

Long-Term Data Retention Task Force Terms of Reference

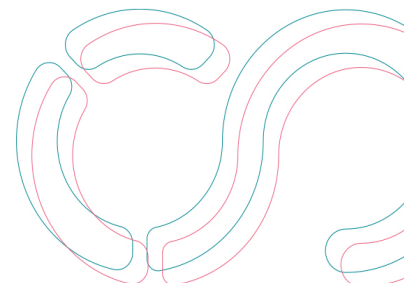
1 TF Background and Scope

The EOSC-A Long Term Data Preservation Task Force worked on defining the vision on Long Term Data Preservation, proposing the adoption and application of clear 'levels of care' (including the differentiation of initial curation from active, long-term preservation), and providing a set of assertions and recommendations for preservation in the context of EOSC. These assertions and recommendations included the need for additional work on retention, appraisal, and re-appraisal of digital objects as a continuous process to evaluate a digital object with a view to retaining it, assigning or changing a level of care, or deleting it. A clear and defined (re)appraisal cycle is a dependency for informed and transparent decisions around the proper level of care. The growth in data production through research and of relevance to research, with an associated increase in costs and demand for limited storage capacity, implies a need for clearer criteria around what should be kept and cared for, including through long-term preservation.

The Long-Term Data Preservation Task Force defined the following vision for optimal preservation of FAIR Digital Objects within the context of EOSC:

- Digital objects that act as inputs to, or outputs from, research are identified, findable and accessible in environments that support good storage practice. These objects are subject to appraisal, and reappraisal over time, to assess their value, their impact and the associated costs, risks and benefits. Ongoing appraisal informs the level of investment in the retention, curation and active long-term preservation of digital objects. The levels of care, and changes to levels of care, provided by repositories and assigned to digital objects are transparent to their funders, depositors and users.

Digital objects are the primary reusable assets of federated research infrastructures and therefore an understanding of the criteria and the decisions that inform retention and the level of care is vital to the future and continued success of EOSC. Topics related to retention and long-term preservation, including data quality and a framework to identify what data are candidates for long-term preservation, are being addressed in the EOSC EDEN Enhancing Digital preservation strategies at European and National level project (HORIZON-INFRA-2024-EOSC-01-04 call). Engagement with the projects that result from EOSC EDEN and other relevant projects (e.g. FIDELIS Establishing A European Network of Trustworthy Digital Repositories HORIZON-INFRA-2024-EOSC-01-03) will be crucial. This Task Force aligns with several objectives of the SRIA related to the co-design and adoption of principles, practices, and frameworks by



communities. Moreover, this TF is aligned with the SRIA's Strategic Objectives SO6 (reuse of European research) and Operational Objective OO1 (EOSC MVE). Additionally, the work of the TF will be fully aligned with the Multi-Annual Roadmap through the liaison to HE projects (in objective 1), fostering the use of FAIR-enabled trustworthy repositories (in objective 2) and long-term preservation (in objective 3).

1.1 Main Objective

The Task Force will engage with the data repository and archives landscape to support an improved understanding of the different scientific communities' status, needs and challenges in data retention, appraisal, and re-appraisal. The work will include discussions on wider typologies and definitions of retention, curation, and preservation, with the objective of reaching a broad understanding and agreement by the EOSC community concerning these definitions. The Task Force will consider recommendations on all forms of retention, curation, and preservation, specifically towards their engagement with the EOSC EDEN and FIDELIS projects. The Task Force will continue to monitor gaps from the projects and formulate advice for new priorities and objectives to be included in the future SRIA and MAR.

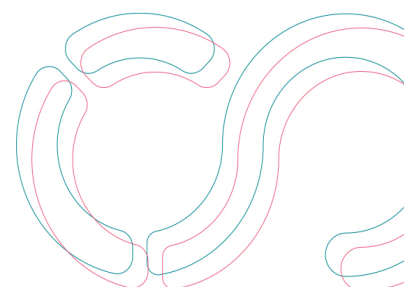
1.2 Key Focus Areas

- **KFA1:** The Task Force will produce a map of the landscape on the status of specific data types' or data communities', needs and challenges in data retention, appraisal, and re-appraisal to make strategic recommendations on continued actions (SRIA, MAR ...). This will include considerations for sensitive data, the CARE and TRUST principles and the recommendations of the previous Task Force.
- **KFA2:** The Task Force will provide feedback and recommendations on retention, appraisal, and re-appraisal of digital objects so that they remain FAIR over time to the EOSC-A Board and to ongoing and starting projects such as EOSC EDEN and Fidelis.

1.3 Expected Membership

Membership is open to a wide range of stakeholders across the digital object management lifecycle but should preferably include:

- Relevant parties from EOSC related projects, specifically EOSC EDEN and FIDELIS projects (HORIZON-INFRA-2024-EOSC-01-03 / -04).
- Other bodies and activities related to the new scope of the TF (national service providers, ERICs/Science Clusters etc).



- The Task Force will welcome input from data curators and data stewards with a diverse representation from organisations (European, National, Institutional), repositories (generalist/specialist), collections/catalogues, and at the digital object type level.
- We encourage active participation from Research Performing Organisations to get a good balance of projects, service providers, and RI representatives.

1.4 List of deliverables

KFA1

- An infographic representing the high-level map of the data retention landscape including the relevant stakeholders, specific data types and data communities. Q4 - 2024
- A report documenting the status of specific data types' or data communities' needs and challenges in data retention, appraisal, and re-appraisal of the stakeholders/communities identified in the map. Q4 - 2025

KFA2

- An expansion of the common standard vocabulary started with the previous Long-Term Data Preservation TF, to include the differentiation of full stack preservation vs. retention to be inclusive of retention, deposit criteria, appraisal/reappraisal, initial curation and active preservation for example. Q3 - 2025
- A conceptual tool to help stakeholders, including repositories, archives, data stewards, etc. to implement a workflow for data retention, appraisal, and re-appraisal. The tool will be developed and tested in collaboration with the EOSC EDEN project and taking into account the feedback collected via the TF representatives and beyond. Q2 - 2025 / Q4 - 2025.

