the table below:

Title	Source	Date
Digital Preservation Handbook	DPC	2015
4C Roadmap	4C	Feb-15
Funding research data management and related infrastructures	Science Europe	May-16
EOSC Pilot Demonstrator	EOSC	2017
Presenting a Framework for Discipline-specific Research Data Management	Science Europe	Jan-18
Long-term Sustainability of Research Infrastructures: Science Europe's Offer to Contribute to Ongoing Efforts	Science Europe	Mar-18
EU Commission Recommendations on access to and preservation of scientific information	European Commission	Apr-18
Practical Guide to International Alignment of Record Data Management	Science Europe	Nov-18
ARCHIVER project D2.1 State of the Art, Community Requirements and OMC Results Report	ARCHIVER	Nov-19
ESFRI cluster project paper	EOSC ESFRI	Dec-19
Solutions for a Sustainable EOSC: A tinman report from the EOSC Sustainability Working Group  Later versions of this document: <ul style="list-style-type: none"> <li>• Solutions for a Sustainable EOSC: An Iron Lady report from the EOSC Sustainability Working Group (October 2020)</li> <li>• Solutions for a Sustainable EOSC: A FAIR Lady (olim Iron Lady) report from the EOSC Sustainability Working Group (November 2020)</li> </ul>	EOSC SWG	Dec-19
Digital Preservation Procurement Toolkit	DPC	2020
Implementing Research Data Management Policies Across Europe: Experiences from Science Europe Member Organisations	Science Europe	Jan-20

Draft Proposal European Partnership under Horizon Europe European Open Science Cloud (EOSC) Partnership	EOSC Executive Board	May-20
Joint paper from regional projects	EOSC	May-20
EOSC Core Operational Costs Study	EOSC SWG	Jul-20
Open Consultation for the Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC)  Later versions of this document: <ul style="list-style-type: none"> <li>• Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC) Version 0.8, 18 October 2020</li> <li>• Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC) Version 0.9, 16 November 2020</li> </ul>	EOSC Executive Board	Jul-20
EOSC Landscape of EOSC-Related Infrastructures and Initiatives Report from the EOSC Executive Board Working Group Landscape	EOSC Landscape WG	Aug-20
ExPaNDS and PaNOSC position paper on EOSC A communication following the SRIA consultation	ExPaNDS and PaNOSC	Sep-20
EOSC-hub briefing paper Provision of CrossBorder Services	EOSC Hub	Sep-20
EOSC Skills and Training Target Hierarchy	EOSC Skills & Training WG	Sep-20

## 2.2. Semi-Structured Interviews

Following the desk-based assessment, interviews with representatives from ESFRI Cluster and Regional projects were arranged and conducted from the 6th October to 2nd November 2020.

DPC chose a semi-structured approach for the interviews to elicit contributions and reflections from the participants as key EOSC stakeholders. The interviews were structured around a consistent set of themes drawing on the interim statement, but with mostly open-ended questions that encouraged flexibility and allowed other relevant ideas around the topics to emerge during discussions.

An initial roster of interview candidates was assessed by the SWG to ensure a sufficiently diverse and representative sample, and to avoid the implicit biases towards digital preservation within the DPC's existing networks.

Interview guides were sent to participants in advance of the interviews to provide an overview of the study, the scope and structure of the interview, and an outline of the main areas and questions to be covered in case they wanted to give written answers or verify factual matters (see Appendix A).

The interviews were one hour in length, conducted in English, and used the same web-based video platform (Zoom), with two DPC staff interviewers and either one or two interviewees attending. One of the DPC staff took the role as designated notetaker. The interviews were recorded with permission and reviewed to confirm and supplement the notes taken.

In total, seven interviews were conducted with twelve individuals from the following stakeholder groups: NI4OS-Europe, ESCAPE, SSHOC, PaNOSC, ENVRI-FAIR, EOSC-Life. These interviews occurred from the 6th October to 2nd November 2020 (see Appendix B).<sup>5</sup>

### 2.3. Use Cases and Candidate Models

The third stage of research began with follow-up interactions with EOSC stakeholders and DPC Membership to establish comparative capabilities and offer a provisional roadmap of service delivery and emerging requirements. In October 2020, a provisional roadmap with interim statement was presented to stakeholders at the EOSC Governance Symposium, followed by the delivery of an interim report (Deliverable 1) summarising the study's progress, forward plans, and revised interim statement for consideration by the Sustainability Working Group.

This final stage of data collection entailed interactions with a purposive sample of representatives from DPC Member organizations, recruiting and selecting candidates for a focus group to help articulate, assess, and compare and potential use cases for preservation services within EOSC. DPC decided that a broad definition of a use case would be adopted for discussing, constructing, and analysing possible use cases scenarios in the focus group. A use case was broadly understood as a situation in which a resource or service could be used to support the daily work of those managing and preserving research data at Research Performing Organizations.

DPC purposively selected and invited participants meeting the following criteria to gather specialized expert knowledge and practical experiences with—and to some degree biases toward—digital preservation: individuals at DPC Member organizations, who work at a Research Performing Organization (RPO), are based in Europe or the United Kingdom, and their role is in the area of research data management or digital preservation. The rationale for the narrowed selection criteria for participants followed critical case thinking in the sense that if challenges occur for this group, then it is very likely that other, less knowledgeable individuals at less developed organizations will face the same.

The decisions to adopt a broad definition of a use case and purposive sampling and were developed in light of the emerging findings from the interviews, ongoing developments in archiving and preservation services within EOSC, and revised interim statement. The complexity in different user needs and requirements among the range of EOSC stakeholders is well established from EOSC research projects, and touched upon through the discussion with the stakeholders in interviews. Rather than add to or reiterate this complexity, the study narrowed its scope for the identification of shared goals or challenges among a key group of stakeholders working in research data management and preservation activities at RPOs.

The findings from the study's interviews with those at ESFRI Cluster and Regional projects pointed to assumptions about research repositories at RPOs as the more suitable, trustworthy repositories for archiving and preserving research data while, at the same time, indicating a lack of clarity on who exactly is responsible for the preservation of research data at these institutions. Therefore, gathering more information regarding the goals, needs and requirements of known individuals working with the management and preservation of research data at RPOs would better identify existing gaps or solutions in EOSC resources or services.

As will be expanded upon in Section 3.3, the findings emerging from the focus groups contribute to the development of the EOSC by illuminating different kinds of preservation challenges faced by those currently working with research data at RPOs (to identify what preservation challenges other,

---

<sup>5</sup> Representatives from EOSC-Pillar, EOSC-Nordic, and EOSC-Synergy were contacted but were either unavailable or unresponsive to the emailed requests during the interview period of the study. The DPC and SWG welcome and encourage any feedback or contribution to this version of the report.

less knowledgeable EOSC stakeholders might encounter) and sharing knowledge of useful resources or models that to better deal with these challenges (to highlight candidate models for EOSC services or resources). Drawing from the knowledge and experience of this purposive sample provides more concrete guidance on how to address the interim statement's questions concerning:

- Where and how might digital preservation be explicit in the EOSC vision (to be effective)?
- What are the existing roles/responsibilities/accountabilities for digital preservation and research data management at RPOs that are relevant and applicable to EOSC?
- Where and how can the risks to data, reputation, sustainability be addressed or mitigated?

These questions were directly asked to the participants as discussion points.

The interactions followed a semi-structured approach. Each web-based session was one hour and followed the same general structure:

- A brief overview and background of EOSC, FAIR, and the study (15 minutes)
- A group activity (or walk through) discussing potential use case scenarios based on representative personas (actors) within the digital preservation community (30 minutes)
- Reporting key points from activity discussion, and more general discussion touching upon the topics or areas surrounding findings and interim statement (15 minutes)
- Survey questions incorporated throughout to supplement and facilitate discussion

The personas used for the group activity were based on EOSC user categories and types developed by the EOSC Skills and Training Working Group, the DigCurV framework, and recent DPC Labour Market Analysis.<sup>6</sup> The following descriptions of the personas were presented to participants:

- Researcher – for example, a primary investigator, post-doc, student, faculty, or staff at a RPO who creates and manages data during research, and their data may be stored or deposited in the research repository.
- Technician (ICT-specific) – specific to Information and Communications Technology (ICT), for example an information technology infrastructure technician, software designer, developer, or engineer whose daily work primarily involves the design, development or implementation of technologies for managing and preserving the research data in the research repository.
- Practitioner (LIS-specific) – specific to Library and Information Sciences (LIS), for example a data curator or data steward whose daily work primarily involves more of the 'people' aspects of research data creation, management, and preservation, working with researchers as well as putting standards and policies into practice, ensuring compliance and quality assurance.
- Manager (generalist or specific) – for example, a digital preservation manager, head of research data, or systems librarian whose daily work primarily involves the managerial role of staff and services, grant applications and procurement.
- Executive – for example, a director whose daily work involves high-level leadership and liaising with various stakeholders.

---

<sup>6</sup> EOSC Skills and Training Working Group, Identifying Digital Skills Sets for EOSC, August 2020: <https://www.eoscsecretariat.eu/news-opinion/identifying-digital-skill-sets-eosc>; DigCurV Curriculum Framework, 2013: <https://www.DigCurV.gla.ac.uk/index.html>; DPC Labour Market Analysis Summary Report, August 2020 (available upon request).

- Educator/Trainer – for example, a training officer whose daily work primarily involves the creation, development or delivery of training on research data management or digital preservation skills.

The group activity was limited to just the Technician, Practitioner, and Manager personas. These three personas were prioritized because they deal most directly with the management and preservation of data at the RPOs, and tend to overlap the most in terms of job titles, roles, and responsibilities. The function of the different personas corresponding with the technology, policy, and 'people' aspects was not to reinforce the conceptual divisions or ignore overlap but rather to dig deeper into the current needs and requirements most associated with each and ensure different aspects were given equal attention.

The interactions were one hour in length, conducted in English, and used the same web-based video platform (Zoom). The DPC project team wanted to encourage participants to speak freely, so we chose not to record the workshop sessions but instead collected data through designated notetakers, group activity sheets, and responses to an online survey (Mentimeter).

In total, there were fifteen participants and four sessions; a group session with ten participants and three follow up interactive sessions with five participants who were unable to attend the group session. These interactions ran from the 9<sup>th</sup> November to 2<sup>nd</sup> December 2020, and more details about the participants are presented below (see also Appendix C and Appendix D).

### Focus Group Participants

An overview of findings about the focus groups participants is provided here to contextualize and qualify later discussions of the findings from the focus group presented in Section 3.3, specifically as they pertain to the articulation of different use case scenarios.

In total, there were fifteen participants from nine DPC member organizations which also classify as RPOs. As one might expect from the DPC membership, most of the participants identified their RPO as having some existing capacity in digital preservation: 80% were at organizations with basic or higher levels of capacity/readiness (see Figure 1 below).



Figure 1. Levels of Digital Preservation Capabilities

All the participants were from Northwest Europe (7 organizations based in the United Kingdom; 1 in the Netherlands; and 1 in Ireland).

Participants were asked to classify their organizational types (participants could choose more than one option). As shown in Figure 2 below, most selected Higher Education Institution (13) and Library or Archive (11). Two participants work at Research Infrastructures or Clusters, which connect with the EOSC as a portal for federated RIs.

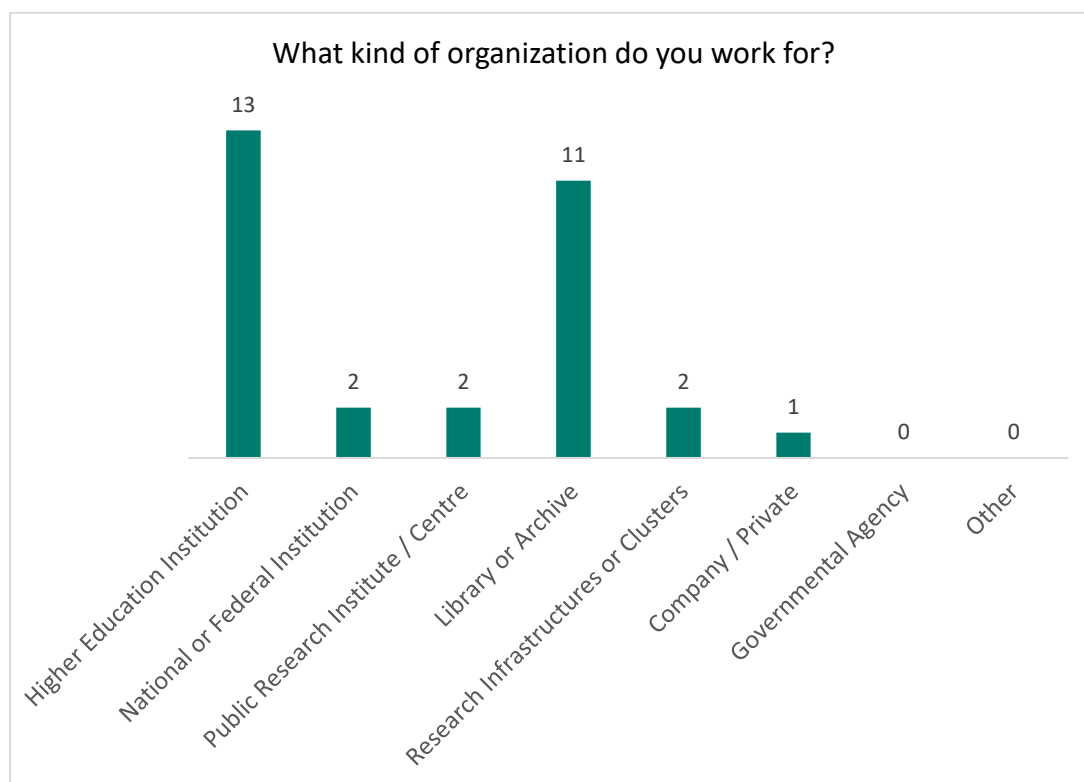


Figure 2. Types of Organizations Represented

There were expected overlaps among the responses for organization types; and indeed nine participants chose more than one type. These included:

- Five participants chose two organization types, all of them being Library or Archive and Higher Education Institution.
- Two participants chose three types. One participant with Library or Archive, Higher Education Institution, and Company/Private. The other with Library or Archive, Higher Education Institution, and Research Infrastructures or Clusters.
- One participant chose four types with Library or Archive, Higher Education Institution, National or Federal Institution, and Public Research Institute / Centre.
- One participant chose five types with Library or Archive, Higher Education Institution, National or Federal Institution, Research Infrastructures or Clusters, and Public Research Institute / Centre.

This demonstrates that digital preservation responsibilities are not limited to one kind of organization; that the majority of digital preservation and research data management work is collaborative; that participants could be employed at more than one organization; and that higher-level positions often involve wearing 'many hats'.

Participants were then asked about their involvement with EOSC projects of initiative (see Figure 3 below).

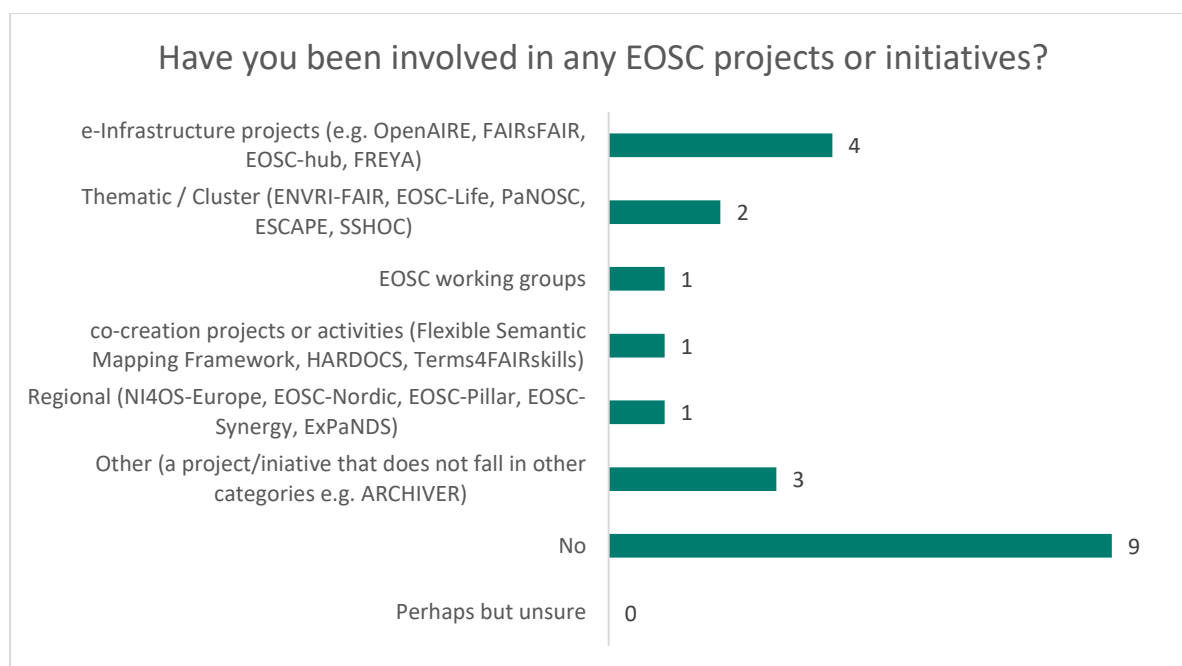


Figure 3. Involvement in EOSC Projects or Initiatives

More than one option could be chosen. However, nine of the fifteen participants (60%) indicated they weren't involved in any current EOSC projects or initiatives. The option 'Perhaps but unsure' was presented in cases where participants may be uncertain, but none selected this option. While this indicates that they are not directly involved in current projects or initiatives, this is not to say that they are not part of the EOSC community of stakeholders. Their participation in the study and active discussion in the sessions indicates that there is an interest and engagement with the EOSC irrespective of direct involvement in projects of initiatives.

Among the six participants who were involved in one or more of EOSC projects or initiatives, those categories of project or initiatives included (12 total):

- Four participants were involved in e-Infrastructure
- Two with Thematic/Cluster
- One with EOSC working groups
- One with Regional
- One with co-creation project or activities
- Three selected Other

In terms of the participants' involvement in more than one category of projects or initiatives, four of the six participants chose two or more:

- Two participants indicated involvement in both e-Infrastructure projects (e.g. OpenAIRE, FAIRsFAIR, EOSC-hub, FREYA) and the Other category.
- Two participants chose three categories.
  - One was involved in EOSC working groups, Regional (NI4OS-Europe, EOSC-Nordic, EOSC-Pillar, EOSC-Synergy, ExPaNDS), and co-creation projects or activities (Flexible Semantic Mapping Framework, HARDOCS, Terms4FAIRskills).
  - One was involved in Infrastructure projects (e.g. OpenAIRE, FAIRsFAIR, EOSC-hub, FREYA); Thematic / Cluster (ENVRI-FAIR, EOSC-Life, PaNOSC, ESCAPE, SSHOC); and selected the Other category.



The research domains and disciplines in which the participants work with data and data creators were not captured in the survey, but it was discussed during conversations about the use cases and will be touched upon in Section 3.3 detailing findings from the use case discussions. It should be noted here that the majority of participants, working at Higher Education Institutions or Library and Archive settings, worked with research data in and across disciplines. There were three participants, however, working with discipline and domain-specific research data (Archaeology and the Arts).

## User Groups and Persona Types

Following the presentation at the start of the session, participants were asked to provide their job titles and identify which EOSC user group and persona they identified with most strongly. There was only one shared job title, Research Data Manager, which three participants listed (illustrated in Figure 4 below).



Figure 4. Word Cloud of Job Titles

It is difficult, if not impossible, to categorize the participants into separate digital preservation or research data management camps based on their job titles alone. However, glancing at the frequency of certain terms within the titles (e.g. digital preservation, preservation, archives, and research data, research, repository) does give some—albeit very limited—indication that in addition to the three participants with 'research data manager' in their title, there are three jobs with 'digital preservation' in the title, indicating there was no dominance of one camp over the other among the participants.

Additionally, participants were asked to choose a user group or persona that they most identified themselves (see Figure 5 below).

## User Groups & Personas

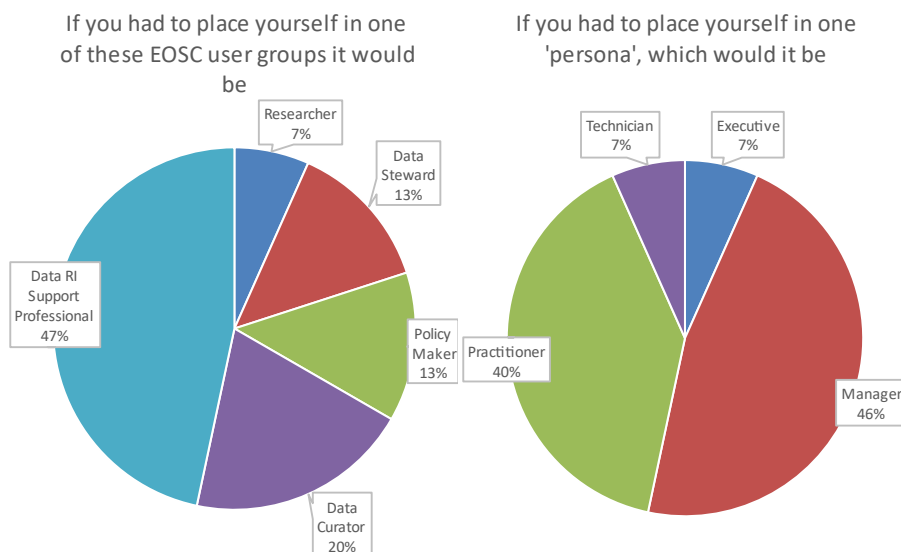


Figure 5. Identified User Groups and Personas from survey responses

It is perhaps surprising that, in the context of the focus group sessions, the majority of participants chose Data RI Support Professional. The desk-based review, as well as the interviews that followed from the first two stages of research, would suggest that Data Steward would have been the preferred term, or perhaps Data Curator which also has a lot of currency within the literature. However, when looking closer at research data itself as the common link between the digital preservation and research data management camps—with the EOSC as a federation of the research data infrastructures—this result demonstrates where the participants see their role as users and providers of data. It is also perhaps a reality that, with 60% of participants having had no involvement in EOSC, they do not readily identify with the terminology of 'stewardship' or 'curation'.

For the identified personas, there was no dominant type. The majority (14) selected either the Practitioner or Manager persona (7 participants for each). Those who identified as Data RI Support Professional in the EOSC user group (7) selected different persona types, suggesting again that their work involves all the various aspects that together contribute to the RI support in the context of EOSC. Although there were seven job titles with 'manager' in the title, only five of those seven chose the Manager persona group. Two participants who identified themselves as Data Stewards also identified themselves as Managers. One participant identified themselves as a Researcher and Practitioner, suggesting their identity as a user of EOSC would primarily be for their research.

Given the evident and acknowledged complexity in and among those in the European research communities relating to the different regions and specific domains, extensively noted and remarked upon by those working to stakeholders and map the larger EOSC Ecosystem<sup>7</sup>, the findings from this small group of fifteen participants drawn from a purposive sample are interesting to explore further. In this way, the findings are not surprising given the known complexity; on the other hand, this was

<sup>7</sup> In the SRIA Version 0.8 (October 2020), EOSC defines the EOSC Ecosystem as "The encompassing set of federated (e-)infrastructures, research infrastructures, stakeholder organizations and projects that contribute to and/or use EOSC." (p. 11). See: <https://www.eoscsecretariat.eu/sites/default/files/eosc-sria-v08.pdf>.

surprising as it was expected that there would be more similarities in the identified groups and types given the narrowed criteria for the purposive sample.

It is worth exploring this further, and Appendix C provides information to encourage other analysis.

## 2.4. Other Inputs

The study also collected and reviewed information from updated versions of EOSC governance documents, resources provided or suggested by participants, and other relevant materials, events and activities including

- EOSC-Pillar D3.1 Summary report of the 'National Initiatives' Survey (July 2020)
- Solutions for a Sustainable EOSC: An Iron Lady report from the EOSC Sustainability Working Group (October 2020)
- Scholarly Infrastructures for Research Software - Draft for consultation (October 2020)
- Version 0.8 of the Strategic Research and Innovation Agenda (SRIA) for EOSC (October 2020)
- EOSC Symposium and follow up conversations with participants (October 2020)
- Version 0.9 of the Strategic Research and Innovation Agenda (SRIA) for EOSC (November 2020)
- Solutions for a Sustainable EOSC: A FAIR Lady (olim Iron Lady) report from the EOSC Sustainability Working Group (November 2020)
- EOSC Enhance Portal release (December 2020)
- ARCHIVER project prototype phase kick-off event (December 2020)

## 2.5. Project Execution and Constraints

Although the activities and deliverables within the study have largely followed the initial specification, the timeline of the study has been significantly impacted by a late start. Initially proposed to start in June the study was not commissioned till mid-August. A simple rectification was to move the key milestones out by three months. As a consequence, milestones are no longer synchronized with key milestones on the roadmap for the Strategic Research and Innovation Agenda. The result for the project team has been tightly compressed deadlines and too little time to digest and elucidate findings, as well as sudden and unexpected requirements for submissions and deadlines which had not been anticipated at the outset. Instead, preliminary finds have been iterated to meet the timelines of the Sustainability Working Group who in turn were responding to a demanding but rigid programme. Follow up interviews and questions were curtailed, and peer review of emerging findings has been scaled back to ensure that deadlines were respected. The project team gratefully acknowledges the generous support of the Sustainability Working Group to this research and also acknowledges the timeline as a constraint on the research.

## 3. Findings

### 3.1. Preservation and FAIR Principles

This section details the findings regarding the assessment of digital preservation as it intersects with FAIR principles in the context of EOSC. It is primarily based on a desk-based analysis of documents which was then refined through interviews and focus group interactions with EOSC stakeholders.

From the outset of this study, it has been apparent that EOSC and the FAIR principles (of Findability, Accessibility, Interoperability, and Reusability) are tightly interdependent. The FAIR principles were part of the foundations of the envisioned EOSC, enabling researchers to perform Open Science and open their research data for sharing. The successful federation of research data infrastructures for EOSC requires the implementation of the FAIR guiding principles so that the data and digital content is discoverable and usable. The EOSC FAIR Working Group, FAIR task groups, and other related initiatives and projects have worked to define and communicate the corresponding FAIR requirements and practices expected for EOSC stakeholders.<sup>8</sup>

There are significant areas where the FAIR principles intersect with preservation, and there are notable examples of how good practice applying FAIR principles also delivers good practice in digital preservation. These include an early focus on persistent identifiers (PIDs), an emphasis on data management planning, and planning for robust storage. As the study progressed from the initial desk-based assessment, further plans for repository audit and certification became apparent. All of these are essential elements of a digital preservation strategy and capability.

In terms of findability and access, the strengths and use of persistent and unambiguous identifiers seems high across EOSC and provides an immediate capability to leverage existing solutions. On the other hand, a deeper dive into the PID landscape reveals this is not a solved challenge. Still, ongoing EOSC projects and initiatives—notably those linked with the EOSC FAIR Working Group—have recognized the challenge and are working to enhance solutions. In the context of FAIR and preservation of research data in and across EOSC infrastructures, the challenge of persistent linking provides a possible entry point for conversations about policy and planning of the wider digital preservation challenges.

The FAIR principles align with an existing culture and expectation of data management planning across EOSC, which overlap with preservation planning. The incorporation of FAIR principles into data policies and guidance can set initial prerequisites for preservation in data management planning—the most obvious prerequisite being the storage of data for any FAIR activity. In this way, FAIR provides an entry point for preservation awareness and early planning among active researchers and an opportunity to guide and assess the creation of, for example, representation information for preservation early in the research data lifecycle.

While there is a general recognition of the value of researchers creating and submitting data management plans (DMPs), and even a sense that responsibilities are made clear through the process, interviewees doubted the accuracy or utility of data management plans. A few commented that DMPs seemed more like a gateway process to funding rather than an actionable, audited vade mecum for data. Interviewees were consistent that there was a lack of accountability or transparency in data management plans—that they were used only occasionally once a project had begun, and offered uncertain support to preservation which was still seen as a matter for the end of the data lifecycle rather than a necessity throughout the lifecycle.

---

<sup>8</sup> For more about the EOSC FAIR Working Group, publications, and outputs, see: <https://www.eoscsecretariat.eu/working-groups/fair-working-group>.

The strengths and limitations of FAIR for creating, maintaining, and preserving metadata also came up in conversations. When asked about FAIR during the interviews, one participant mentioned insufficient ways to automatically preserve metadata for findability. His work on building an analysis platform where search and delivery are shared, with the results being saved back into the system, was limited by the timelines for making data findable and accessible across current infrastructures. He was cognizant of the need for preservation considerations, and potential implications with current approaches, but explained that current investments of time and resources were focused on ensuring the FAIRness of current metadata.

Another interviewee ranked interoperability as the greatest of the FAIR challenges—it was far more difficult to make data interoperable than findable or accessible. He felt achieving interoperability as a more aspirational than an achievable goal when the sharing data across the domains of physics, astronomy, and astrophysics is already challenging. When asked about FAIR interoperability for archiving and preserving, the same participant echoed the constraints of time and resources.

An interviewee working with humanities data also commented that interoperability over time depends on data being FAIR at the outset and that the success will come from re-use as the 'proof of the pudding'. The principle of reusability, framed in this way, is one (but not the only) measurement of the effectiveness of the others. The idea of interoperability over time—diachronic interoperability—may be an area where the present steps to enable cross-disciplinary resource discovery might have long-term implications for preservation. At the same time, this circles back to the idea of aspirational versus achievable goals within the scope of EOSC data, current investments of time and resources, and a degree of hesitation expressed by the participant of wanting to avoid 'FAIR fatigue'—worrying that these efforts without concrete embodiment of the EOSC vision and vision of FAIR data will fall through.

A similar pattern of aspirational versus achievable goals for preservation emerged with respect to the audit and certification of repositories. The certification and other forms of assessment of FAIR services is a welcome step and is articulated within the EOSC vision: but, as currently envisioned, it is insufficient to the scale of the challenge if taken in isolation. Further analysis suggested a misalignment or lack of consideration of the path dependency associated with digital preservation actions required by the application of FAIR principles. One participant at a focus group suggested that, whilst it was useful to certify a FAIR repository, there needed to be a similar audit and certification for data objects, and by extension documentation that identified accountabilities through the data lifecycle. He anticipated that the EOSC interpretation of FAIR meant that repositories would likely be presented with data that they had limited resource to preserve, and little practical chance of saving.

In summary, while the findings show the benefits of EOSC developing FAIR among services and areas where the implementation of the principles might benefit both data management and preservation planning, there seems to be a gap in EOSC's interpretation of the FAIR principles as they pertain to preservation. As currently articulated, EOSC's implementation of FAIR—notably through the work and contributions of FAIRsFAIR—helps researchers and data managers assess their awareness of the requirements for making data FAIR prior to uploading them into a repository, but it does not provide a consistent programme for preserving data.<sup>9</sup> The implementation of FAIR principles in general and specifically concerning digital preservation appears to be ultimately in the hands of those providing and managing the data infrastructures, with some of them commenting on limited time and the current challenges of interoperability.

---

<sup>9</sup> One particularly relevant contribution of FAIRsFAIR for this discussion of developed guidance and services for implementing FAIR is the set FAIR evaluation service: <https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#/>.

### 3.2. The Current State of Preservation in EOSC

This second section summarizes the main findings from the study's assessment of the current state of preservation in EOSC, with each sub-section corresponding to the three key areas identified and refined in the Interim Statements. These were drawn from the desk-based assessment, interviews, and focus group interactions with EOSC stakeholders.

#### Digital Preservation is not explicit in the EOSC Vision.

Whilst there were references to the preservation and archiving of research data throughout the governance documents reviewed for the study's desk-based assessment—notably in the Strategic Research and Innovation Agenda (SRIA) published on 20<sup>th</sup> July 2020—closer readings and analysis of these documents found that digital preservation is only implicit in the EOSC vision.<sup>10</sup>

During the interviews with representatives from the ESFRI Research Infrastructures Cluster and Regional projects, several of them corroborated the advantage of making digital preservation more explicit in the EOSC vision. They confirmed that researchers in every discipline see the need for digital preservation, and all agreed that digital preservation involves more than the long-term storage of data. They further commented that long-term open data archives and preservation services are required to enable sustainable access to data and thus essential to a sustainable EOSC.

Additionally, the desk-based assessment found a strong (perhaps rhetorical) tendency to privilege data, a term which is used inconsistently. EOSC stakeholders are right to be concerned with the preservation of data but need to recognize the width of that challenge implied by this maximal definition. Data, publications, correspondence, software, applications, libraries, code, micro-service dependencies, execution environments and operating systems will all also need to be preserved or recreated depending on scientific use cases. During the interviews, two interviewees in particular detailed the need for software preservation, expanding on how this preservation is arguably more impactful. One interviewee, who works on a data stewardship work package, felt strongly that support for preservation needs to extend for software for reproducibility of scientific findings.<sup>11</sup> Software is a significant digital preservation challenge as the certification of code repositories and the validation of emulation or virtualization services are still immature.

In general, most agreed that there is an overall need for more explicit digital preservation policy and strategy. There is an acknowledgement within the EOSC vision that preservation is important and a consensus that it is a core requirement in every discipline: but specific preservation functions remain obscured and miscommunicated. When asked how digital preservation might be made more explicit in the EOSC vision, one interviewee suggested that an articulation of the main digital preservation objectives and challenges could help guide and assess research infrastructure requirements and capabilities. Another participant added that she felt it was critical for them (as representatives of stakeholders groups and researchers) to be involved with or aware of how EOSC will establish requirements for preservation within the existing policies and frameworks.

There was some confusion and misperception by stakeholders regarding EOSC providing or funding end-to-end digital preservation solutions.<sup>12</sup> When asked about the present view taken by EOSC

<sup>10</sup> The SRIA Version 0.8 (October 2020) added a definition and brief discussion of the preservation of data in EOSC. See: <https://www.eoscsecretariat.eu/sites/default/files/eosc-sria-v08.pdf>.

<sup>11</sup> Ibid. The SRIA Version 0.8 also includes a discussion of different data types in EOSC, notably a section on the preservation of software.

<sup>12</sup> For example, in 'Turning Open Science and Open Innovation into reality: ICDI Position paper on EOSC Partnership Strategic Research and Innovation Agenda' (September 2020) it states that "EOSC should cover the full lifecycle from data production/ discovery to curation, long-term preservation, archiving, and reuse" (p. 7). See: <https://doi.org/10.5281/zenodo.4062291>.

Governance Board that the EOSC community is ultimately responsible for the preservation of research data, the interview participants largely accepted this responsibility but noted that within that broad community there are different views of who should be responsible.

In summary, clearer articulation of data in digital preservation within EOSC's strategic mission, along with an effort to spell out objectives, challenges, and implications for the preservation of research data will help strategic alignment in and across EOSC infrastructures.

### Preservation roles, responsibilities, and accountabilities are opaque

The most significant findings from the desk-based assessment and interviews concerned preservation roles and responsibilities that are that unclear and accountabilities that are uncertain. These findings, summarized below, lead to a significant number of questions about digital preservation capability in EOSC.

The general sense, derived from listening to and reading the notes from the interviews, was that researchers are solely responsible for data creation and management until the data is stored or transferred to institutional, disciplinary or national repositories. Interviewees from the ESFRI Research Infrastructures Cluster and Regional projects discussed the limitations of those projects for ensuring the quality of research data and compliance of data management plans. One explained that cluster projects could provide high-level guidance, but data responsibility sits with the partners and infrastructures that comprise the larger ESFRI. Cluster sustainability is not guaranteed as there exists no specific infrastructure planned for the long-term (though some planning is anticipated to be done). As one participant put it, even the guarantees for infrastructures are problematic—that 'data buckets are not preservation functions, and long-term preservation needs domain knowledge'—because domain knowledge cannot be recreated afterwards.

The concept of data stewardship is presented throughout EOSC documents, often as an ambassadorial role between the researcher and other institutional department and staff such as computing services, institutional repositories, libraries or archives. The EOSC vision anticipates that there will be new professional opportunities created through the Web of FAIR Data and Services.<sup>13</sup> Those working in EOSC FAIR initiatives and projects see FAIR data stewardship as a critical way to support FAIR data interoperability.<sup>14</sup> Yet it is unclear if such stewardship roles are also professionally responsible for long-term preservation and archiving of research data in EOSC, as the title implies. Moreover, if they are assumed to hold this responsibility for long-term preservation, little is known about how it will be supported.

Consequently, there is a lack of clarity on preservation responsibility in EOSC or the specific preservation skills and training needed for staff to advise and guide data creators effectively. One participant reflected on her experiences with researchers within her country, finding that

---

Another example comes from the 'ExPaNDS and PaNOSC position paper on EOSC: A communication following the SRIA consultation' (September 2020) which states "The data produced by our instruments at PaN facilities are a valuable resource for long term reuse. However, our data policies typically guarantee only a 5-10 year preservation period. After that, it would be valuable if the EOSC would take over the responsibility for finding facilities/services for storage and curation of FAIR data to keep it available for further long-term use." (p. 2). See: <https://expands.eu/2020/09/21/expands-and-panosc-position-paper-on-eosc/>

<sup>13</sup> For instance, the SRIA Version 0.8 notes "When open science becomes the 'new normal', scientists will extend their requirements accordingly, and new roles and responsibilities will have to be created (e.g. data scientists, data stewards, etc.)" (p. 47), October 2020: <https://www.eoscsecretariat.eu/sites/default/files/eosc-sria-v08.pdf>

<sup>14</sup> For example, the FAIRsFAIR project document 'D3.4 Recommendations on practice to support FAIR data principles' (2020) outlines specific recommendations aimed primarily at research communities and research support personnel including data stewards and research software engineers. See: <https://doi.org/10.5281/zenodo.3924132>

researchers in smaller social sciences and humanities institutes are very ready to share data but are unsure how to find and navigate the best strategies for preserving it. She also found that, in general, researchers had a greater preference for giving their data to institutional repositories because they believe it could ensure better data preservation as there is staff with a job to curate and control the data and, should anything go awry, will be held accountable.

Furthermore, there are expectations of libraries and archives in institutional repositories to take on digital preservation. Conversations with the interviewees and later interactions with those in the digital preservation community highlighted that the problems arise not in the unwillingness of those in institutional repositories to take this challenge on, but rather the staffing and resource limitations that are implied. There is a need for digital preservation skills development and training for staff. For example, one participant could think of only a few examples of staff able to interact confidently with the data creators and thus provide confidence about their preservation capability. Her experience found it difficult to get University staff involved and skilled sufficiently—while there is the potential for professional opportunities to be created through the Web of FAIR Data and Services the lack of a recognized career structure remains a profound challenge.

Conversations with participants raised a significant concern to whether capability exists to deliver on commitments, even when these are clearly articulated. As one participant commented, the lack of clarity hampered the development of recognized roles and career structures which ought to have emerged by now. Another commented that preservation is—or more accurately should be recognized as—a 'real professional' job which makes a distinct and essential contribution to the research data lifecycle. Yet, it is hard to show how this is represented in career structures. Several interviewees noted that a coordinated and central approach to the provision of preservation skills should be a genuine priority for EOSC.

Digital preservation is path dependent and involves the continuous interaction of policy, technology and resources. When discussing current capabilities for preservation in the ESFRI Cluster and Regional projects, interviewees stressed the knowledge and skills necessary for utilizing technologies in particular research domains such as astronomy and astrophysics. The conversations with interviewees on organizational capacity and staff capabilities touched upon the ongoing development of knowledge, skills, and resources necessary for effective preservation of various kinds of data in and across the research infrastructures. Participants felt that it is very difficult for the EOSC community to take on the staffing and resourcing of digital preservation, especially when coordinating services across Europe.

The interviewees felt organizational viability across the EOSC community varied in terms of governance, organizational structure, staffing and resourcing of digital preservation activities. One said that digital preservation training for their librarians at their university was needed. Another from an ESFRI Research Infrastructures Cluster project said preservation skills are not well represented in the overall infrastructure, and more spend is given to equipment and instruments than people committed to taking care of the data afterwards. She added that in the more matured clusters and science communities, there are projects led by domain experts who are familiar and knowledgeable with digital preservation policy and strategies. For example, in the context of EPOCH (European Research Network of Excellence in Open Cultural Heritage), the work of national component repositories where the expertise and capacity exists.<sup>15</sup> However, the participants within these groups added that there are still gaps in skills and facilities to enable preservation.

The existing knowledge, strengths, and skills within the EOSC community were also discussed. EOSC has a substantial community around it and relatively good infrastructures of communication. While it is agreed that collaboration is key for the success of the federation of infrastructures in EOSC,

---

<sup>15</sup> The European Research Network of Excellence in Open Cultural Heritage (EPOCH) project webpage: <https://www.brighton.ac.uk/csirus/what-we-do/research-projects/epoch.aspx>.



collaboration is complicated when dealing at the scale of EOSC. Some of the issues in scoping preservation within EOSC arise because the services are ordered around community needs. This proximity to the community comes at the cost of complexity but is a very sound investment for data creators and users, creating much better preservation outcomes. Service providers like the ADS (Archaeology Data Service) and DANS (Data Archiving and Networked Services) have demonstrated that proximity and accountability to the research community has a significant benefit to sustainability.<sup>16</sup> In this way, EOSC should try to retain this community focus, establishing a framework of subsidiarity which leverages the latent economies of scale with granular capabilities. At the same time, EOSC faces a challenge of making sure decisions are made at the right level, with accountability and transparency to optimize shared outcomes.

Communication across EOSC is challenging, and this creates a further preservation challenge. Participants felt engaged with and felt part of the larger EOSC research data community but that there is room for more engagement with the wider digital preservation community. Participants acknowledged collaborative development and implementation of the EOSC was complicated given its scale. One commented that genuine goodwill brings together a lot of key people and a lot of people can see the value to the EOSC vision—no one doubts that collaboration is the key to a lot of success and development for preservation, but politics are not properly aligned yet, especially in respect to ministries and Member States. While there was disagreement on the level of detail needed for guidance at the EOSC level, all felt that the role of EOSC as guiders—not managers or dictators—was essential for successful implementation at the community level.

Significant comments were made about the unfulfilled promise of data management plans as far as accountabilities were concerned. There is more or less universal recognition of the importance of data management plans and recognition that these can be used to raise preservation questions at the right point in project lifecycles—even before any data has been created. But interviewees could not identify any accountabilities within data management planning, nor was it clear which part of the EOSC ecosystem was mandated to audit their implementation. This is an obvious example of how accountabilities and reporting could make a dramatic improvement. One participant at the focus group described data management plans as the simplest way to turn 'FAIR data into FAIR data lifecycles'.

### Risks to data, reputation, and sustainability

As stated in the SRIA and SWG FAIR Lady (olim Iron Lady) report, Horizon 2020 (H2020) is the financial instrument implementing the core of actions around which the EOSC implementation roadmap unfolds. H2020 has supported the development of services for Open Science, and the European Commission supporting the implementation of the EOSC as a federated model on this basis. Interview participants were aware of plans for the envisaged continuation of funding EOSC activities beyond 2020 through Horizon Europe.<sup>17</sup>

Within the context of the EOSC initiative, interviewees were also well aware of how their current cluster and regional projects were supported through EOSC funding streams, but varied in their level of knowledge of proposed funding models for EOSC.<sup>18</sup> Several commented on the problems with

---

<sup>16</sup> Archaeology Data Service (ADS): <https://archaeologydataservice.ac.uk/>; Data Archiving and Networked Services (DANS): <https://dans.knaw.nl/en>.

<sup>17</sup> Additionally, there is Digital Europe and other programmes, tools, institutions, businesses or funding bodies contributing to funding EOSC both directly and indirectly through investment in national infrastructures. SRIA Version 0.8 (October 2020): <https://www.eoscsecretariat.eu/sites/default/files/eosc-sria-v08.pdf>.

<sup>18</sup> As explained in the SWG FAIR Lady (olim Iron Lady) report, two families of funding models have been discussed, with both co-existing and potentially applied to different sides of the platform or targeting different clusters of roles and players in order to sustain EOSC. These are: 1. transaction-based models and 2. patronage/membership-based 'learning' funding.

budgeting for the long term with project-based funding. When asked about current funding and funding models for long-term preservation, interviewees explained the project-based funding of their current projects but none provided (or were aware of) any examples specific to the preservation of research data, nor were any preservation models or tools brought up.

The research benefit to preserving data was evident to all the participants across disciplines and regions, especially when discussed in the context of making research data FAIR. It was obvious to them all that the uses to which the FAIR digital material is put (for various research activities) represents a benefit to the users (as researchers). Their concern was how to address other benefits of preservation and assess costs in order to secure the resources required—of time, money, people—to create and maintain a sustainable digital preservation programme. In other words, the central issue was one of measuring benefits and calculating costs for business plans—in both qualitative and quantitative terms.

A participant commented that most of the partners in one of the ESFRI Research Infrastructures Cluster projects have no budget for preservation and questioned whether it is even possible to 'make a guess' on the costs. Others agreed that parameters and actual measures were lacking; several interviewees felt that sustainability is not guaranteed. Although there are digital models and tools such as the Digital Heritage Network Digital Sustainability Cost Model (2017), 4C Curation Costs Exchange Roadmap (2015), Curation Costs Exchange (CCEX) community platform (2014), AVP Costs of Inaction Calculator (2014), or KRDS Benefits Analysis Framework and Toolkit (2011), none were mentioned.<sup>19</sup> These resources, which are helpful for starting out with higher level benefits analysis, curations costs, and cost models, tackle some—but not all—of the qualitative and quantitative aspects of digital preservation. Cost modelling remains a challenging activity given that digital preservation will always be a moving target tied to changing technologies and evolving requirements and, as supported by the interview discussions, challenging to separate digital preservation costs from other business costs. For this reason, the findings suggest an ongoing need to share and develop digital preservation models in the context of EOSC to guide stakeholders and encourage transparency in the costs of digital sustainability, which will be addressed again later in this report's recommendations.

One area where the costs of digital sustainability appeared markedly unclear was that of personnel and staffing costs. Participants found the costing of preservation in terms of personnel and staffing costs difficult to calculate or estimate, including one interviewee who was currently working on a project on sustainability. Nevertheless, given the expressed importance of knowledge, skills, and expertise to developing and implementing a preservation programme, participants agreed that there should be deeper investigations on how to calculate the costs of preservation roles and responsibilities so that they can be included and aligned with EOSC funding programmes. For this reason, which shall be expanded upon later in the report's recommendations, further investigations into preservation roles and responsibilities and their associated costs (e.g. case studies, labour market data analysis) are necessary to establish accurate costing parameters for preservation programmes and services provided in EOSC.

At the fore of the conversations surrounding who should fund preservation was the importance of the European Commission and national infrastructures to provide funding streams. The attention to national infrastructures emerged at multiple points across the interviews. One commented that

---

<sup>19</sup> Network Digitaal Erfgoed (Digital Heritage Network) Digital Sustainability Cost Model report (January 2017): [https://www.netwerkdigitaalervoed.nl/wp-content/uploads/2018/03/Factsheet\\_Cost\\_Digital\\_Preservation\\_sum\\_DEF.pdf](https://www.netwerkdigitaalervoed.nl/wp-content/uploads/2018/03/Factsheet_Cost_Digital_Preservation_sum_DEF.pdf); Collaboration to Clarify the Costs of Curation (4C): <https://www.4cproject.eu/>; Curation Costs Exchange (CCEX): <https://www.curationexchange.org/>; AV Preserve (AVP) Costs of Inaction Calculator: <https://coi.weareavp.com/>; Keeping Research Data Safe (KRDS) Benefits Framework and a Benefits Analysis Toolkit: <https://beagrie.com/klds-i2s2.php>.

unless the Member States provide the funding and are on board that there is no real prospect for EOSC to provide an ongoing platform for funding. To her, Member State agreement is key; a lot of infrastructure is being proposed without knowing for sure that the Member States are willing to support it in the longer term. Another participant in an Associated Country explained almost all their funding comes from the government—that the science ministry is very important and having preservation written into the Open Science research platform means there is a strong policy driver. She added that becoming part of the EU will make it easier not just for funding but also for policy-level work.

Two participants felt that an EOSC business model with preservation as one of the added-value services was problematic and risky. Their experiences with core grant models for preservation—not charging users per se—found that even with free-to-use services for preservation that people do not use them without some kind of evaluation and flagging beforehand. Neither had encountered models where data or its preservation was sold; instead, they had government-funded infrastructure.

The risk of data loss or reputational harm or sustainability for EOSC arises from the implicit preservation requirements and implied responsibilities. Conversations with the EOSC stakeholders show that they widely recognize that data is 'born vulnerable', but there is a lack of clarity and depth of insight into how to address this issue, which creates reputational risk. This finding, presented in the Interim Statement, was situated in the context of the DPC's Global List of Digitally Endangered Species (BitList).<sup>20</sup> The BitList notes that digital materials are 'Critically Endangered' in the presence of two conditions: when there are material, technical challenges to preservation; and when no agency takes responsibility for them or those agencies are unwilling or unable to meet preservation needs. This is the second-highest alert level and is a precarious position to place emerging EOSC data infrastructure.

The BitList also describes how good practice can reduce the alert level pertinent to any given set of digital materials. This is another important corollary in the context of EOSC where there is significant evidence of good practice by stakeholders (such as DANS, CERN, and UKDS for example) as well as a significant opportunity to model good practice through policy development, training and procedural development. Therefore, an opportunity exists to address these challenges.

The findings from interviews corroborated and elaborated the initial finding of risks to data, reputation, and sustainability within the EOSC infrastructure resulting from lack of digital preservation capacity. Several interviewees provided examples where valuable data has already been lost. One shared an example from CERN, taking the view that continuing to lose data in this way was a reputational risk as well as an operational impediment to research. The larger the infrastructure, the greater the cost, the greater the reputational risk.

Research has shown that the data within EOSC partners are technically complex, heterogeneous and substantial in scale. The role and nature of data vary widely, and so there are different perspectives and insights into the kinds of preservation action that may be required. While not all data are equal, in every discipline represented in the interviews, there was a sense that data would need to be retained for medium to long term. Knowing which elements matter, and what metadata they might require is a subject specialist skill that in many cases can only be captured at the point of creation. Thus, EOSC cannot rely on a single generic canon of 'digital preservation practice'. Instead, workflows should leverage large scale infrastructures while remaining faithful to discipline-specific requirements.

Reflecting on organizational capability, it seems that, so long as roles, responsibilities, and accountabilities remain uncertain, the risk to reputation will remain and, in some senses, be

---

<sup>20</sup> DPC Global List of Digitally Endangered Species (The BitList). See: <http://doi.org/10.7207/DPCBitList20-01>.

disproportionate to the inconvenience of data loss. Efforts to deflect or transfer the reputational risk by waiting for others to step up seem shallow and unlikely to succeed; rather than waiting for the moment when this occurs, likely when a notable data loss or technological breakdown occurs, the risk to reputation should be tackled with urgency so that it can be minimized.

The risk to sustainability is apparent, with no clear plan for ongoing and sustainable funding for digital preservation in EOSC. At the time of writing this report, the business plans for maintaining EOSC (EOSC-Core, EOSC-Exchange) are still in development and debate. The archiving and digital preservation services selected by the ARCHIVER project and other common services will be made available through EOSC-Exchange, but these services are only one part of the larger preservation equation. A business model that presumes the existence of a core grant is risky unless Member States are properly committed to the expenditure that will arise. In other words, elaborating a sustainable business model is challenging for digital preservation and EOSC in general but is essential nonetheless.

### 3.3. Articulation and Assessment of Use Cases

This third section details the findings specifically from the focus groups comprised of participants representing DPC members working with the preservation and management of research data at research performing organizations. Unlike the preceding discussions in 3.1 and 3.2, it draws entirely from notes and data collected during focus group sessions, along with results from an online survey that accompanied the sessions.

Focus group sessions followed the same structure: they began with an overview of FAIR, EOSC, and the FAIR Forever study followed by a brief presentation on the aims, objectives, and structure of the use case activity to provide enough contextual background to guide but not bias their input. Section 2.3 of this report provides more detail about the design of the focus groups, including how the work of the EOSC Skills and Training Working Group, DigCurV Framework, and DPC Labour Market Analysis informed and shaped the persona groupings, the rationale for only focussing on the Technician, Practitioner, and Manager types for the activity, and findings about the participants.

This section begins with findings from the construction of use case scenarios in the focus groups, with discussion points and conclusions that emerged presented throughout. Five use cases are presented for consideration, each with a summary of their SWOT (Strengths, Weaknesses, Opportunities, Strengths) assessment.

#### Use Case Group Activity

##### Persona Goals and Needs

Figure 6 below summarizes views expressed during the group activity, illustrating key goals and needs which participants identified and discussed. There are separate lists for each of the three personas of Technician, Practitioner and Manager. The text in the bullet points was provided by the designated notetakers from each group, reflecting how they were communicated and discussed. Consequently, some elements points encompass more than one goal or need.

# Persona Goals and Needs

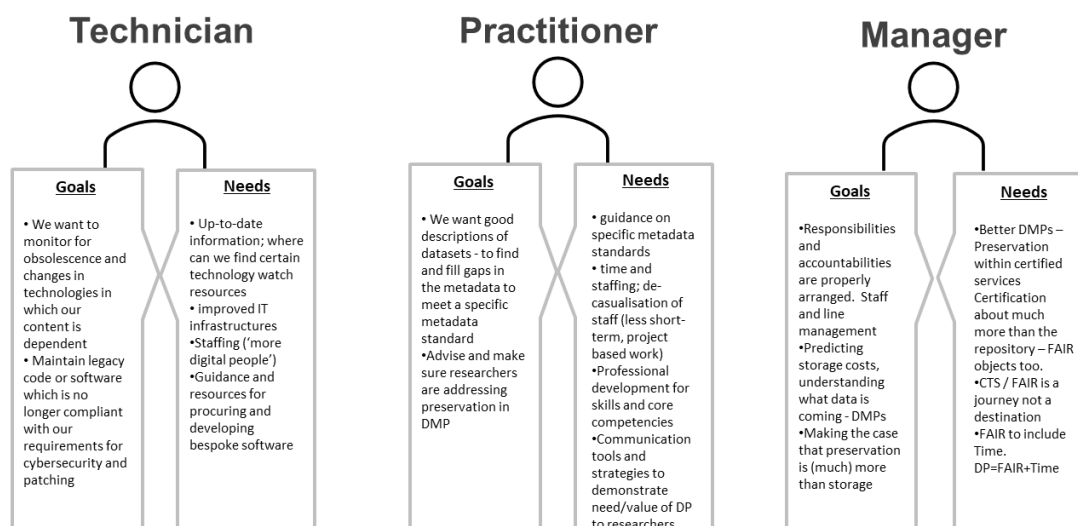


Figure 6. Persona Goals and Needs from group activity

The goals, needs, and areas covered during the activity, those listed above, set the foundation for the articulation of use cases for subsequent assessment and development of potential candidate model services and recommendations for EOSC.

Key findings and points raised within each of the groups are noted below:

- **Conversations in the Technician group** brought attention to active data issues. One commented if you are open with research data, and share the data early, that you need to be able to do that properly in the context of sustainability and preservation, particularly as it concerns persistent identifiers (PIDs) and digital object identifiers (DOIs). He added that in terms of what EOSC can provide, they should work to make sure there is an ability to cite a specific version of data and be confident that it will be that version that continues to be cited over time even as new versions or iterations of the data are created.
- **Conversations in the Practitioner group** addressed assumptions of data stewardship roles and responsibilities, reflecting on the resources and support needed to carry out such duties effectively. In other words, that sustainability involves the everyday work of maintenance and quality assurance and, therefore, a need for both financial and professional development support. Comments arose on the challenges with data management plans (DMPs) coming across as checklists of 'what' without communicating the 'why' behind preservation. One brought up domain and discipline specific issues, specifically how researchers in the Arts may not perceive their outputs as data or may not wish to make them open and, related to this, how the preservation of data should be communicated to all researchers regardless of their plans to make the data FAIR and open.
- **Conversations in the Manager group** centred around organizational and cultural challenges of digital preservation at all levels. One commented that there is a clear need for research communities (whether talking about the 'community' of researchers at the RPO, or in discipline-specific, regional or national contexts) to take preservation seriously, and for EOSC to address it more clearly. In terms of what EOSC might provide through resources and

services, they responded there needs to be an area of EOSC specific to digital preservation, and also a collaborative working group or task group dedicated to addressing the challenges in and across the work of EOSC communities. There was a strong steer that data management plans lacked accountabilities and that although a repository could be certified, if the data being presented was poor quality, then preservation was impossible. Managers feared that, without strong accountabilities for preservation from the point of creation through deposition, repositories are being set up to fail. There was scepticism that EOSC's vision addressed this.

The following issues, presented in Figure 7, were considered as they applied to the goals listed by participants.

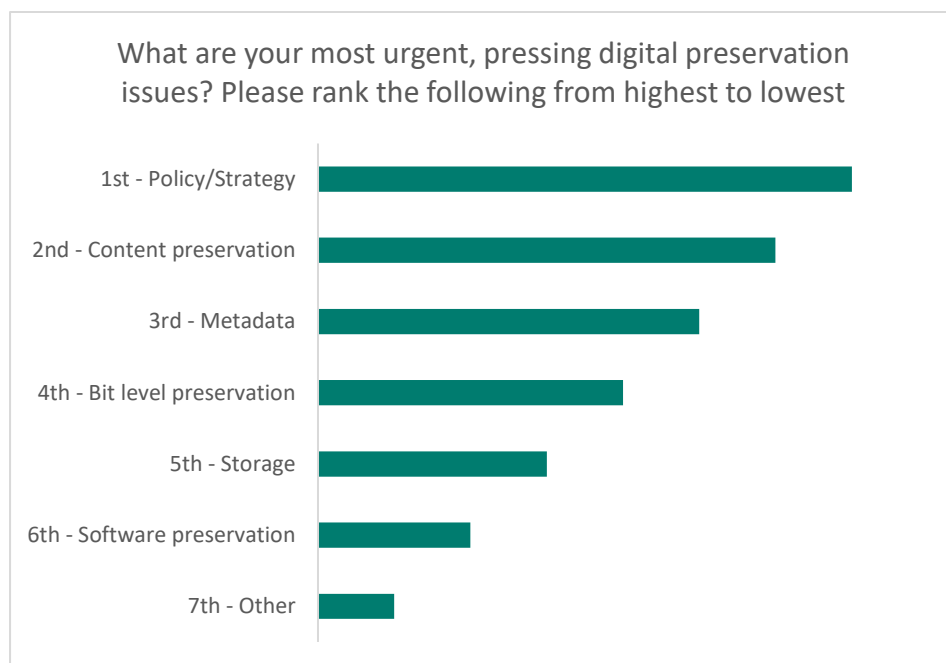


Figure 7. Most Urgent, Pressing Preservation Issue responses from the survey

Policy and strategy were frequently mentioned during discussions with participants—with participants discussing digital preservation policy at various levels:

- Participants mainly discussed organizational or institutional policies that apply to digital preservation and research data management.
- They also discussed the need for a digital preservation policy on its own, to make clear its importance and outline the scope, challenges and, most importantly for this research, the roles and responsibilities along with strategies and actions necessary to ensure continued access to digital materials (for as long as necessary)
- They also talked about challenges with forming national digital preservation policy with language becoming general but it still being useful for those at organizations without a policy (to help shape and direct) and those with a policy (for accountability)

The following obstacles presented in Figure 8 below were similarly considered in relation to the needs addressed by participants through the activity.

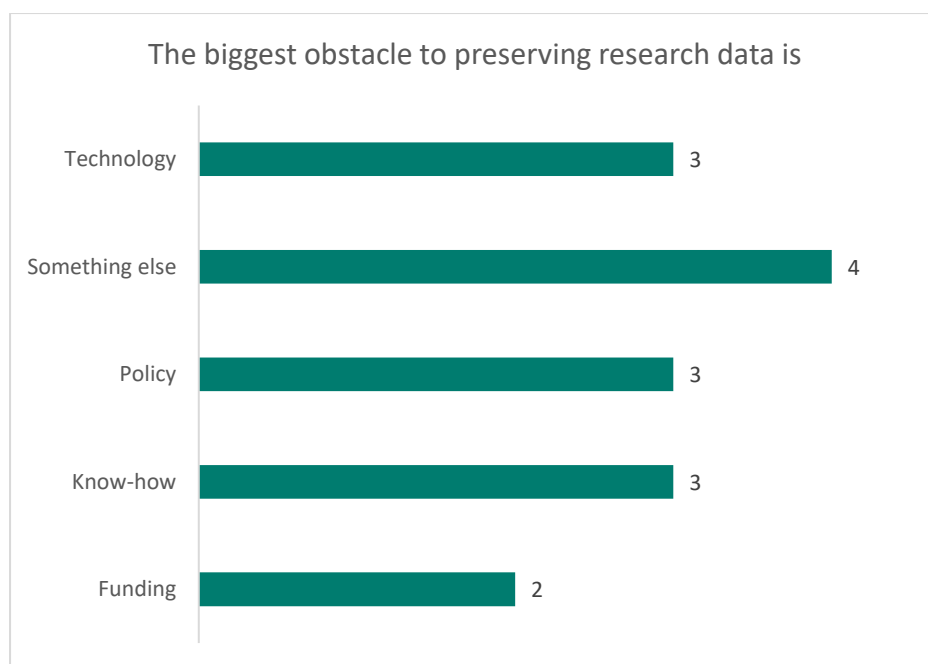


Figure 8. Biggest Obstacle to Preserving Research Data responses from the survey

There was no consensus on one biggest obstacle for preserving research data. Indeed, four participants chose 'something else', explaining that they saw the main obstacle as

- lack of professional recognition
- systematic changes in research culture
- lack of understanding of the need for digital preservation
- lack of documentation (incl. documentation regarding software)

Together, these obstacles circle back to discussions about policy and strategy as a clear goal to consider and also advocacy through policy—the need for making the importance of digital preservation and work of those doing it.

### Persona Solutions and Gaps

The group activity also inquired on what was going well to address the goals and needs, or solutions, which would benefit those in the European research community dealing with the same and possible gaps if not addressed or included in EOSC resources.

The findings from this activity are presented in Figure 9's diagram below with the suggested solutions, represented in the ovals, as provided by the designated notetakers from each group to reflect how they were discussed.

## Persona Solutions and Gaps

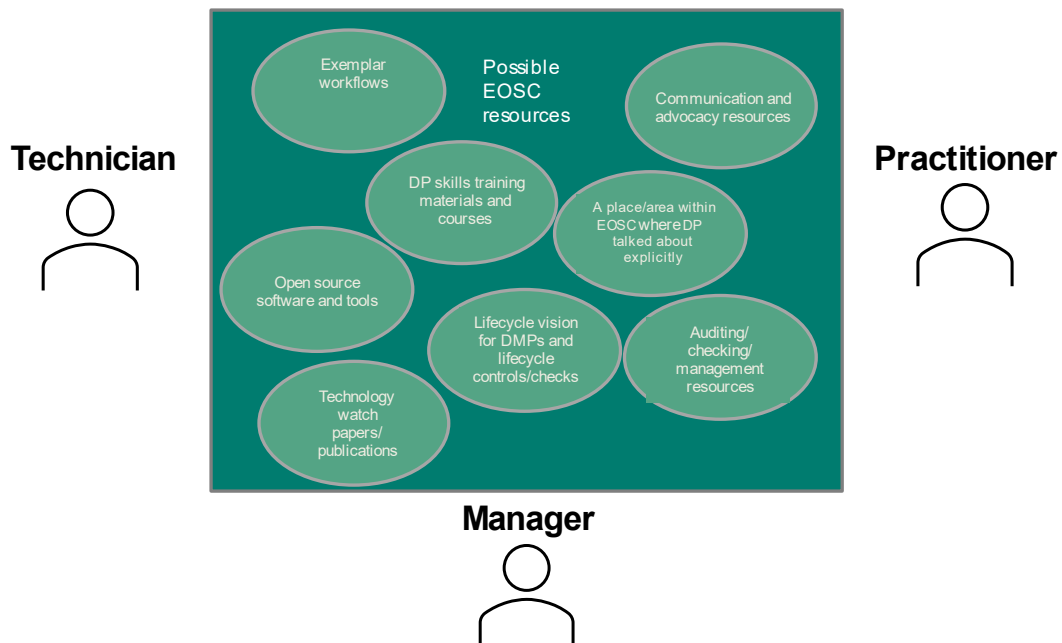


Figure 9. Persona Solutions and Gaps from group activity

The diagram positions these suggested solutions together in the rectangle representing 'Possible EOSC resources' to consider how they will—or could—be made available and accessible as services or resources through the EOSC.

The suggested solutions were more general than specific in two ways. First, they were general enough to be useful and beneficial to more than one persona's goals. For example, exemplar workflows for content preservation can include actions performed by technicians and practitioners (e.g. migration, documentation, quality control) that are developed and periodically reviewed by managers. For this reason, the diagram groups the solutions together in the rectangle.

Second, each of the suggested solutions was general enough to encompass more than one kind of resource as defined by EOSC.<sup>21</sup> In particular, the suggested 'place/area within EOSC where digital preservation talked about explicitly' was general enough to include any or all of the other suggested solutions. Participants commented that it was challenging to think of use case scenarios for EOSC at this stage where the system and services are still in development.<sup>22</sup>

Moreover, in the group discussion following the activity, participants from the different persona groups commented on the challenges of working out what digital preservation services would make

<sup>21</sup> An EOSC resource is any component available by EOSC; a component is a discrete actor, service, policy, data or infrastructure that can be considered an asset of or constituent part of EOSC; a service a means of delivering value for the end-user by facilitating outcomes the end-user wants to achieve.

<sup>22</sup> The activity was on 20th November, before updates to the portal from EOSC Enhance and ARCHIVER selection of winner consortia for its prototype phase.



sense to deliver given the complex picture of services nationally, locally, and discipline or domain-specifically.

To help refine this, a survey question asking participants to rank the following categories of EOSC resources (Figure 10) was used to consider how they related to those produced in the activity.

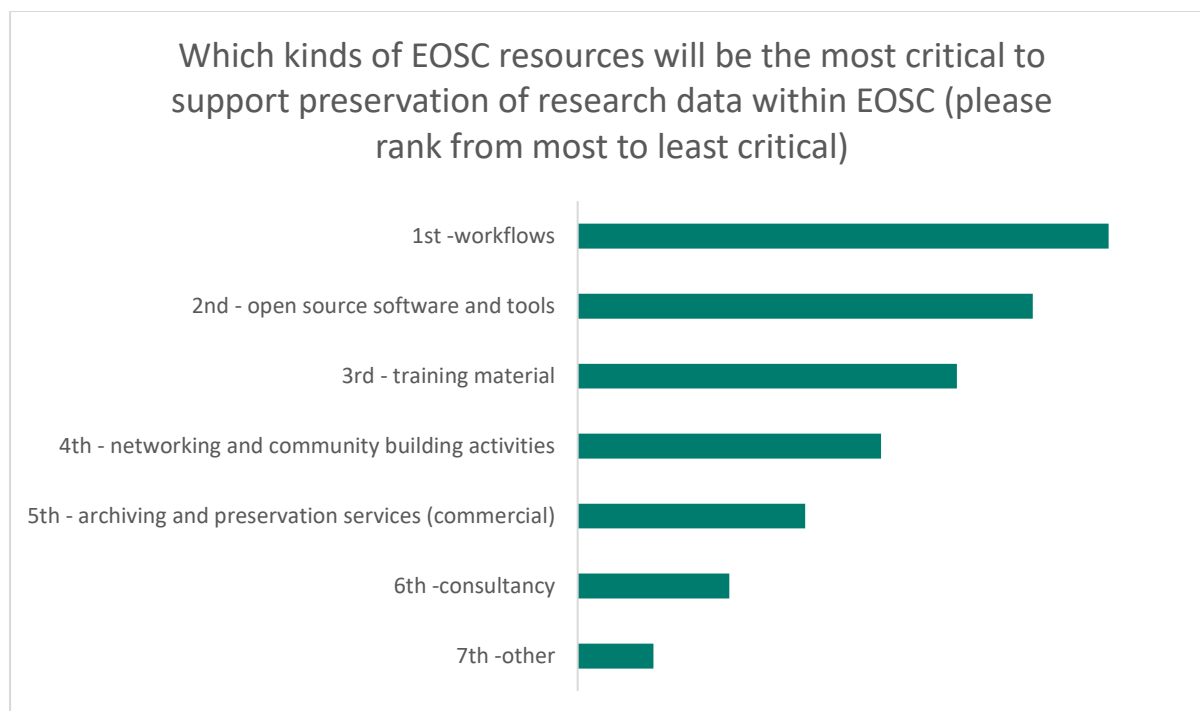


Figure 10. Most Critical EOSC Resources for Preservation of Research Data survey responses

Workflows dominated, mirroring the 'exemplar workflows' noted in the activity. Open-source software and tools were also mentioned, especially for software preservation. Training material was viewed as essential to not only building the preservation skills of staff but also pointing to materials for researchers interested in preservation. Networking and community building activities ran throughout discussions, as they connected to 'communication and advocacy resources'.<sup>23</sup>

### From Personas to Preservation Use Cases: Candidate Services Models for EOSC

The 'persona' approach offers a general framework to pinpoint particular concerns, goals, needs and requirements for the preservation of research data, drawing on the experience of those directly working with research data management and preservation at research performing organizations.

There continue to be ongoing developments with the EOSC to shape the services and resources that it will make available. Presently, the discovery of resources within EOSC occurs primarily through the EOSC Portal, where discovery is based on a categorization of services. While there have been recent changes to the portal through the EOSC Enhance project, at the time of conducting this stage of research the categories included: networking, compute, storage, sharing and discovery, data management, processing analysis, security and operations, and training and support.

The examples of preservation solutions found in this study point to potential preservation gaps in EOSC that, if not addressed, will impact the long-term preservation of research data federated and shared through EOSC. For this reason, five examples were selected and developed further as generic candidates for EOSC services and resources.

<sup>23</sup> Those in the 'Other' category included Registries of Digital Repositories, Rights Responsibility Lifecycle, and Standards but they were not addressed during the activity or discussion.

## Articulation and Assessment of Use Cases for Candidate Services in EOSC

This section presents five use cases for candidate services within EOSC, which were derived and developed from the focus group sessions and use case scenario activity. In each case, a further SWOT assessment was conducted by DPC staff to identify the strengths, weaknesses, opportunities and threats associated with each.

### Use Case 1: A Legacy Code Software Preservation Service within EOSC

**Brief Description:** This generic preservation use case concerns the goal to preserve legacy code or software to a sufficient standard that underlying processes are reproducible and research outcomes credible; providing mechanisms to document and assess external dependencies; and supporting good practice in the design of preservation-ready code within research. The actors interested in achieving the goal would use a service or programme in EOSC that focusses specifically on strategies for preserving legacy code or software. The service or programme would include guidance and next steps to work through what is needed, places to locate identification schemas for source code and documentation regarding software, and also to keep up to date to monitor for obsolescence and changes in technology in which the content is dependent. The service would, in turn, be monitored and validated to ensure that performance is sufficient to the task and that preservation actions are demonstrably sound and properly documented. Accomplishing this goal will allow the reliable use and re-use of the software for research, allowing credible reproducibility for as long as this may be required.

**Associated Persona:** Technician (primary)

**EOSC Target Users:** Software Engineer, Data RI Support Professional, Data Curator, Data Steward, Researcher, EOSC Enabler<sup>24</sup>

**EOSC Stakeholder Groups:** Resource Providers, Service Providers, Research Funding Organizations, Research Performing Organizations

#### Strengths:

- Federated EOSC infrastructure enabling sharing of resources and services
- Existing specialized knowledge and expertise within the EOSC research community (Software Design and Development)
- Aligns with current EOSC research software initiatives and projects underway (FAIR 4 Research Software Working Group)
- Improved reproducibility for researchers

#### Weaknesses:

- Dependency on knowledge and expertise of specialized group within the EOSC research community
- Outside the current scope of services. The EOSC Strategic Research and Innovation Agenda (SRIA) states that the archiving of software has to be part of a general-purpose software archive outside of EOSC; and recommends the Software Heritage Archive for archiving research software to consider leveraging software preservation initiatives at the institutional-level
- Federation of resources and services depends on buy-in of resource and service providers

---

<sup>24</sup> EOSC Target Users are those as outlined by the EOSC Skills and Training Working Group. For example, the EOSC Enabler refers to a researcher of a discipline with technical skills who might design and partially implement discipline-specific applications using this EOSC service. For more details of each of the user types, see: EOSC Skills and Training Working Group, Identifying Digital Skills Sets for EOSC, August 2020: <https://www.eoscsecretariat.eu/news-opinion/identifying-digital-skill-sets-eosc>.

### **Opportunities:**

- Sharing and co-creation of legacy code and software with cloud-based infrastructures currently outside of EOSC
- Development and collaboration with the digital preservation community

### **Threats:**

- Focus on a specific material type or kind of research output over others
- Inability to adapt/respond to changing technology and obsolescence
- Legal restrictions (e.g. intellectual property rights)

## **Use Case 2: A Mechanism to Ensure Accountability and Implementation of Preservation in DMPs within EOSC**

**Brief Description:** This generic preservation use case concerns the safeguarding research data throughout the research lifecycle, identifying and assigning preservation actions and accountabilities within data management plans (DMPs). The actors interested in achieving this goal would use a mechanism or service within EOSC to review, check, and audit DMPs, ensuring the preservation actions are sufficient to requirements, implemented successfully and in a timely fashion, and respond accordingly if not. It would provide resources for developing different automated machine-actionable DMPs and, perhaps more importantly, keep users up to date on the current state of strategies and actions for auditing data management planning, assigning actions and accountabilities to actors. The mechanism would include templates and tools for controls and checks to mitigate risks of loss or damage as well and ensure the integrity and reliability of data. Repositories would thus be able to plan more effectively for the infrastructure they need; funders would have confidence that data objects were FAIR throughout their lifecycle; data creators and their institutions would be able to demonstrate compliance; and data users would be able to understand the full biography and provenance of data.

**Associated Persona:** Practitioner (primary)

**EOSC Target Users:** Researcher, Data Curator, Data Steward, Data RI Support Professionals, EOSC Enabler

**EOSC Stakeholder Groups:** Resource Providers, Service Providers, Research Funding Organizations, Research Performing Organizations, Governance

### **Strengths:**

- Aligns with FAIR guiding principles for scientific data management and data stewardship, and current EOSC data management initiatives and projects underway (Skills and Training Working Group, FAIRsFAIR project and FAIR Evaluation Services)
- Makes FAIR principles actionable
- Quality control and accountability mitigating risks of data loss or damage
- Existing specialized knowledge and expertise within the EOSC research community (Research Data Management)
- Clarification and consistency of responsibilities

### **Weaknesses:**

- Inapplicable with certain limits of sovereignty where Member States may decide data management has to follow specific rules
- Limited resources at research institution to hire and train staff to implement the mechanism
- Unclear limits or obligations related to accountability

- Scalability
- Assessment requires strong disciplinary layer to ensure effectiveness

**Opportunities:**

- Planned European Commission proposal/mandate for mandatory development and implementation of DMPs (even if not making research data open)
- Development and collaboration with the digital preservation community
- Move towards professionalization of digital preservation responsibilities and activities

**Threats:**

- Lack of comparability/interoperability with digital preservation strategies
- Dependency on external funding
- May become an overhead or burdensome to small, fast-moving research

**Use Case 3: A Business Case Factory or Service for Preservation Cost Modelling within EOSC**

**Brief Description:** This generic preservation use case concerns the goal of building business cases to acquire funding and resources supporting digital preservation activities. The actors interested in achieving this goal would use a business case factory service within EOSC for information and guidance on how to create and develop a business case for the preservation of research data research particular to their organizational or institutional setting, research funding organization, or other public or private funding body. This resource would include resources and tools for auditing organizational readiness and predicting costs such as digital preservation models, maturity models, cost modelling for curation and preservation, shared templates, and other resources.

**Associated Persona:** Manager (primary)

**EOSC Target Users:** Data Curator, Data Steward, Funder, Policy Maker, EOSC Enabler, Researcher

**EOSC Stakeholder Groups:** Resource Providers, Service Providers, Research Funding Organizations, Research Performing Organizations, Governance

**Strengths:**

- Federated EOSC infrastructure enabling the sharing of resources and services for preservation cost modelling
- Aligns with current EOSC funding and sustainability initiatives and projects (Sustainability Working Group)
- Existing knowledge and expertise within the EOSC community (Research Funding Organizations)

**Weaknesses:**

- Ongoing development of minimum viable funding model for EOSC
- Limited information of costing models for preservation from fragmented and complex nature of the European research-funding landscape
- Scalability
- Dependency on certain funding streams

**Opportunities:**

- Development and collaboration with the digital preservation community
- Opportunities for awareness-raising and introduction of external funding streams

**Threats:**

- Limited uptake or use due to varied and complex funding sources
- Dependency on external funding

#### Use Case 4: A Programme to Support Researchers with Preservation at the Point of Creation within EOSC

**Brief Description:** This generic preservation use case concerns the goal to safeguard research data by incorporating preservation strategy and actions at the point of creation or conceptualization rather than at the point of transfer or deposit. The actors interested in achieving this goal would engage in a programme provided through EOSC for information and guidance on data management and preservation planning, outlining the reasons and benefits of early preservation interventions along with general guidance and next steps. In the generic EOSC Portal, this service could provide general resources on filling out preservation sections of a DMP and overviews of different storage options, bit-level and content preservation actions, preservation metadata standards, and preparing for DMP audits to ensure the preservation actions are implemented, and respond accordingly if not.<sup>25</sup> More tailored programmes could be developed and provided through the EOSC thematic and regional portals.

**Associated Persona:** Researcher (primary)

**EOSC Target Users:** Researcher, Data Curator, Data Steward, Data RI Support Professional, Research Software Engineer, Data Sci. Data Analytics, EOSC Enabler

**EOSC Stakeholder Groups:** Resource Providers, Service Providers, Research Funding Organizations, Research Performing Organizations, EOSC Governance

#### Strengths:

- Aligns with FAIR guiding principles for scientific data management and current EOSC data management initiatives and projects underway (Skills and Training Working Group)
- Early interventions for quality control and mitigating risks of data loss or damage
- Draws on specialized knowledge and expertise within EOSC research community (discipline and domain knowledge of Researchers)

#### Weaknesses:

- Limited resources for developing digital preservation skills and training programmes
- Researcher buy-in or incentivization and rewards for early interventions

#### Opportunities:

- Planned European Commission proposal/mandate for mandatory development and implementation of DMPs (even if not making research data open)
- Development and collaboration with the digital preservation community
- Move towards professionalization of digital preservation responsibilities and activities

#### Threats:

- Lack of control over data

---

<sup>25</sup> Both Use Case 3 and Use Case 4 would include support for cost modelling. The two differ in that the envisioned Use Case 3 service (Business Case Factory or Service for Preservation Cost Modelling) would offer specific support for cost modelling amongst various EOSC users whereas the Use Case 4 service (Programme to Support Researchers with Preservation at the Point of Creation) would offer support specific to researchers including (but not limited to) cost modelling resources.

- Dependency of external funding

### Use Case 5: A Mechanism to harmonize Digital Preservation Policy Across Institutions and communities of practice within EOSC

**Brief Description:** This generic preservation use case concerns the goal of developing and implementing digital preservation policy at various levels. The actors interested in achieving this goal would come to EOSC to find a documented set of intentions, expectations, goals, results and requirements for digital preservation in the context at EOSC, providing core information and guidance to support the creation, monitoring, optimization, and synchronization of digital preservation policy across multiple European research organizations and institutions. The mechanism will provide a resource and a service to various stakeholders. It will provide an EOSC-endorsed document outlining the current risks and challenges for the preservation of research data across EOSC with expected roles of different EOSC stakeholders so that responsibilities are transparent and accountabilities actionable. The sharing of policies within the mechanism will also streamline the creation and coordination of policy. Furthermore, the mechanism could also be developed to enable organizations to publicly 'sign-up' to a shared policy and monitor the implementation of policy.

**Associated Persona:** Manager

**EOSC Target Users:** Data Curator, Data Steward, Funder, Policy Maker, EOSC Enabler, Researcher, Data RI Support Professional, Research Software Engineer, Data Sci. Data Analytics, Citizen

**EOSC Stakeholder Groups:** Resource Providers, Service Providers, Research Funding Organizations, Research Performing Organizations, EOSC Governance

#### Strengths:

- Shared goals and strategy for shared development output
- Transparency and accountability of roles and responsibilities
- Technical and policy-level convergence

#### Weaknesses:

- Dependency on partners involved (technical, organizational, financial)
- Inapplicable with certain limits of sovereignty where Member States may follow specific rules
- Scalability

#### Opportunities:

- Awareness-raising, training, education, and improved community-specific support
- Development and collaboration with the digital preservation community
- Awareness-raising and professionalization of digital preservation responsibilities and activities
- Emergence and alignment of procedures as implementation of policy

#### Threats:

- Lack of comparability/interoperability of digital strategies
- Changing internal strategies of EOSC

## 4. Conclusion, Recommendations, and Roadmap

The findings from the three stages of data collection and analysis—from the desk-based assessment, the interviews with representatives of the ESFRI Cluster and Regional projects, and the focus group sessions on use case scenarios for preservation services in EOSC—support and extend the three key arguments presented in the Interim Statement:

- Digital preservation is not explicit in the EOSC vision: it needs to be.
- Roles, responsibilities, and accountabilities for preservation in EOSC are opaque: they should be clarified.
- There is a risk to data, reputation and sustainability: EOSC cannot achieve its goals in the long-term unless they are addressed.

There are clear points of intersection between the FAIR principles, data management and preservation planning for EOSC. These include a persistent identifier policy for FAIR data, a culture of data management planning, certification of trustworthy repositories compliant with FAIR standards, and plans for a Web of FAIR Data and Services to support the development and implementation of digital preservation plans.

At the same time, critical elements of digital preservation are not covered by EOSC’s interpretation of the FAIR principles and so are largely absent from the EOSC vision. Digital preservation involves the series of managed activities necessary to ensure the continued access to research data for as long as necessary, which encompass actions and interventions throughout the lifecycle—not just at the creation of FAIR data or the transfer and ingest to a certified archival repository.

While the expected outcomes of the ARCHIVER project—including the ‘research ready’ commercial services, trustworthy repository services, and set of derived ‘rules of participation’—will greatly benefit a number of EOSC stakeholders and improve the preservation of research data by its adopters, these should not be the only digital preservation services or resources provided through EOSC. This study’s findings, especially those from the focus groups, show that immediate, pressing preservation needs and requirements of those currently working with research data at research performing organizations extend beyond the availability of data archiving and preservation commercial services available through EOSC.

The five use cases highlight potential gaps in preservation services that must be addressed within EOSC; otherwise, data remains at risk. They also present opportunities for improving the preservation of research data in its various forms, identifying the kinds of resources that will support and enhance those opportunities, highlighting the role of those working in research performing organizations as both provider and user within the EOSC ecosystem.

With these particular stakeholders and use case preservation services in mind, the next sections provide recommendations on the emergent roles and responsibilities for the preservation of digital research data followed by possible enhancements to the preservation of research data corresponding to the identified action areas for EOSC as provided in the most recent Strategic Innovation and Research Agenda.<sup>26</sup>

### 4.1. Long-term Preservation Roles and Responsibilities

The EOSC ecosystem is comprised of federated e-infrastructures, research infrastructures, stakeholder organizations and projects that contribute to or use EOSC. It is also comprised of individuals and groups of stakeholders who can affect, be affected by, or perceive themselves to be affected by EOSC activities or decision.

---

<sup>26</sup> SRIA Version 0.9 (provided by SWG contact).

The EOSC Landscape Working Group identifies three groups of these stakeholders:

- **Research funders** – Horizon Europe and other programmes (such as Digital Europe), tools, institutions, businesses or funding bodies contribute to funding EOSC both directly and indirectly through state-of-the-art expertise and investment in national infrastructures
- **Policymakers**. - The Steering Board including EU Member State and Associated Country representatives to make strategic decisions on the development and evolution of the EOSC
- **Other relevant stakeholders**
  - e-infrastructures, such as PRACE, GEANT, OpenAIRE, EUDAT, EGI, also referred to as delivering horizontal services.
  - Research infrastructures, such as ESFRI Research Infrastructures Cluster and Regional projects, also referred to as delivering vertical or thematic services.
  - Data and research initiatives, such as RDA, offer global platforms for sharing expertise.
  - Cloud providers, including commercial parties such as Amazon, offering services to research.
  - Cloud community.
  - Research communities
  - Research-performing organizations.

This study reviewed the provided EOSC governance documents not just to understand digital preservation in the EOSC vision but also to see where more policy and guidance can affect EOSC actions and decisions for the medium to long term. It identified a significant need to clarify the roles and responsibilities and accountabilities for preservation within the EOSC ecosystem. This will mitigate the risks of data loss, reputation, and sustainability by offering greater transparency and accountability among different stakeholders. The organizational and operational aspects of preservation laid out in this report show that there is a clear need for EOSC Governance to contribute to policymaking for digital preservation and guiding the direction of the envisioned EOSC. Modelling this policy development itself, EOSC Governance can guide and sustain policy towards successful implementation at the organizational and community levels.

This is most easily done through the establishment of a Digital Preservation Task Force within EOSC. This would have three immediate purposes:

- To establish a high-level digital preservation policy across EOSC
- To establish a workplan for policy development and implementation within EOSC services and partners
- To act as a point of contact and high-level advocate for preservation within EOSC

In the longer term, three further roles would emerge:

- To monitor policy implementation across EOSC partners
- To act as a virtual centre of excellence for digital preservation within the EOSC community
- To act a point of contact between EOSC and other digital preservation communities outside of the research data community

The uncertainty over roles, responsibilities and accountabilities for preservation found from the stakeholders in this study extends to the other parts of the EOSC ecosystem. Another key finding from the study—relating to those working in EOSC Research Infrastructures and Research Performing Organizations—is that researchers, disciplinary and cross-disciplinary research communities, Member States and EOSC are all expecting commitments from each other but lack the support to make commitments of their own. EOSC faces a challenge of making sure decisions are made at the right level, with accountability and transparency so that shared outcomes can be optimized.



For the above reasons, we recommend the development of a framework and strategy at the governance level that delineates responsibilities and leadership for preservation within the context of research data in EOSC. Suggested roles and responsibilities should be put into some form of EOSC governance documentation. It should include, at the least, an identification of the key roles and responsibilities for preservation documented within the SRIA or in another higher-level document. There should also be the creation of a working group, task group, or other governance structure for coordinating the work on preservation actions within EOSC to ensure that it will be properly addressed throughout EOSC.

To help inform and guide this, the next section uses the findings from the research to offer foundations for a framework of digital preservation roles, responsibilities and accountabilities.

## Foundations for a framework of preservation roles and responsibilities

Examples of suggested digital preservation roles and responsibilities within EOSC are presented below, noting areas of accountability corresponding wherein the specific responsibilities within the auditing process are assigned to each of the roles.

### **EOSC Board**

The EOSC Board will ultimately own the reputational and financial risks of the EOSC Association and the implementation of the EOSC strategy. It has a role to ensure that digital preservation functions are clarified and articulated. While the EOSC Association statutes Article 6.3 makes it clear that Directors of the Board are excluded from membership of operational and advisory bodies, another simple way to address this would be to have one member EOSC Board member take on a role of rapporteur on behalf of the Board to directly communicate and liaison with a Digital Preservation Task Group.

### **Policy Makers**

The role of policy makers includes the responsibility to make digital preservation explicit in the EOSC vision across the wider community, providing an overarching policy and strategy to guide EOSC stakeholders with the development and implementation of preservation actions across the lifecycle. They hold accountability for revisiting, reviewing, and revising guiding policy.

Use Case 5, A mechanism for digital preservation policy across institutions within EOSC, would offer valuable resources to help research policy makers with these responsibilities.

### **Research Funders**

The role of research funders includes the responsibility of review and auditing throughout the data lifecycle. Because digital preservation is path dependent, the certification and audit of preservation commitments and their implementation should begin early in the data lifecycle and recur through key intervention points.

We recommend that funders assert this role transparently to ensure better accountability for arranging, scheduling, and performing audits of preservation actions within data management plans. If possible, funders should also assess the capability of digital preservation providers used to store and manage deposited research data—irrespective of whether they are at public institutions or research performing institutions, or provided privately by commercial services.

We suggest that funders establish a system of incentivization or obligation for performance, providing or withdrawing funding based on periodic auditing of performance against data management plans. For instance, funders might offer additional or supplementary funding for exemplar preservation plans which detail specific costs of storage and labour, or are part of the

EOSC community use cases. They might also withhold funding if certain preservation actions are not being taken, or withdraw funding if data is not deposited for preservation.

Opinions differ to which approach is most appropriate. Responses to the survey given to participants during the use case scenario sessions, shown in Figure 11 below, varied.

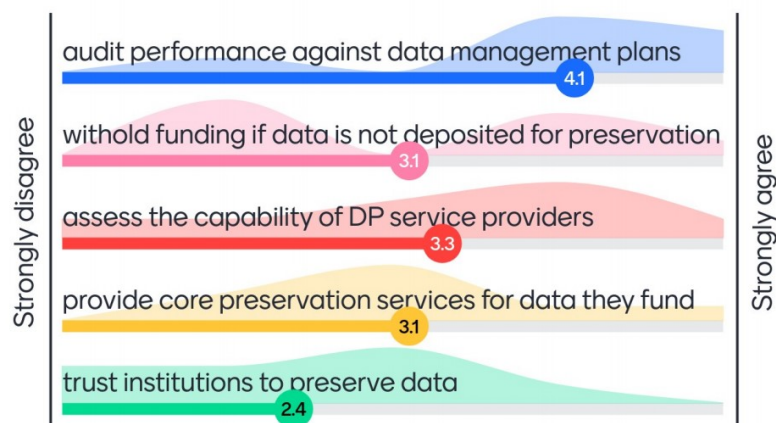


Figure 11. Responses to Statements on Responsibility of Funders from the survey

With this designated responsibility for review and auditing, accountability for data loss or damage would fall become immediately clear.

It is therefore recommended that these expectations and requirements for preservation performance are made clear to recipients of funding, noting researchers' responsibility in the periodic review of plans or audits of data objects.

Additionally, we recommend that funders commission repositories to undertake audits and check if progress is satisfactory. The repository can then report to the funder if progress is unsatisfactory or there are any issues, and the funder can then respond with whatever obligations were established with the researcher at the start.

Three of the use cases for EOSC preservation services presented in this report would offer valuable resources to help research funders with these responsibilities

1. A mechanism to ensure accountability and implementation of preservation in DMPs
2. A business case factory or service for preservation cost modelling
3. A programme to support researchers with preservation at the point of creation

## Researchers

The role of researchers includes the responsibility to prepare data for preservation; to plan and implement basic preservation actions throughout the data lifecycle, beginning as early as possible (at point of conceptualization), to account for those preservation actions within a data management plan, and to participate in the periodic review and auditing throughout the lifecycle as arranged by funders or institutions.

We recommend that this role for researchers is transparent in the EOSC vision to better ensure accountability for the implementation of preservation actions within data management plans. Researchers should provide plans with basic preservation actions including the selection of an archive or repository ahead of time, storage of multiple copies in different locations, DOIs and metadata with detailed provenance, and integrity checking throughout the project.

With this designated responsibility for review and auditing, accountability for data loss or damage would fall in their hands if they have not implemented the actions within their plan, or have not participated in period reviews or audit as specified by funders. If there is data loss or damage for reasons beyond the scope or control of the researcher as indicated in the preservation plan, they will not be held accountable.

In line with the survey responses shown in Figure 12 below, we also suggest that researchers foster and develop preservation skills within the context of their research area and within wider communities of practice. This will not only ensure their data will be accessible and reusable over time but will also provide their knowledge and disciplinary insight into ways to tackle the challenges of metadata and interoperability in and across scientific domains and disciplines.

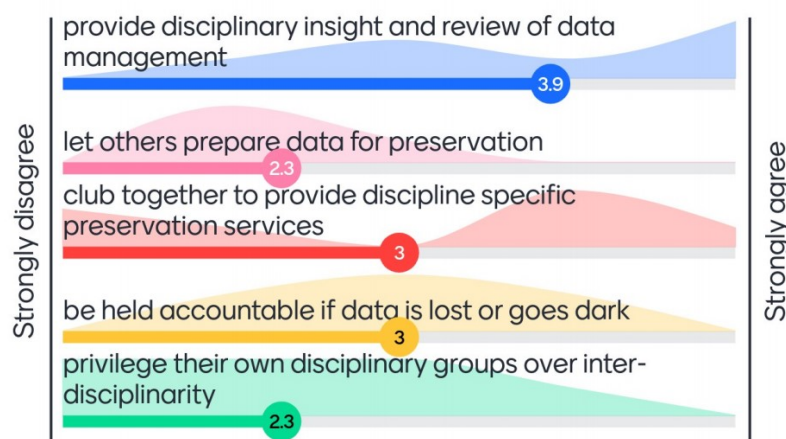


Figure 12. Responses to Statements on Responsibility of Researchers from the survey

Four of the use cases for EOSC preservation services presented in this report could offer valuable resources to help research funders with these responsibilities:

1. A legacy code or software preservation service
2. A mechanism to ensure accountability and implementation of preservation in DMPs
3. A business case factory or service for preservation cost modelling
4. A programme to support researchers with preservation at the point of creation

### Research Performing Organizations and Institutions

The role of research performing organizations includes many responsibilities from helping researchers prepare data, to storing and managing deposited data during a project, to selecting and archiving data after a project ends. The institution should be, in essence, a trustworthy repository and provider of FAIR data as well as a trustworthy provider of information and preservation services.

If research data is deposited in the institution, the role of the institution should include the responsibility of review and auditing throughout the data lifecycle. This may occur separately or with funders depending on the project and outsourcing to third parties (see Figure 13 below). However, we recommend that funders should commission repositories to conduct audits, and repositories should undertake these audits. In other words, repositories have a function to undertake the audits to work with both researchers and funders; working with researchers to address key areas of preservation in plans and working with funders to review and report back satisfactory or

unsatisfactory progress to respond according to obligations such as withholding funds.



Figure 13. Responses to Statements on Responsibility of Institutions from the survey

We recommend that this role of institutions be made clear in the EOSC vision so that accountabilities are clear at the organizational level. However, the findings of this study show this role and responsibilities for preservation within these institutions to be largely assumed. There are key roles within the institution that should too be recognized and made clear in the EOSC vision.

While the findings from discussion and survey of participants in the use case sessions revealed that some see preservation as a shared team effort (six of the fifteen participants responded 'It really is an equally distributed team effort'), accountability is more effective when it is given to individuals rather than an institution. Even if responsibilities are evenly distributed, it should still be made it clear how everyone holds accountability. Therefore, we suggest that an overview of different preservation role types within research performing organizations and institutions be more explicit in the EOSC vision.

Additionally, the accountability of the different roles will offer an avenue for further investigation and assessment of current preservation capabilities as well as functions, salaries and funding streams for sustaining the preservation of research data in the context of EOSC.

With this mind, the different roles associated with the personas offer a starting point, for example

- The role of Executive is responsible for final approval and oversight of policy and strategy.
- The role of Manager is responsible for the strategic implementation of preservation planning and actions through managing staff and services, grant applications and procurement. They can lead the auditing of performance against data management plans.
- The role of Practitioner is responsible for providing correct information and guidance on preservation planning (preservation literacy), implementation of preservation actions, and quality assurance. They can help researchers with completing data management plans and prepare data for preservation.
- The role of Technician is responsible for designing, developing, improving the information technology infrastructure for preservation planning and actions. They can help design and automate the data management planning and auditing process.

All of the use cases for EOSC preservation services presented in this report could offer valuable resources to help research funders with these responsibilities:

1. A legacy code or software preservation service
2. A mechanism to ensure accountability and implementation of preservation in DMPs

3. A business case factory or service for preservation cost modelling
4. A programme to support researchers with preservation at the point of creation
5. A mechanism for digital preservation policy across institutions within EOSC

## 4.2. Strategies, Activities and Enhancements to the Preservation of Research Data

After identifying the key roles in the ecosystem and suggested recommendations, an essential part of improving the preservation strategy is identifying the support and enabling services that allow them to improve in their role or develop and embrace new opportunities.

The use cases presented in this report offer five examples of preservation services or resources to support the different preservation roles within EOSC:

1. A legacy code or software preservation service
2. A mechanism to ensure accountability and implementation of preservation in DMPs
3. A business case factory or service for preservation cost modelling
4. A programme to support researchers with preservation at the point of creation
5. A mechanism for digital preservation policy across institutions within EOC

These five use cases should be developed as recommendations. They should be understood as additional to the existing elements of the EOSC infrastructure which support digital preservation functions.

We further recommend that these services be packaged so that an effective catalogue or listing of digital preservation services be included in the EOSC—whether a landing website, portal, registry, catalogue, or something else.

### Recommendations for technical challenges and prerequisites

The findings of this study identified good FAIR practices and approaches within EOSC that address some but not all of the technical challenges of digital preservation. The following list provides a summary of DPC's recommendations for strategies, activities, and possible enhancement to the preservation of research data corresponding to the seven actions areas relating to technical challenges and prerequisites for EOSC:

#### 1. Persistent Identifiers.

- Continued collaboration on developing PID policy, working with DOI and others to address preservation for mature PID infrastructures
- Resources and services within EOSC supporting versioning and tracking through the data lifecycle, including tools and standards supporting machine-actionable, preservation specifications/schemata for kernel information to support machine-actionable PIDs

#### 2. Metadata and Ontologies

- Continued work in coordinating and contributing to activities around metadata and ontologies outside of EOSC including in RDA and the Digital Preservation Coalition–
- Provide or embrace/stimulate existing preservation standards and registries of metadata schemas and ontologies including PREMIS, DPV and METS
- Make sure that the EOSC guidelines for minimum metadata description for resource discovery are supported by relevant and essential preservation metadata including elements for provenance and fixity, as well as sufficient representation information necessary for the creation of archival information packages.

### 3. FAIR metrics and certification

- Provide iterative metrics and assessments as EOSC and FAIR requirements developed to measure and direct the improvement of preservation capability
- Establish an initiative to support the assessment and improvement of the RDA FAIR Data Maturity model:
  - Map approaches maturity modelling within EOSC so they are consistent with preservation maturity models such as DPC Rapid Assessment Model (RAM).
  - Contributing to ongoing CoreTrustSeal+FAIR preservation work for alignment of repository certification schemas with FAIR
  - Test FAIR metrics in digital preservation and research data management communities
- Similarly, convene experts for a preservation task group to ensure key components of collaboration are sustained.
- A Digital Preservation Committee of the EOSC Board (rather than time-limited Task Force) to establish responsibility for assessment frameworks are updated and maintained over time, and to manage ongoing preservation risks.

### 4. Authentication and authorization infrastructure

- Continued work concerning AAI and long-term access to FAIR data objects provided by commercial or semi-commercial e-science service providers
- Address relevant preservation requirements for the design of security and access control protocols established in the short to medium term. For example, encryption keys may need to be escrowed in the event of loss; and security patching may impact the reproducibility of applications. These are relatively modest requirements but if overlooked, will constrain the operational competence of EOSC.

### 5. User environments

- Composing and providing resources for a digital preservation user environment with researchers in mind. While researchers are recognized as the primary users of EOSC, there appears to be the tendency for researchers to be naturally focused on research data in the present (not thinking about preserving in the long term beyond a checklist of preservation requirements in data management planning). For this reason, a digital preservation user environment needs to be made visible within EOSC.
- Further research and projects investigating digital preservation use cases within EOSC

### 6. Resource provider environments

- The ARCHIVER project has undertaken valuable work, and similar work into assessment and procurement of archiving and preservation services should continue between service providers and their clients to optimize and where possible lower the barriers to procurement on both sides.
- Build community and professional networks as providers in community-owned 'marketplace' for support activities with advice and resources. This would enhance the EOSC Portal to ensure more efficient onboarding of resources and integration with existing research (management and preservation) community catalogues and repositories.
- Furthermore, the establishment of a Digital Preservation Working Group can identify where EOSC can provide support for the maintenance of repositories of semantic artefacts, and governance frameworks for such repositories, taking into account common practices and stages of semantic resource development and usage of different communities.

### 7. EOSC interoperability framework

- Use open specifications

- Enable discovery of preservation resources and provide open access and tools to integrate across communities
- Provide tools for quality validation of metadata records and content of digital objects
- Improve search tools for more general-purpose and domain-specific metadata
- Map interoperability efforts explicitly as a form and basis for preservation rather than a simply advanced form of resource description or discovery.

## Recommendations for social, financial, legal, educational, cultural challenges and prerequisites

The findings of this study identified substantial gaps and raised concerns relating to the organizational, social, and cultural changes needed to support the people behind preservation in and across EOSC. Accordingly, the following list summarizes DPC's recommendations for strategies, activities, and possible enhancement to the preservation of research data corresponding to the seven actions areas relating to technical challenges and prerequisites for EOSC:

### 8. Rules of Participation

- Make digital preservation explicit in governance, oversight and authority. Incorporate roles, responsibilities and accountabilities for digital preservation where relevant in policy and framework of stakeholders
- Appoint a senior digital preservation risk owner at board level to manage down the emergent digital preservation risks within EOSC
- Ensure that accountability for digital preservation is agreed in Rules of Participation and is adhered to through periodic checks to provide assurance and trust in the resources and services produced

### 9. Landscape Monitoring

- Connect and map preservation goal, aims, and imperatives with key stakeholders within the EOSC landscape where roles and responsibilities framework can be integrated for a full set of actors and actions. Elaborate process and responsibilities (recognizing that EOSC implementation must be driven bottom-up to meet the varying needs of the different stakeholder communities)
- Digital preservation requires a form of technology watch as well as monitoring of changing community needs. These should be explicitly included within the risk register of the EOSC Association.

### 10. Funding Models

- Develop and provide a business case factory or service for preservation cost modelling that aligns with current EOSC funding programmes, tools, institutions, businesses and bodies contributing to EOSC, and opens opportunities for introducing external funding streams.
- Further in-depth studies and piloting are needed with continued sustainability initiatives and projects (Sustainability Working Group) to gather data on costings for preservation and accounting services' operational costs. Undertake a labour market analysis of digital preservation and research data management professions.

### 11. Skills and Training

- A concerted effort in skills and capacity development is needed to build and support critical Digital Preservation skills, to have these skills specifically addressed with training materials and opportunities.
- Aligning with the priorities and aims for the Skills & Training WG –to develop the next generation of Open Science and data professionals—coordinating training and curricula to include data preservation skills and development

- Recognize and highlight digital preservation professional paths, and maturity models for skills assessment such as DPC's Rapid Assessment Model (RAM) and the National Digital Stewardship Alliance (NDSA) Levels of Preservation
- Create/designate a Digital Preservation Skills & Training advocate for the EOSC Knowledge/Education Hub for knowledge sharing, developed in close cooperation with universities and higher education centres

## 12. Rewards and Recognition

- EOSC should establish explicit goals for the re-use of data and provide professional recognition in the form of case studies of good practice which demonstrate the benefits of re-using data, and the value of the steps that make this possible. EOSC could sponsor such an activity in the context of the Digital Preservation Awards or another relevant framework.

## 13. Communication

- EOSC should participate actively and directly with the global digital preservation community, engaging in two way communication and mutually supportive networking, especially with those sectors of the European economy not aligned with research or research data management.
- The digital preservation community should likewise make an effort to track and facilitate learning from EOSC.

## 14. Widening to the private and public sectors and going global

- Join and build international networks of digital preservation in and outside of Europe to bring in researchers and research performing organizations in private and public sectors. For example, EOSC could contribute to World Digital Preservation Day.

## 4.3. Summary Recommendations and Roadmap

The initial study design, approved for funding in August 2020, proposed seven areas for recommendations. This study makes nineteen recommendations for action, presented below, arranged by the seven areas and tabulated with respect to owners of the recommendation to offer particular responsibilities and accountabilities corresponding to each. Each is assigned one of three priority levels based on its importance and urgency:

- Urgent: of highest importance and urgency (for immediate action)
- High: of high importance and urgency (as soon as possible)
- Medium: of moderate importance or urgency (when feasible)

## Strategies, activities and possible enhancements to the preservation of research data

1. For the EOSC Secretariat: of urgent priority, establish a working party or task group, reporting directly to the EOSC Association Board with respect to digital preservation
2. For the EOSC Secretariat: of high priority, formalize terms of reference and host an initial meeting of a digital preservation task group to establish an iterative work plan
3. For the EOSC Secretariat: of medium priority, establish an operational basis for partnership to deliver the candidate model services proposed in this report:
  - A legacy code or software preservation service
  - A mechanism to ensure accountability and implementation of preservation in data management plans
  - A business case factory or service for preservation cost modelling
  - A programme to support researchers with preservation at the point of creation
  - A mechanism for digital preservation policy across institutions within EOSC



4. For Research Repositories: of urgent priority, adapt workplans to include quality improvement mechanisms where these do not already exist, including DPC Rapid Assessment Model, establishing thereby a strategic framework to achieve baseline certification for primary preservation services, or identifying preservation pathways for data

### Emergent roles and responsibilities for the preservation of digital research data

5. For the EOSC Association Board: of urgent priority, designate a Senior Digital Preservation rapporteur on behalf of the Board to directly communicate and liaison with a Digital Preservation Task Group to monitor and oversee EOSC's responses to digital preservation risks
6. For Funders: of urgent priority, articulate to all grant holders the clear view that adherence to FAIR principles requires data to be monitored and preservation issues to be addressed over the entire life of a project not simply at the point of project completion
7. For Funders: of high priority, audit preservation pathways for all research outputs to identify critically endangered content
8. For Funders: of high priority, initiate a process to establish accountabilities and obligations with respect to implementation of data management plans
9. For Funders: of medium priority, establish mechanisms to engage expert communities of practice in the validation of data management plans
10. For Research Repositories: of medium priority, provide strategic framework for audit of data management plans

### Implication and intersection of FAIR principles on long term preservation of research data and data management planning

11. For EOSC Secretariat: of medium priority, establish mechanism to align EOSC implementation and interpretation of FAIR with the path-dependent and continuous quality improvement cycles of digital preservation
12. For the Digital Preservation Community: of urgent priority, provide a place for EOSC to share lessons and articulate emerging requirements outwith the research data 'bubble'

### Costing parameters and funding models for the preservation of research data

13. For EOSC Secretariat: of medium priority, establish and verify business models for preservation services
14. For Research Repositories: of medium priority, identify costs of action versus inaction with respect to high value, critically endangered content
15. For Funders: of medium priority, establish and identify costs of action versus inaction with respect to high value, critically endangered content

### Training and dissemination needs to promote and sustain a culture of data preservation

16. For EOSC Secretariat: of high priority, establish an ongoing basis for partnership in the digital preservation community, expanding beyond the research data community
17. For Funders: of medium priority, establish more sustained digital preservation training for researchers and repository managers

### Potential benefits and risks of EOSC data preservation services

18. For EOSC Association Board: of high priority, obtain strategic control of digital preservation risks to EOSC

19. For EOSC Association Board: of medium priority, establish a strategic trajectory for management of digital preservation risks, embedding these within reviews and enhancements.

Table 1. Recommendations Tabulated by Owners

For the EOSC Secretariat	<b>Recommendation One:</b> of urgent priority, establish a working party or task group, reporting directly to the EOSC Association Board with respect to digital preservation.
	<b>Recommendation Two:</b> of high priority, formalize terms of reference and host an initial meeting of a digital preservation task group to establish an iterative work plan.
	<b>Recommendation Three:</b> of medium priority, establish an operational basis for partnership to deliver the candidate model services proposed in this report: <ul style="list-style-type: none"> <li>• A legacy code or software preservation service</li> <li>• A mechanism to ensure accountability and implementation of preservation in DMPs</li> <li>• A business case factory or service for preservation cost modelling</li> <li>• A programme to support researchers with preservation at the point of creation</li> <li>• A mechanism for digital preservation policy across institutions within EOSC</li> </ul>
	<b>Recommendation Eleven:</b> of medium priority, establish a mechanism to align EOSC implementation and interpretation of 'FAIR' with the path dependent and continuous quality improvement cycles of digital preservation.
	<b>Recommendation Thirteen:</b> of medium priority, establish and verify business models for preservation services.
	<b>Recommendation Sixteen:</b> of high priority, establish an ongoing basis for partnership in the digital preservation community, including beyond the research data community.
For the EOSC Association Board	<b>Recommendation Five:</b> of urgent priority, designate a Senior Digital Preservation Rapporteur on behalf of the Board to directly communicate and liaison with a Digital Preservation Task Group, to monitor and oversee EOSC's responses to digital preservation risks.
	<b>Recommendation Eighteen:</b> of high priority, obtain strategic control of digital preservation risks to EOSC.
	<b>Recommendation Nineteen:</b> of medium priority, establish a strategic trajectory for management of digital preservation risks, embedding these within reviews and enhancements.
For Funders	<b>Recommendation Six:</b> of urgent priority, articulate to all grant holders the clear view that adherence to FAIR principles requires preservation actions to be monitored and managed over the entire life of a project not simply at the point of completion.
	<b>Recommendation Seven:</b> of high priority, audit preservation pathways for all research outputs to identify critically endangered content.
	<b>Recommendation Eight:</b> of high priority, initiate a process to establish accountabilities and obligations with respect to implementation of data management plans.

	<p><b>Recommendation Nine:</b> of medium priority, establish mechanisms to engage expert communities of practice in the validation of data management plans.</p>
	<p><b>Recommendation Fifteen:</b> of medium priority, identify costs of action versus inaction with respect to high value, critically endangered content.</p>
	<p><b>Recommendation Seventeen:</b> of medium priority, establish more sustained digital preservation training for researchers and repository managers.</p>
<p>For Research Repositories</p>	<p><b>Recommendation Four:</b> of urgent priority, adapt workplans to include quality improvement mechanisms where these do not already exist, including DPC Rapid Assessment Model, establishing thereby a strategic framework to achieve baseline certification for primary preservation services, or identifying preservation pathways for data.</p>
	<p><b>Recommendation Ten:</b> of medium priority, provide strategic framework for audit of data management plans.</p>
	<p><b>Recommendation Fourteen:</b> of medium priority, identify costs of action versus inaction with respect to high value, critically endangered content.</p>
<p>For the Digital Preservation Community</p>	<p><b>Recommendation Twelve:</b> of urgent priority, provide a place for EOSC to share lessons and articulate emerging requirements outwith the research data 'bubble'.</p>

## Appendices

### Appendix A: Interview Guide



#### **FAIR Forever: Interview Guide for Participants**

Thank for your participation in the FAIR Forever project. The purpose of this interview is to ask some questions about your thoughts and experiences with FAIR, EOSC, and the management of research data as they relate to digital preservation. We define digital preservation broadly to include all the managed activities necessary to ensure access to digital materials for as long as necessary, including changes in technology, policy and user requirements. We hope to use the information gathered from the discussion to help assess the current strengths, weaknesses, opportunities and threats to the preservation of research data across EOSC, and determine the feasibility of establishing shared approaches, workflows and services that would benefit EOSC stakeholders.

The interview should take approximately an hour, and this interview guide outlines the main areas and questions we would like to cover. The guide functions mainly as a means to return to the pre-determined topics as the need arises. It is also important to note that most of the questions are open-ended; they are intended to encourage flexibility and allow other relevant ideas around these topic areas to emerge during discussion, so every question below does not need to be answered. If for any reason you are not comfortable with certain questions or topics, you are welcome to suggest alternatives or omit them from the discussion.

Subject to your agreement, we would like to record our interview so that we don't miss any important information and make sure that our notes are accurate. This recording would be kept confidential and deleted upon completion of the project. Should you have any questions or prefer that it not be recorded, please do let us know.

#### **1. Background**

- *Tell me a little about the [Working Group/Research Infrastructure/Research Organizations/Research Performing Institution/etc]*
- *How did you first get involved? Tell me a little about your responsibilities here (if you could sum your role/job up in a sentence).*

#### **2. Data and Digital Preservation**

- *What kind of data do you work with?*
- *What about tools and software?*
- *Are there any challenges managing and sharing that data/tool/software?*
- *Do you offer any guidance in terms of data creation, management, planning or practices?*
- *What is digital preservation? What is your understanding of digital preservation (in theory and/or practice)?*

- *How do you track and assess changing user needs and expectations?*
- *When does/should it occur? Do you use a lifecycle model? Approximately how many years before the data is moved to an archive (and/or managed as semi-current)?*
- *Who is involved (or should be involved)?*
- *Do you have a digital preservation plan? Is it incorporated into another plan or documents?*

### **3. FAIR Principles**

- *Where and how does FAIR come into play? How do you apply FAIR principles to preservation activities?*
- *Are the FAIR principles easy to incorporate and implement?*
- *Where are they challenging?*

### **4. Funding**

- *What are the current funding models for digital preservation?*
- *What are the plans for privately funded projects' data? Could you give any examples? Are these more of the norm or an exception?*
- *How do/would/should you acquire funding for the continued long-term preservation of the research data?*
- *Have you done any cost modelling for digital preservation?*
- *Would you be interested in following up?*
- *What are your positions with respect to carbon costs and footprint?*

### **5. Strengths and Gaps**

- *Have you undertaken any formal or informal maturity modelling for digital preservation capability?*
- *What are the strengths of current staffing and resources for digital preservation?*
- *What are the skills gaps?*
- *What are the training and dissemination needs?*
- *What, if any, are the long-term plans after the projects end? Is there any disaster/exit plans? Are there any concerns about sustainability?*
- *Would you be interested in following up on maturity modelling?*

### **6. EOSC**

- *What is your understanding of EOSC? How does its vision connect with your goals and those of your organization/institution/group?*
- *Why should there be EOSC data preservation services?*
- *What are the benefits and risks to EOSC data preservation services (regarding data, reputation, sustainability)?*

## Appendix B: Interview Participants

Table 2. List of Interview Participants

Interview Date	Participant ID	Org/Inst	Country	EOSC RI Cluster/Region
26/10/20	1	University of Helsinki	Finland	ENVRI-FAIR
26/10/20	2	BGS	UK	ENVRI-FAIR
12/10/20	3	CLARIN ERIC	The Netherlands	SSHOC
09/10/20	4	ASTRON	The Netherlands	ESCAPE
09/10/20	5	ASTRON	The Netherlands	ESCAPE
09/10/20	6	CERN	Switzerland	ESCAPE
12/10/20	7	ESRF	France	PaNOSC
12/10/20	8	CERIC	Italy	PaNOSC
06/10/20	9	RCUB	Serbia	NI4OS
06/10/20	10	Serbian Academy of Sciences and Arts	Serbia	NI4OS
02/11/20	11	ELIXIR	UK	EOSC-LIFE
02/11/20	12	ELIXIR	Spain	EOSC-LIFE

## Appendix C: Focus Group Participants

Table 3. List of Use Case Session Participants

Focus Group Date	Participant ID	Org/Inst	Country	Recruitment
20/11/2020	13	TU Delft / 4TU.ResearchData	The Netherlands	referral/snowball
20/11/2020	14	DRI / Royal Irish Academy	Ireland	direct
20/11/2020	15	Essex	UK	direct
20/11/2020	16	Oxford Bodleian	UK	direct
20/11/2020	17	Loughborough	UK	direct
20/11/2020	18	St Andrews	UK	direct
20/11/2020	19	St Andrews	UK	referral/snowball
20/11/2020	20	Oxford Bodleian	UK	direct
20/11/2020	21	DRI / Royal Irish Academy	Ireland	direct
20/11/2020	22	ADS / University of York	UK	direct
25/11/2020	23	University of Cambridge	UK	referral/snowball
02/12/2020	24	University of Cambridge	UK	direct
25/11/2020	25	University of Cambridge	UK	referral/snowball
26/11/2020	26	UAL	UK	referral/snowball
26/11/2020	27	UAL	UK	direct

## Appendix D: Focus Group Survey Responses

Table 4. Levels of Digital Preservation Capabilities

In general, where would you place your organization in terms of digital preservation capabilities?	Count
0 - minimal: there is minimal awareness of the need to support digital preservation activities	0
1 - awareness: the organization is aware of the need to support digital preservation activities	3
2 - basic: digital preservation activities are supported and resourced at a basic level	5
3 - managed: digital preservation activities are managed, supported, and resourced	3
4 - optimized: digital preservation activities are proactively managed, enhanced and developed	4
NA - I am not sure or my work does not involve digital preservation	0

Table 5. Types of Organizations Represented

What type of organization do you work for?	Count
Higher Education Institution	13
Library or Archive	11
Research Infrastructures or Clusters	2
National or Federal Institution	2
Public Research Institute / Centre	2
Company / Private	1
Governmental Agency	0
Other	0

Table 6. Involvement in EOSC Projects or Initiatives

Have you been involved in any EOSC projects or initiatives?	Count
Perhaps but unsure	0
No	9
Other (a project/initiative that does not fall in other categories e.g. ARCHIVER)	3
Regional (NI4OS-Europe, EOSC-Nordic, EOSC-Pillar, EOSC-Synergy, ExPaNDS)	1
co-creation projects or activities (Flexible Semantic Mapping Framework, HARDOCS, Terms4FAIRskills)	1
EOSC working groups	1
Thematic / Cluster (ENVRI-FAIR, EOSC-Life, PaNOSC, ESCAPE, SSHOC)	2
e-Infrastructure projects (e.g. OpenAIRE, FAIRsFAIR, EOSC-hub, FREYA)	4

Table 7. Identified EOSC Target User Groups

If you had to place yourself in one of these EOSC user groups it would be	Count
Data Curator	3

Data RI Support Professional	7
Data Steward	2
Policy Maker	2
Researcher	1

Table 8. Identified Persona Type

If you had to place yourself in one 'persona', which would it be?	Count
Executive	1
Manager	7
Practitioner	6
Technician	1

Table 9. Job Titles

Job Title	Count
Digital Archives Manager	1
Digital Preservation Manager	1
Digital Preservation Officer	1
Director	1
Head of Digital Preservation	1
International Projects Manager	1
Open Research Systems Manager	1
Repository and Preservation Manager	1
Repository Manager	1
Research Coordinator	1
Research Data Coordinator	1
Research Data Manager	3
Head of Innovation	1

Table 10. Current Responsibility for Digital Preservation at the Organization

Based on your knowledge and experience, who typically takes on the most responsibility for digital preservation at your organization?	Count
It really is an equally distributed team effort	6
Manager	2
Practitioner	2
Technician	1
(blank)	4

Table 11. Biggest Obstacle to Preserving Research Data

The biggest obstacle to preserving research data is:	Count
Funding	2



Know-how	3
Policy	3
Something else	4
Technology	3

Table 12. Making Digital Preservation Explicit in EOSC Vision

Based on your knowledge of EOSC, do you think digital preservation should be made more explicit in the EOSC vision?	Count
Yes	13
(blank)	2

Table 13. Accountability for Digital Preservation at the Organization

Who should be held the most accountable for digital preservation (if/when there is a loss of data, who should answer for it)?	Count
Executive	3
Manager	1
Other	2
Unsure (accountability is unclear)	3
(blank)	6

Table 14. Current Funding of Digital Preservation at Organization

How are digital preservation activities at your organization currently funded?
Central budget; Project funding centrally
Centrally by Institution; Outside funding
Funding Bodies
Minimally Core budget
Projects; Some central; Library operations
Strategic funding; Project funding; Core budgets

Table 15. Recommended Funding for Digital Preservation Resources in EOSC

Who should provide funding for digital preservation resources in EOSC?
EOSC itself; funders
Everyone:)
Funders
Funders
Should be funded as part of research grants and by institutions
European Commission
Don't know

Institutions
--------------

*Table 16. Recommended Funding for Digital Preservation Services in EOSC*

<b>Who should provide funding for digital preservation services in EOSC?</b>
Everyone
Funders
Funders
EOSC is a federation of services, so the service that join need to be self-sufficient.
European Commission
Don't know
EOSC members