

Context, approach, and progress of the EOSC-A Long Term Data Preservation Task Force

EOSC Symposium



Objective of the session

The Long-Term Data Preservation Task Force (LTDP TF) is addressing the complex issue of long-term preservation within the EOSC and seeks to propose clear, implementable recommendations at the European, national, and institutional levels.

This session gives a short **overview** on the TF approach on preservation in the context of EOSC as input for an **interactive discussion** with you on the challenges and recommendations to address digital objects' preservation through EOSC.

Agenda:

- Short introduction to the TF
- Open discussion
- Wrap-up

LTDP-TF References

LTDP in the Strategic Research Agenda

“Moreover, long-term open data archives and preservation services are required to enable a **sustainable EOSC and the sustainable access to data**. Data preservation not only refers to the long-term storage of data, but also includes ensuring the preservation and maintenance of data, as well as its context, understandability, interpretability, authenticity and integrity.”

“The interim findings of the FAIR Forever study, conducted by DPC on behalf of the Sustainability Working Group, noted that **digital preservation is not explicit in the context of EOSC** and the roles, responsibilities and accountability for digital preservation are currently not clearly defined. “

“**Clearer roles and responsibilities are needed**, including the assessment of capability as well as functions, salaries and funding streams for preservation.”

Source: SRIA v1.0, 21/06/2021 (https://www.eosc.eu/sites/default/files/SRIA_2022_01.pdf)

LTDP-TF References

FAIR Forever recommendations with urgent priority

Recommendations for the EOSC Secretariat and EOSC-A with urgent priority status:

- Establish a working party or task group, reporting directly to the EOSC Association Board with respect to digital preservation
- 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

Source: FAIR Forever study, conducted by DPC on behalf of the Sustainability Working Group, 17/02/2021

(<https://doi.org/10.5281/zenodo.4574234>)

CoreTrustSeal Curation & Preservation Levels Discussion Paper

<https://doi.org/10.5281/zenodo.6908019>

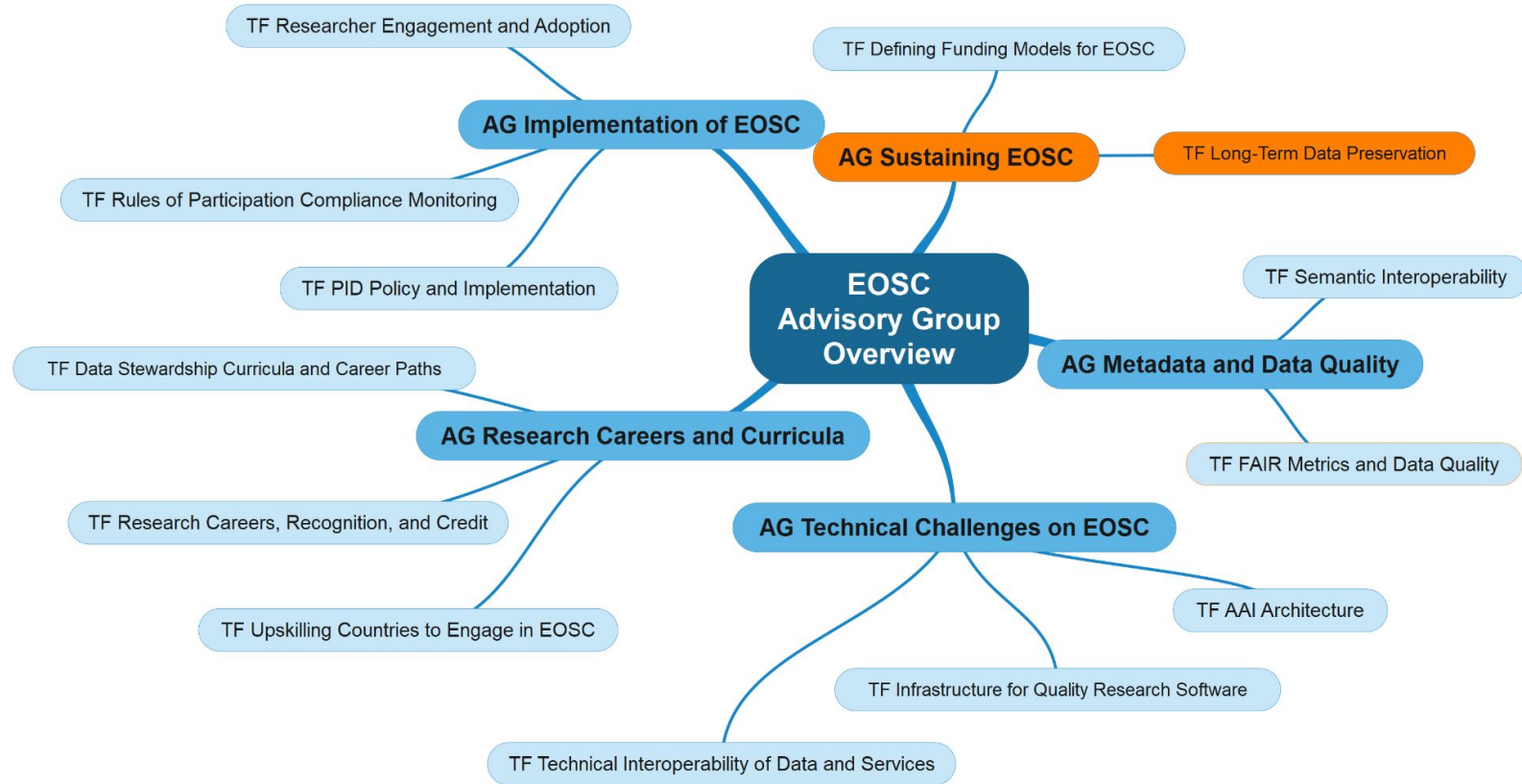
Review of CoreTrustSeal Applicability to non-Preservation (Trustworthy) Data Services

<https://doi.org/10.5281/zenodo.6865991>

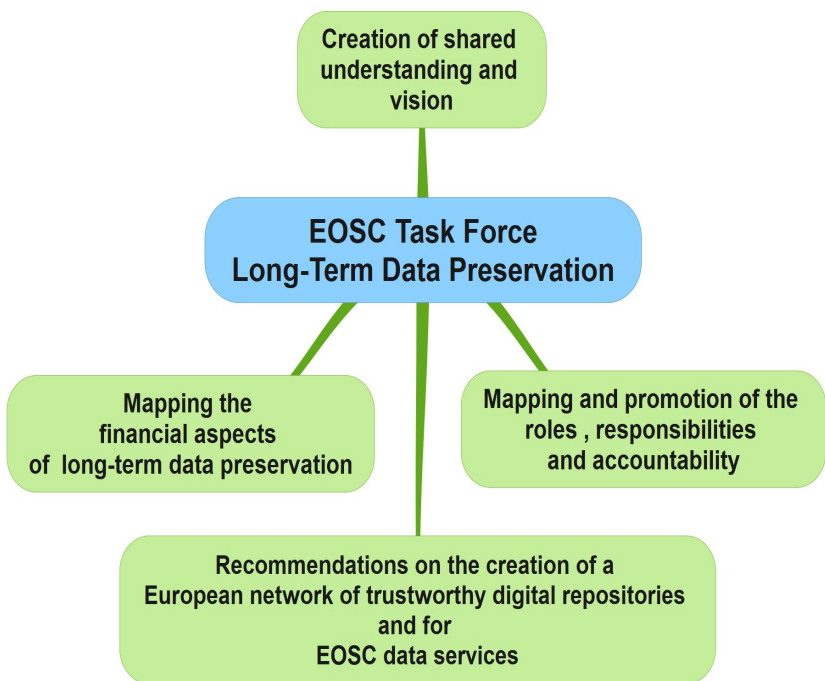
FAIR Forever? Long Term Data Preservation Roles and Responsibilities, Final Report





<https://doi.org/10.5281/zenodo.4574234>

AG Sustaining EOSC – LTDP TF

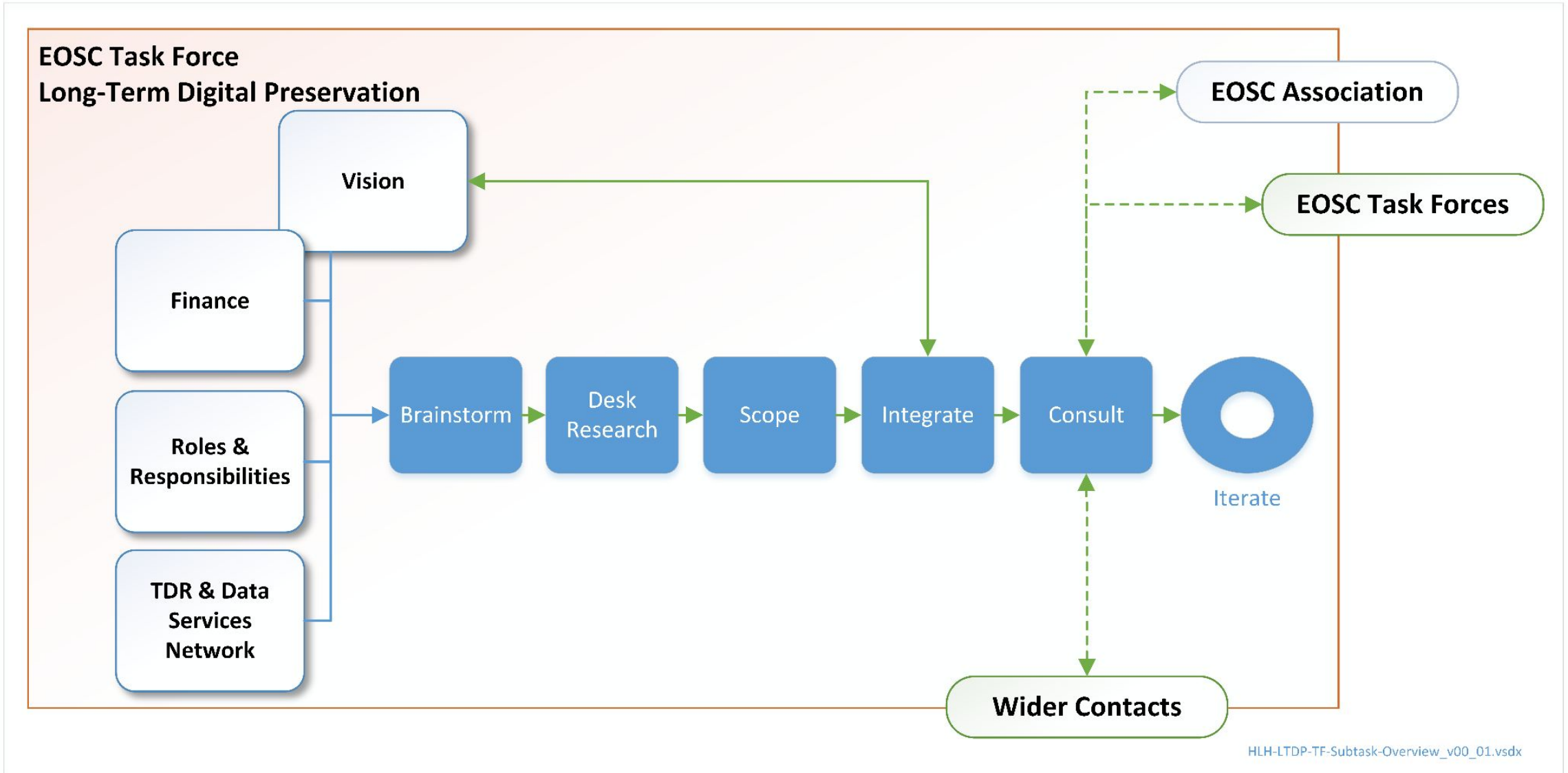


Task force structure



<p>Co-chairs:</p> <ul style="list-style-type: none"> • Hervé L'Hours (UK Data Archive) • Roxanne Wyns (KU Leuven) 	<p>EOSC-A liaison:</p> <ul style="list-style-type: none"> • Bob Jones (BoD) • René Buch (CTO)
<p> Shared Understanding and Vision Subgroup lead: Chris De Loof (Belnet) Pirjo-Leena Forsström (CSC), Rui Fernandes (C4G), Gerardo Ganis (CERN), Ville Tenhunen (EGI), Andras Holl (MTA), Maciej Brzeźniak (PSNC), Lara Lloret Iglesias (CSIC), Riccardo Smareglia (INAF), Eileen Gibney (UCD)</p>	<p> Mapping and promotion of roles Subgroup lead: Mariusz Majdański (IG PAS) Andrea Lammert (DKRZ), Sangeetha Shankar (DLR), Christian Cuciniello (EC), Marcello Maggi (INFN), Florina Piroi (TU WIEN), Mojib Wali (TU GRAZ), Sabine Crépé-Renaudin (CNRS), Bregt Saenen (Science Europe)</p>
<p> Mapping of financial aspects Subgroup lead: Paul Stokes (IISC) Jean-Yves Nief (CC-IN2P3), Jiri Novacek (CEITEC), Toni Andreu (EATRIS), Martina Stockhause (IPCC DDC)</p>	<p> Recommendations TDR Subgroup lead: Didi Lamers (Radboud university) Olesia Dubois (Sciences Po Paris), S. Venkataraman (OpenAire), Pierre-Yves Burgi (OLOS.swiss), Lluís Anglada (CSUC), Cécile Cavet (Univ. Paris), Ingrid Dillo (DANS), David Antos (CESNET), Draženko Celjak (SRCE), Matthew Viljoen (EGI)</p>

Task force approach



LTDP TF Minimalist Lifecycle

EOSC Task Force
Long-Term Digital Preservation
Minimalist Lifecycle Perspective



HLH-LTDP-TF-MinimalLifecycleSimple_v00_01.vsd

Overview discussion paper

Soon open for feedback and consultation

- Overview and discussion paper that presents the context, activities and ongoing issues surrounding the EOSC-A LTDP-TF to be released for public consultation, seeking comments from a wide range of stakeholders across the digital object management lifecycle
- Feedback to be integrated into an updated overview
- Consultation on the recommendations addressing digital objects' preservation through EOSC at the
 - European
 - National
 - Institutional level



LTDP TF: Preservation Perspectives

Outcomes, Systems and Actions

Preservation Outcomes: these concern digital objects that, having been curated for FAIRness and other desirable characteristics, are maintained to retain those characteristics for as long as necessary. These outcomes are targeted at a defined and 'designated community' of users.

Preservation Systems: accept the deposit of digital objects for storage and access (making them 'repositories' in the broad sense) and also curate them with a long term FAIR-enabling perspective on the objects and their designated community (making them candidate to become 'trustworthy digital repositories'). The designated community's knowledge base and technological needs are understood and monitored over time. These systems have sufficient resources, including personnel and financial resources, to be sustainable in terms of their organisational, technology and security infrastructure.

Preservation Actions: are the changes to digital objects' (metadata and data) that are intended to keep them FAIR over time. These actions include technical steps such as emulation or transformations to modern, in-demand, file formats and metadata schemas; but also actions to ensure that the conceptual content of data and metadata, including semantic artefacts such as ontologies, continues to be understood and re-usable.

LTDP TF: Preservation Perspectives

FAIR + TIME

Importance of:

- FAIR (Findable, Accessible, Interoperable and ReUsable) digital objects
- Trusted Repositories and CoreTrustSeal
 - CoreTrustSeal 2020-2022 □ Mapped to FAIR
 - FAIR + Time □ Requires preservation
 - Revised during FAIR IMPACT Project
- Transparency over current levels of FAIRness and trustworthiness (current status, plans for improvement)



- F1. (meta)data are assigned a globally unique and persistent identifier.
- F2. data are described with rich metadata (defined by R1 below).
- F3. metadata clearly and explicitly include the identifier of the data it describes.
- F4. (meta)data are registered or indexed in a searchable resource.

R13. Data discovery and identification

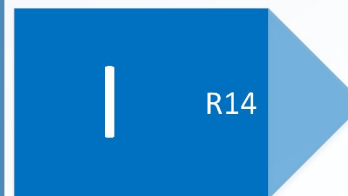


- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
 - A1.1 the protocol is open, free, and universally implementable (vs context)
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

R15. Technical infrastructure

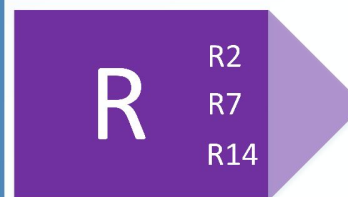
R16. Security

R10. Preservation plan



- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

R14. ReUse



- R1. meta(data) are richly described with a plurality of accurate and relevant attributes.
 - R1.1. (meta)data are released with a clear and accessible data usage license.
 - R1.2. (meta)data are associated with detailed provenance.
 - R1.3. (meta)data meet domain-relevant community standards.

R14. ReUse

R2. Licenses

R7. Data integrity and authenticity

R14. ReUse

Vision in progress

Overview discussion paper

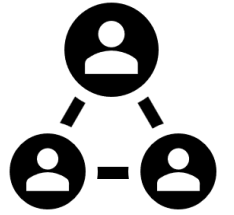


- Ideally, all data of interest to EOSC researchers retain their value and remain FAIR for a defined community for as long as necessary (FAIR+Time)
- But resources are not infinite so a nuanced approach in the vision and direction of the TF needed
- Vision concepts:
 - FAIR
 - Appraisal and reappraisal
 - Retention, curation, long-term preservation
 - Transparency on level of care provided by repositories



Roles

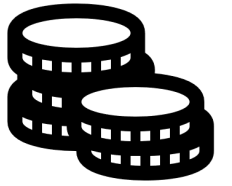
Overview discussion paper



- Gaps in understanding the roles of various actors
- Problems in identifying and financing those with the skills to take responsibility for curation and preservation
- A need for support from specialised training roles that provide general and specialist guidance
- Different roles, responsibilities and accountability on institutional, national and European level
- Additional responsibilities for data creators, technicians, data stewards etc, all incur costs Finance subgroup

FAIR Forever study as the basis for the further identification of roles

Finance



Overview discussion paper

- Examine and collate existing cost models (e.g. curationexchange.org, 4C Project)
- Look at preservation-specific costs vs. other data services costs and define subset of preservation costs
- What are the costs for a community if a digital object of value for reuse is not cared for
- Problem of short term funding
- Resources are not infinite, so appraisal and reappraisal is critical over time Also incurs costs
- Practices will differ between generalist and specialist domain repositories Cooperation for sharing expertise, services and costs
- Balance between potential of automation and maintaining enough resources of human expertise
- Identifying the likely costs of preservation at a sufficiently granular level is challenging. The same goes for calculating the costs of inaction with risks of data loss

Network of Trust

Overview discussion paper

- Vision can only be achieved through an ongoing engagement with curation and preservation professionals
- In need of a body that integrates the expertise of, and provides a voice for trustworthy digital repositories
- Reviewing and providing feedback on relevant initiatives (e.g. Towards a European network of FAIR-enabling Trustworthy Digital Repositories (TDRs) - A Working Paper produced by a group of representatives from the repository community)



August 30, 2022

Working paper Open Access

Towards a European network of FAIR-enabling Trustworthy Digital Repositories (TDRs) - A Working Paper

Philippp Conzett, Ingrid Dillo, Françoise Genova, Natalie Harrower, Vasso Kalaitzi, Mari Kleemola, Amela Kurta, Pedro Principe, Olivier Rouchon, Hannes Thiemann, Maaïke Verburg

This working paper is a bottom-up initiative of a group of stakeholders from the European repository community. Its purpose is to outline an aspirational vision of a European Network of FAIR-enabling Trustworthy Digital Repositories (TDRs). This initiative originates from the workshop entitled "Towards exploring the idea of establishing the Network". The paper was created in close connection with the wider community, as its core was built on community feedback and the first draft of the paper was shared for community-wide consultation. This paper will serve as input for the EOSC Task Force on Long Term Digital Preservation. One of the core activities mentioned in the charter of this Task Force is to produce recommendations on the creation of such a network.

The working paper puts together a vision of how a European network of FAIR-enabling TDRs could be based on the community's needs and its most important functions: Networking and knowledge exchange, stakeholder advocacy and engagement, and coordination and development. The specific activities hosted under these umbrella functions could address the wide range of topics that are important to TDRs. Beyond these functions and the challenges they address, the paper presents a framework to highlight aspects of the Network to further explore in the next steps of its development.

This Working Paper was created with input from the European repository community (17th June - 4th July 2022)



<https://doi.org/10.5281/zenodo.7034315>

LTDP TF Network of TDRs

Towards a European network of FAIR-enabling Trustworthy Digital Repositories

Working paper by stakeholders from:

- FAIRsFAIR <https://www.fairsfair.eu/>
- SSHOC <https://sshopencloud.eu/>
- EOSCNordic <https://www.eosc-nordic.eu/>
- Workshop <https://doi.org/10.5281/zenodo.5849658>

Working paper for public feedback <https://doi.org/10.5281/zenodo.7034314>

LTDP Trust Network subgroup working on a response:

- Supporting, contributing, aligning
- FAIR IMPACT Synchronisation Force as reference point for aligned actors and implementation

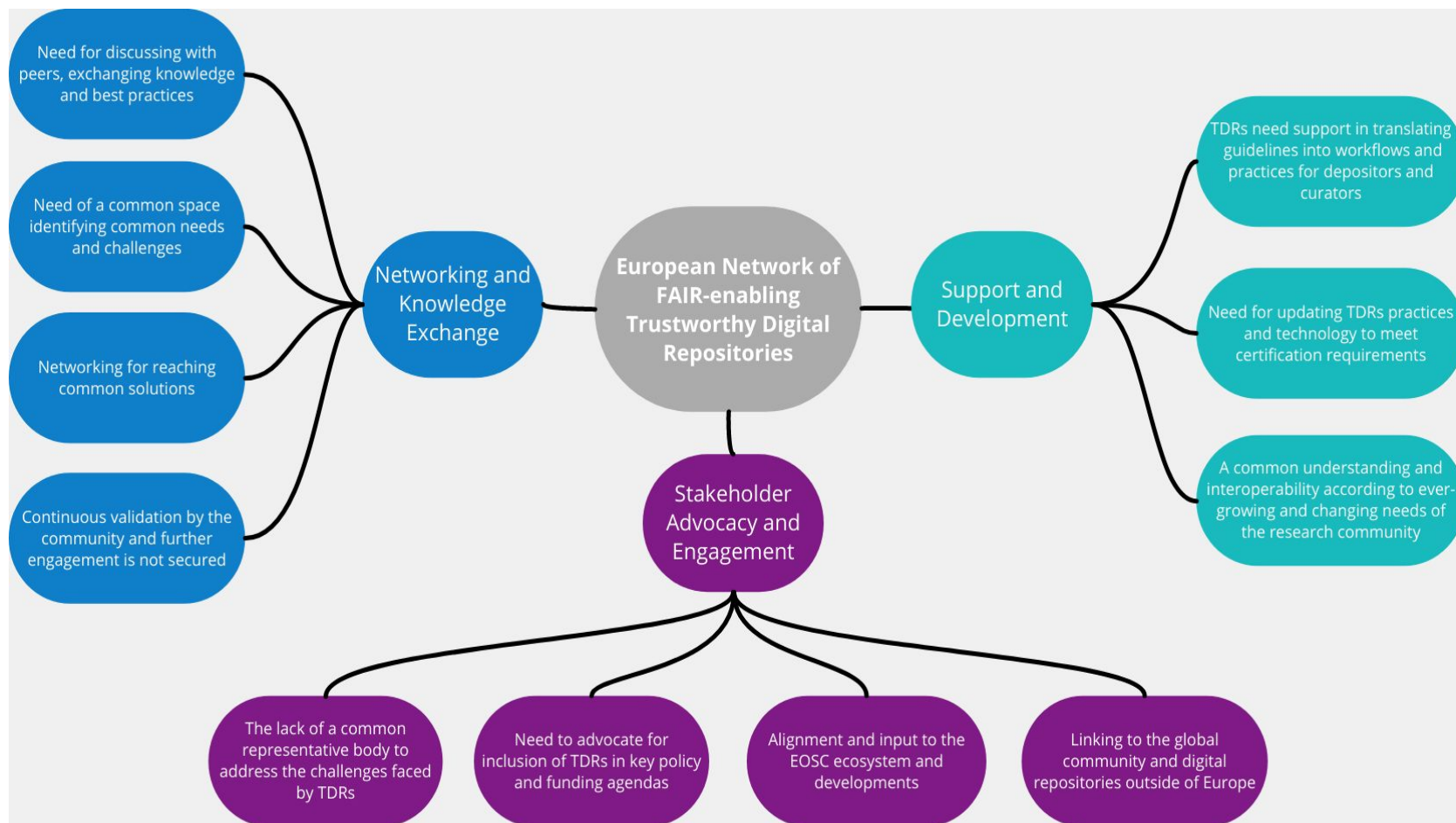


LTDP TF Network of TDRs

Towards a European network of FAIR-enabling Trustworthy Digital Repositories

The network addresses:

- Networking & Knowledge Exchange
- Support & Development
- Stakeholder Advocacy & Engagement



LTDP TF Network of TDRs

Towards a European network of FAIR-enabling Trustworthy Digital Repositories

Why the network is important:

- Beyond limited project timeframes and perspectives
- Empower trustworthy and FAIR-enabling repositories in the EOSC ecosystem.
- A 'voice' for European TDRs within EOSC
- Community led, knowledge exchange, peer to peer support, network spaces
- Continuous validation of standards and practices

Discussion

- **Institutional level**
 - Roles requirements
 - Finance requirements
 - Network of TDRs requirements
- **National level**
 - Roles requirements
 - Finance requirements
 - Network of TDRs requirements
- **European level**
 - Roles requirements
 - Finance requirements
 - Network of TDRs requirements



Discussion - Institutional level

Roles requirements

- Xxx
- Xxx

Finance requirements

- Xxx
- Xxx

Network of TDRs requirements

- Xxx
- Xxx



Discussion - National level

Roles requirements

- Xxx
- Xxx

Finance requirements

- Xxx
- Xxx

Network of TDRs requirements

- Xxx
- Xxx



Discussion - European level

Roles requirements

- Xxx
- Xxx

Finance requirements

- Xxx
- Xxx

Network of TDRs requirements

- Xxx
- Xxx



