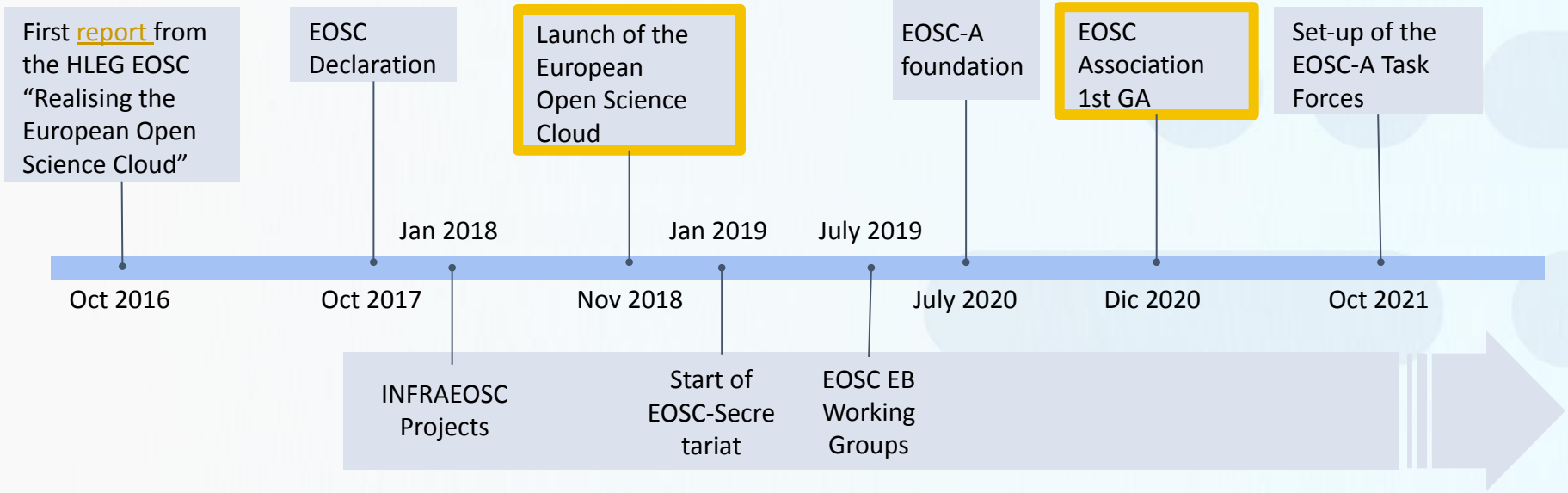
EOSC-A
Overcoming Data Challenge

Ignacio Blanquer
Board of Directors EOSC-A

January the 26th, 2022
Science Mesh Workshop

The European Open Science Cloud (EOSC) initiative

- EOSC aims to offer a virtual environment for open access to services to store, share, process and reuse research data and other research digital objects, such as software
  - The EOSC initiative was proposed in 2016 by the EC as part of the European Cloud initiative, funded through the so-called H2020-INFRAEOSC-2018-2020 (€ 157M).
  - Main objective of HEU destination 2 with 89M€ in the HORIZON-INFRA-2021/22-EOSC-01 calls.



EOSC Executive Board Working Groups

- EOSCsecretariat.eu was a project addressing the set-up of an operational framework supporting the overall governance of the EOSC.
- The EOSC Executive Board set up 6 Working Groups, with some relations to Open Software and interoperability.

Landscape

Mapping of the existing research infrastructures which are candidates to be part of the EOSC federation.

- *Inventory of initiatives related to multiple dimensions of EOSC.*

FAIR

Implementing the FAIR data principles by defining reqs. for cross-disciplinary interoperability.

- *FAIR practices for Data.*

Architecture

Defining the technical framework required to enable and sustain an evolving EOSC federation of systems.

- *EOSC as a federation of systems.*
- *Produced the Interoperability Report*

Rules of participation

Designing the RoP that define the rights & obligations among EOSC users, providers and operators.

- *Software as key asset for Data reuse.*
- *EOSC SW licenses*

Skills & Training

Providing a framework for a sustainable training infrastructure to support EOSC in all its phases and ensure its uptake.

- *Definition of the Research Software Engineer and Data Science role in EOSC.*

Sustainability

Providing a set of recommendations concerning the implementation of an operational, scalable and sustainable EOSC federation.

- *Recommendations for PCP to include Open Software.*

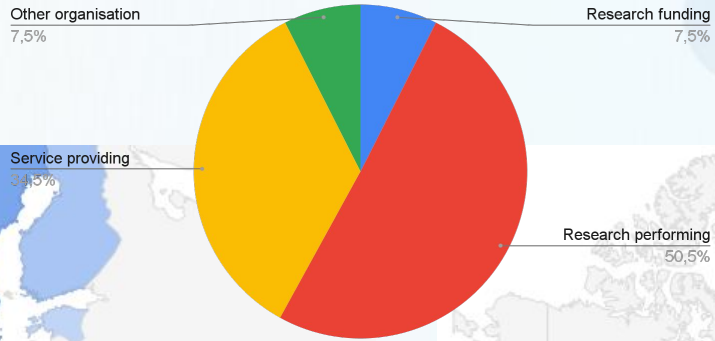
EOSC Governance Model 2021-2027

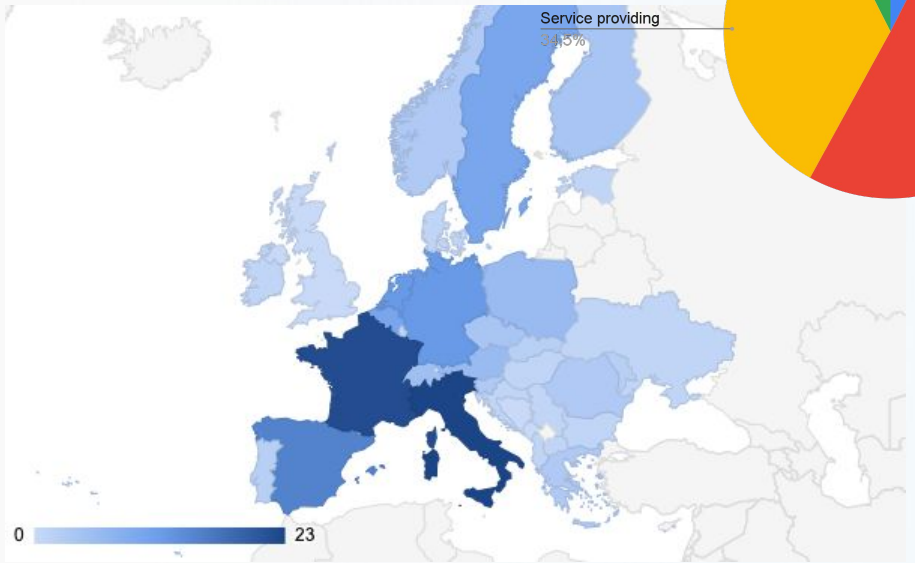
The new governance model agreed with EU countries for the next EOSC implementation phase after 2020 will be tripartite including:

- The EU represented by the **Commission**
- The European research community represented by the **EOSC Association**
- EU countries and countries associated with Horizon Europe represented through a **Steering Board** to be set up in 2021 outside of the EOSC Association

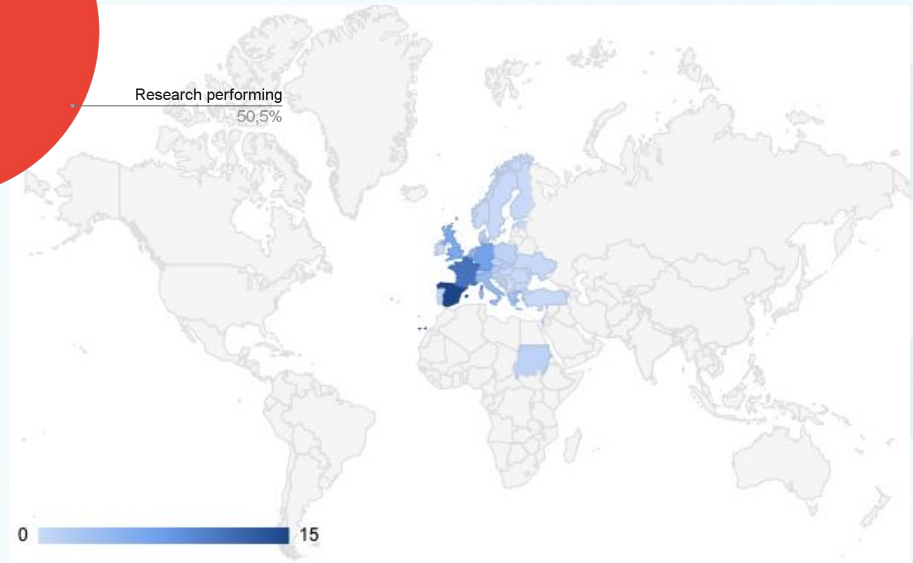
EOSC membership by geographical spread

160 Members





73 Observers



The Strategic Research and Innovation Agenda (SRIA)

- A key document for the implementation of EOSC is the "Strategic Research and Innovation Agenda" (SRIA) .
- It states three General Objectives
  - GO1. Ensure that Open Science practices and skills are rewarded and taught, becoming the 'new normal'.
  - GO2. Enable the definition of standards, and the development of tools and services, to allow researchers to find, access, reuse and combine results.
  - GO3. Establish a sustainable and federated infrastructure enabling open sharing of scientific results.

The Strategic Research and Innovation Agenda (SRIA)

The specific objectives (SOs), which are reflected in the critical success factors include:

...

- SO5 The EOSC Interoperability Framework supports an increasing range and quantity of FAIR digital objects including data, software and other research artefacts;
- SO6 Provide an increased number of services and resources to ensure that European research is discovered and reused within and across disciplines to extract new knowledge;
- SO7 EOSC is operationalised and provides a stable and valuable infrastructure supporting researchers addressing societal challenges;

...

The Strategic Research and Innovation Agenda (SRIA)

The operational objectives (OOs) include:

- OO1 Deliver and operate all the necessary components of the Minimum Viable EOSC to share openly research data, publications, software, tools and services while attracting increasing numbers and categories of users (public and private) … by 2025;
- ...
- OO5 Provide the technical components of a FAIR ecosystem for uptake and customisation by the communities by 2023 (including open specifications, standards, schemas, application programming interfaces (APIs), metadata frameworks supporting FAIR digital objects and their automated processing);
- ...
- OO7 Co-develop a first generation of a robust pan-European network of infrastructures for software source code (including incentives for the effective documentation and sharing of research software) by 2025;

  ...

Working Groups EOSC-A

Implementation of EOSC

- Rules of Participation compliance monitoring
- PID policy and implementation
- Researcher engagement and adoption

Technical challenges on EOSC

- Technical interoperability of data and services
- Infrastructure for quality research software
- AAI Architecture

Metadata and data quality

- Semantic interoperability
- FAIR metrics and data quality

Research careers and curricula

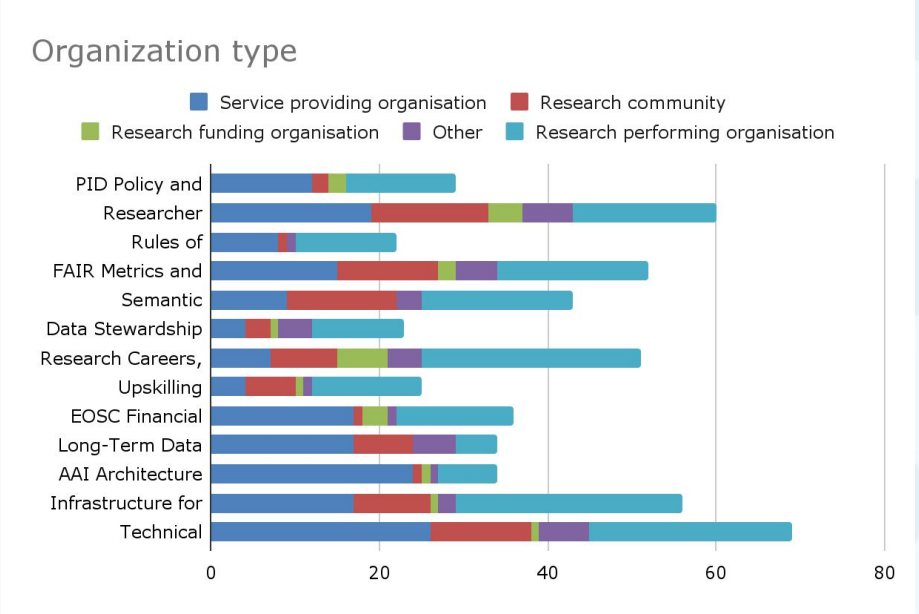
- Data stewardship curricula and career paths
- Research careers, recognition and credit
- Upskilling countries to engage in EOSC

Sustaining EOSC

- Defining funding models for EOSC
- Long-term data preservation

Task Forces activities

- EOSC Association members & observers are the primary members of AGs.
- Around 450 applicants have been selected.
- Representatives of the key implementation projects should be members of relevant AGs to present work in progress and receive advice and steers from the community.
- The ideas and priorities of the EOSC Association Advisory Groups should feed into the Descriptions of Work of the upcoming Horizon Europe projects.
- Draft remits can be found in the web site.



The EOSC-Association Task Force on Technical Interoperability

- Coordinated by Álvaro López (CSIC-ES) and Eva Sciacca (INAF-IT).
- Draft Charter available in the EOSC-A website.
- Objective of the TF
  - The Task Force will take the EOSC Interoperability Framework (EIF) recommendations around technical architecture as their starting point to help develop the EOSC Core and Exchange.
  - The EOSC Core and Exchange will create interoperability whilst preserving a diversity of components, both on the semantic and the technical level.
  - This Task Force will enable the EOSC-Future project, other EOSC projects and the EOSC resource providers to harmonize their approaches to achieve interoperability.

- 4 subgroups defined, addressing:
  - Landscape, Overview and Scouting:.
  - Data technical interoperability:.
  - Services technical interoperability:.
  - Technical Architecture recommendations:

- Five deliverables expected:
  - A first principles document. (May 2022)
  - A landscape overview (capabilities and gaps) of the EIF. (July 22-July 23)
  - A draft technical architecture description of the EIF. (Nov 2022)
  - A technical architecture description of the EIF, including examples of adaptation hints for major existing solutions. (Nov 2023)

Conclusions

- EOSC aims at fostering the development of Open Science by enriching publications, data and software according to the FAIR principles in order to make them usable by machines and scientists and Federating infrastructures to do better and more rewarded research.
- EOSC-A has created several Task Forces focusing on Interoperability of data, software and services improving the reuse and sustainability of research assets.
- Important links:
  - Newsletter: https://eosc.eu/newsletter
  - Join the EOSC-A: https://www.eosc.eu/join-association
  - Contact: info@eosc.eu
  - Twitter: @eoscassociation

EUROPEAN OPEN SCIENCE CLOUD

THANK YOU

@eoscassociation

info@eosc.eu

EOSC and ScienceMesh - Overcoming data challenges:

EOSC Long-Term Data Preservation Task Force (LTDP TF)

Pierre-Yves Burgi

Deputy CIO at the University of Geneva

Member of the EOSC LTDP TF

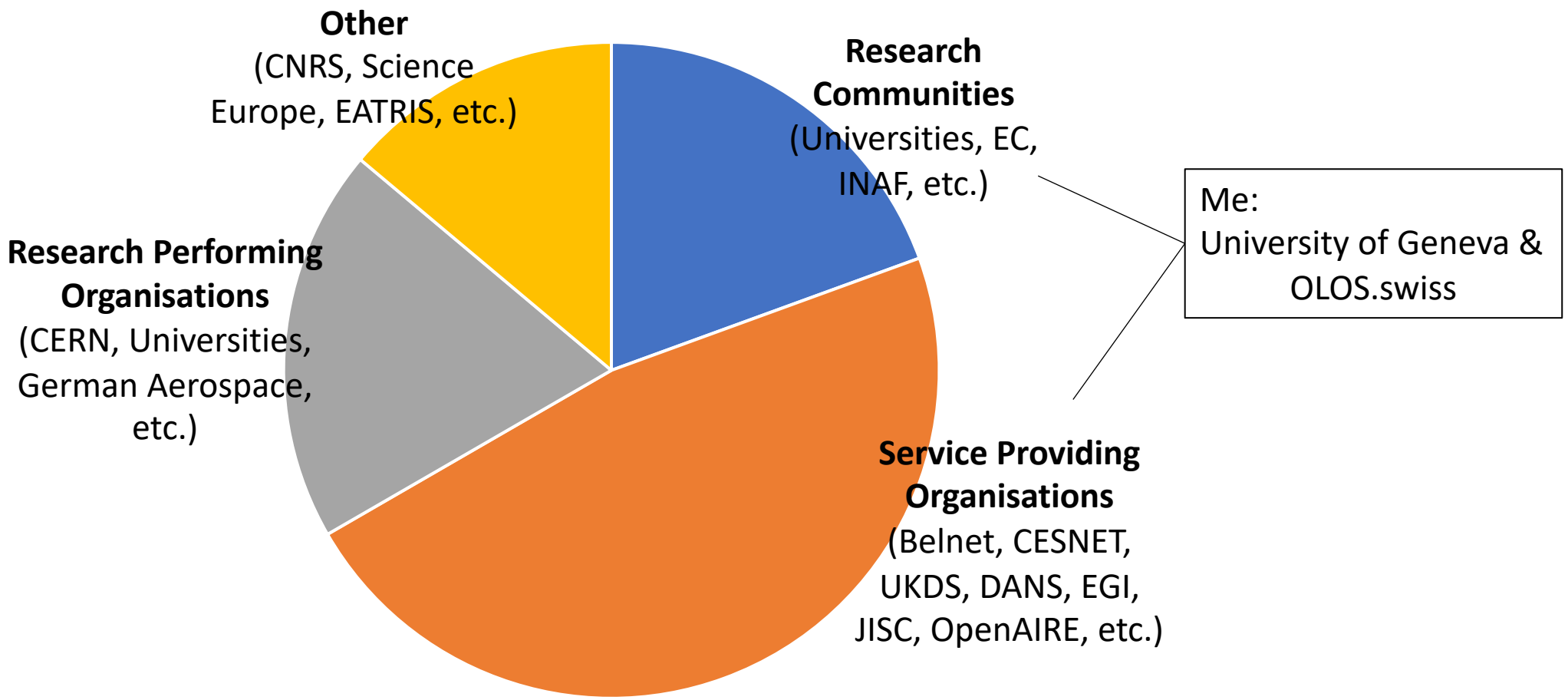
Pierre-Yves.burgi@unige.ch

LTDP TF Aims at

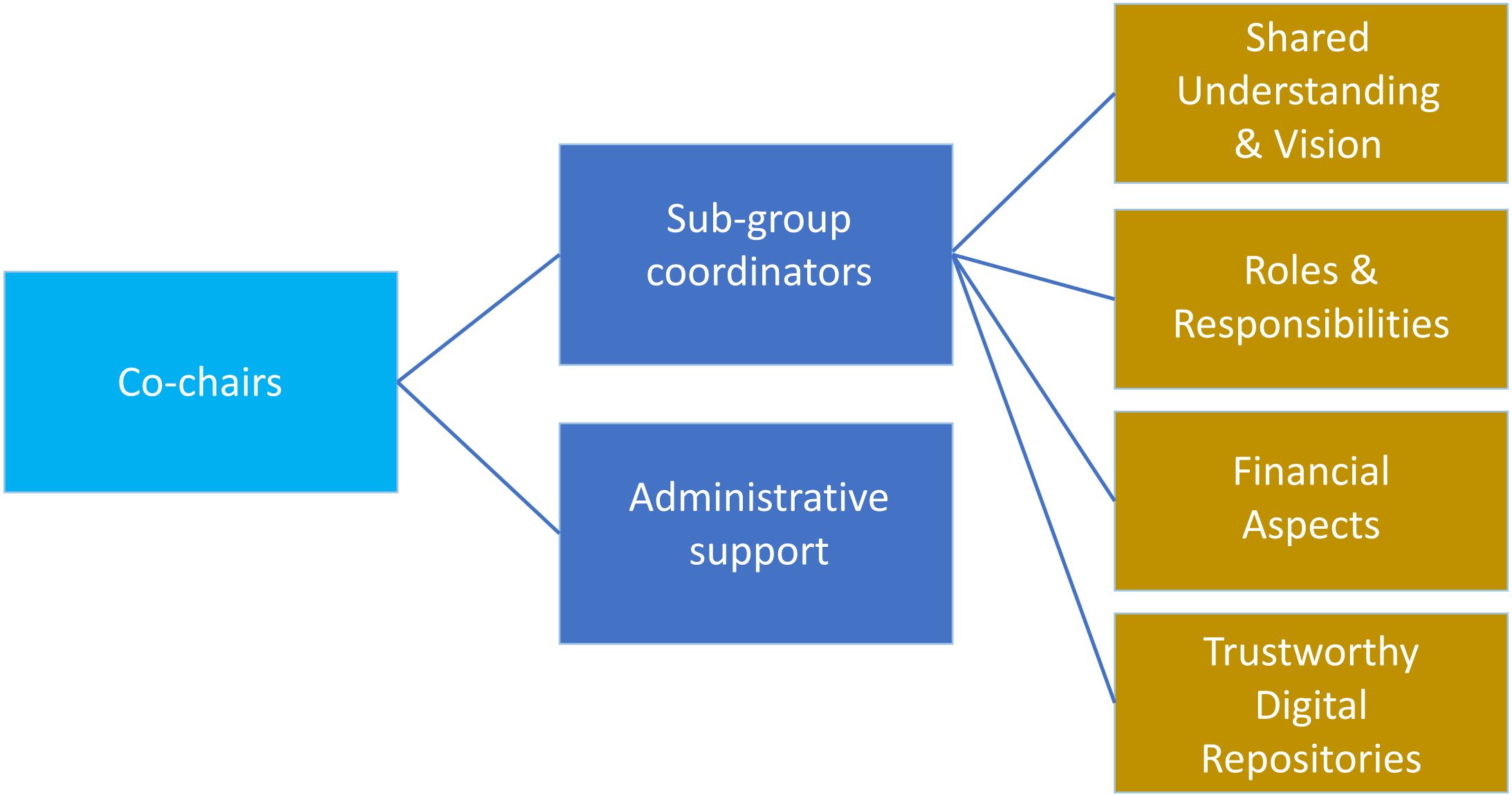
- **Providing recommendations for the EOSC board** on the vision and sustainable implementation of long-term data preservation policies and practices
- Making suggestions to later **strategy execution**
- Addressing the **roles and responsibilities** of the different stakeholders, the **financial aspects** of long-term preservation and the necessary **service infrastructure**

LTDP TF Members

36 members from 16 countries



LTDP TF Organization



LTDP TF Sub-group “Shared Understanding & Vision”

- Define what we mean by digital preservation in the context of EOSC
- Map the broad LTDP landscape (use cases, stakeholders, maturity levels, etc.)
- Provide suggestions policies and strategies
  - Retention and curation
  - Collaborations and synergies
  - Risk analysis and data governance

Expected Output

- Recommendations on the vision and sustainable implementation of LTDP policies and practices
- Suggestion for a strategy execution

LTDP TF Sub-group “Roles and responsibilities”

- Investigate the responsibility levels defined in the Strategic Research and Innovation Agenda (SRIA)
- Stakeholders identification in the different stages of the research data life cycle and their respective roles and responsibilities

Expected Output

- Mapping and promotion of the roles, responsibilities and accountability of the actors within the EOSC ecosystem with respect to long-term data preservation

LTDP TF Sub-group “Financial Aspects”

- Question the value of data, their price to preserve them (including indirect costs – eg. Carbon footprint)
- Identify business cases

Expected Output

- Mapping of the financial aspects of LTDP, including the financial responsibilities of the different stakeholders
- Mapping of the business models behind repositories

LTDP TF Sub-group “Trusted repositories”

- Establish an inventory of the operating European repositories
- Assess the percentage of the repositories that will be certified by 2025

Expected Output

- Recommendations on the creation of a European network of trustworthy digital repositories following FAIR-enabling principles with disciplinary and geographical spread
- Recommendations for EOSC data services to connect to this network
- Roadmap to further mature the LTP aspects of these repository services

LTDP TF Methodology

- Desk research to identify reports (eg. FAIR lady report, FAIR Forever?, Practical Guide to Sustainable Research Data), project outputs, etc. that can be used to build on draft recommendations
- Open consultation (webinar/workshop)
- Revised recommendations based on “decision-making on consensus”

Planned duration: 24 months (till December 2023)

LTDP TF Dependencies

Input from, collaborates with and creates synergies with:

- TF Defining funding models for EOSC
- TF FAIR metrics and data quality
- TF Technical interoperability of data and services
- TF PID policy and implementation
- TF Data stewardship curricula and career paths
- TF Researcher engagement and adoption
- TF Infrastructure for quality research software

LTDP TF Web Page

https://www.eosc.eu/advisory-groups/long-term-data-preservation

THANK YOU !

EOSC Task Force on Infrastructures for quality research software

Roberto Di Cosmo

Director, Software Heritage

Inria and University of Paris, France

Co-chair of the TF

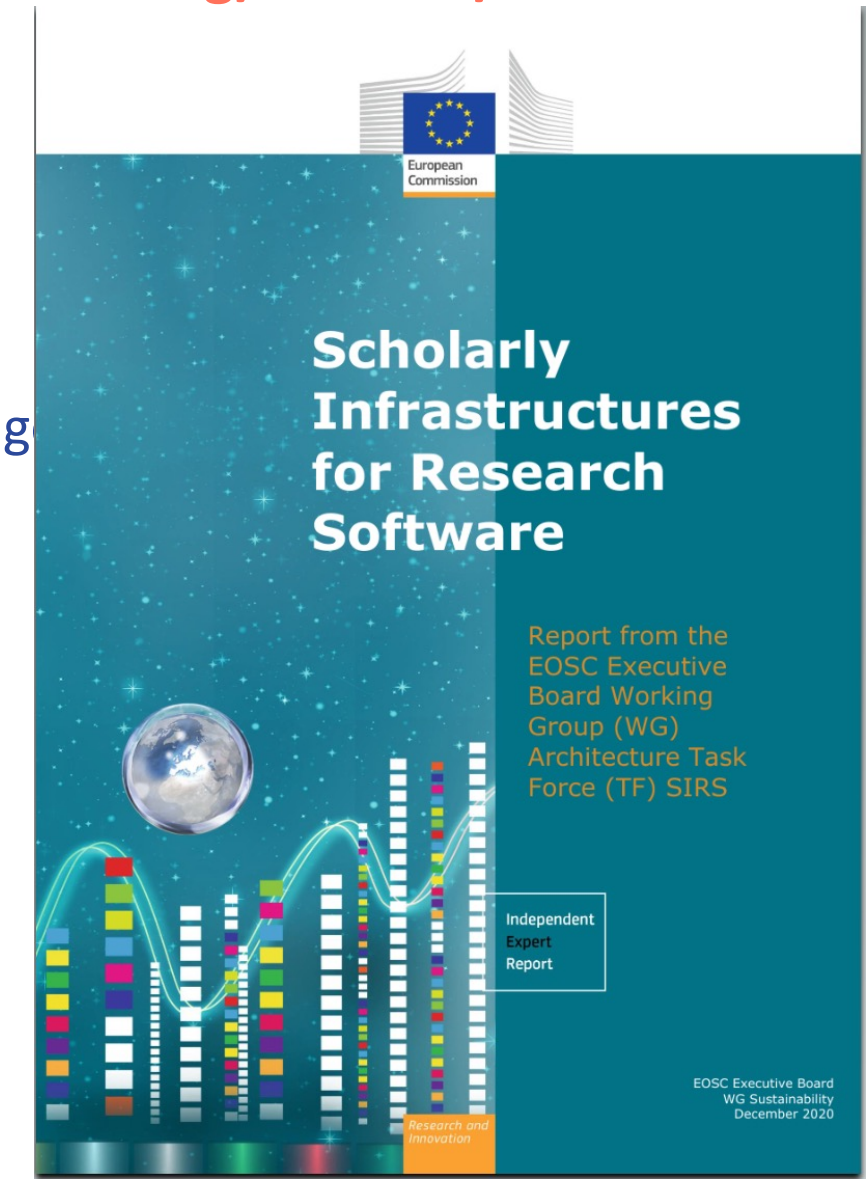
roberto@dicosmo.org

The origin: EOSC SIRS report (12/2020) doi.org/10.2777/28598

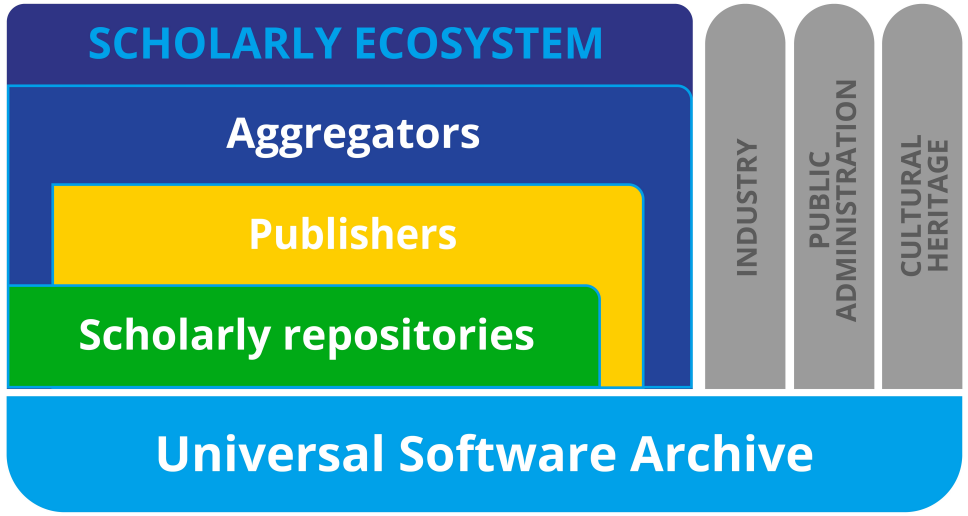
Chair: Roberto Di Cosmo, Software Heritage
Co-Chair: José Benito Gonzalez Lopez, Zenodo

- Focus on **Software Source Code**
- Four Pillars **Archive, Reference, Describe, Credit**
- State of the Art
  - Best Practices & Open Problems
  - Cross Cutting Concerns
- The Road ahead
  - Requirements & Criteria
  - 13 Workflows / Use Cases examples
- Recommendations
  - Standards & Tools
  - Policy recommendations
  - Long term perspectives

- **Archives**
  - HAL
  - Software Heritag
  - Zenodo
- **Publishers**
  - Dagstuhl
  - eLife
  - IPOL
- **Aggregators**
  - OpenAIRE
  - scanR
  - swMATH



Overall Architecture of SIRS and recommendations



Software Heritage

- **Scholarly ecosystem**
  - Aggregators collecting data from...
  - Academic publishers
  - Scholarly repositories
- **Software Heritage** connects with the global software development ecosystem

Main challenges identified :

- the long–term archiving
- global referencing
- interoperable descriptions
- crediting of authors.

Actionable recommendations to:

- Strengthen interactions between archives, publishers and aggregators,
- Agree on software metadata standards, like CodeMeta
- Generalize the use of extrinsic identifiers and intrinsic identifiers (SWHID)
- Ensure appropriate citations for research software source code,
- Foster standardization through policy and guidelines, and
- Ease adoption of the processes and tools for the research community at large

The Task Force on Infrastructures for Quality Research Software

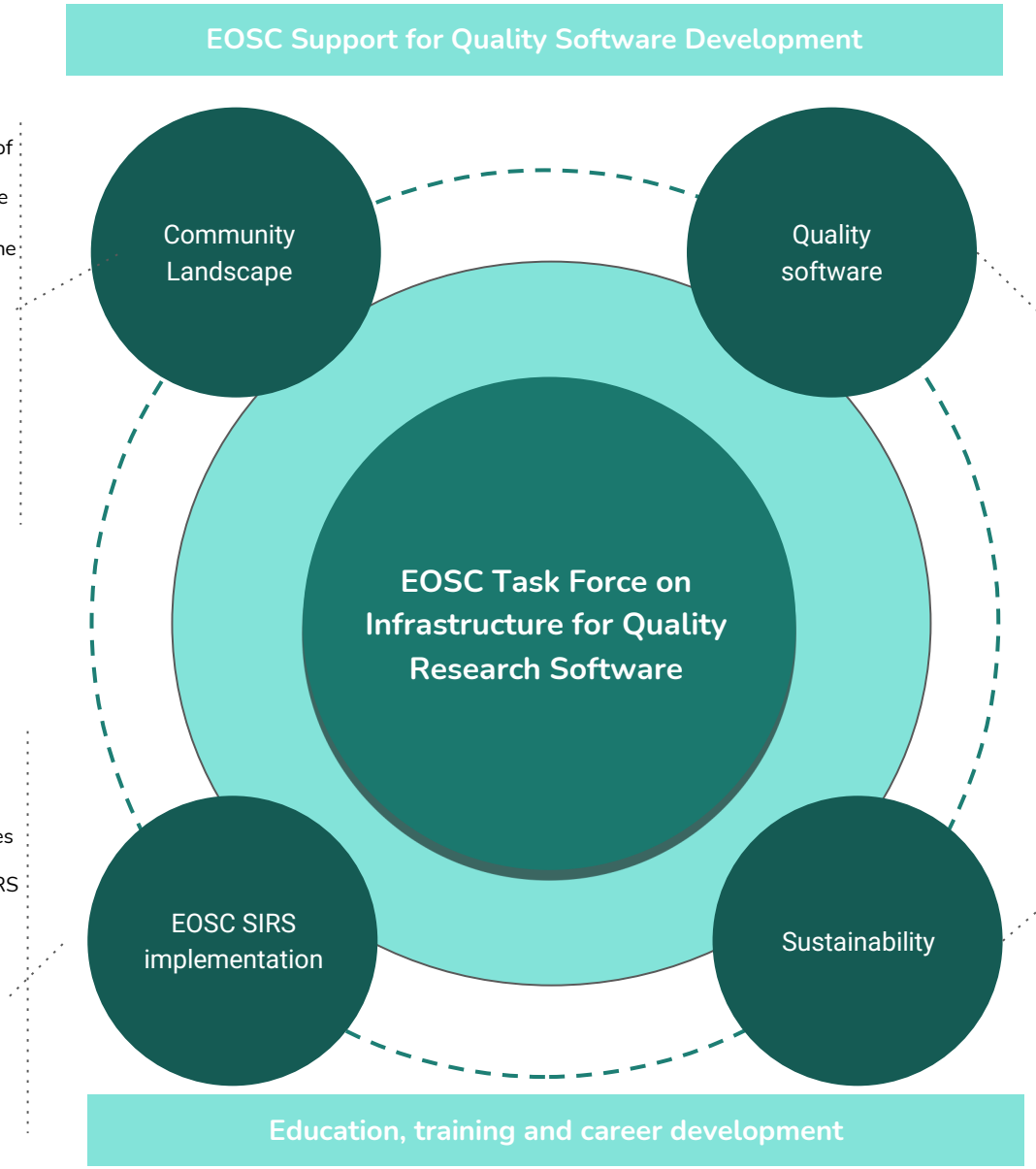
- Coordinated by Roberto Di Cosmo (INRIA–FR) and Isabel Campos (CSIC–ES)
  Draft Charter available in the EOSC–A website.
- **Objectives**

- Foster development and deployment of tools and services to properly archive, reference, describe with proper metadata, share and reuse research software

- Improve the quality of research software, both from the technical and organizational point of view and in particular the software used in the services offered through EOSC

- Increase recognition to software developers and maintainers of research software as a valuable research result, on a par with publications and data

Core Activities



1. Explore tools, standards and platforms that are used in state of the art software development
2. Identify open source development communities with which the research communities have broad connections
3. Formulate actionable recommendations, taking into account the diversity of software size, complexity, organization and longevity that one finds in the research world

1. Identify standards-based practices to be used for guidelines and procedures that would foster the development and deployment of high quality research software.
2. Provide requirements to infrastructure managers to enable the provision of platforms able to connect quality software and quality data.
3. Identify software development best practices to enable researchers to develop and run their research software on a diversity of computing infrastructures.

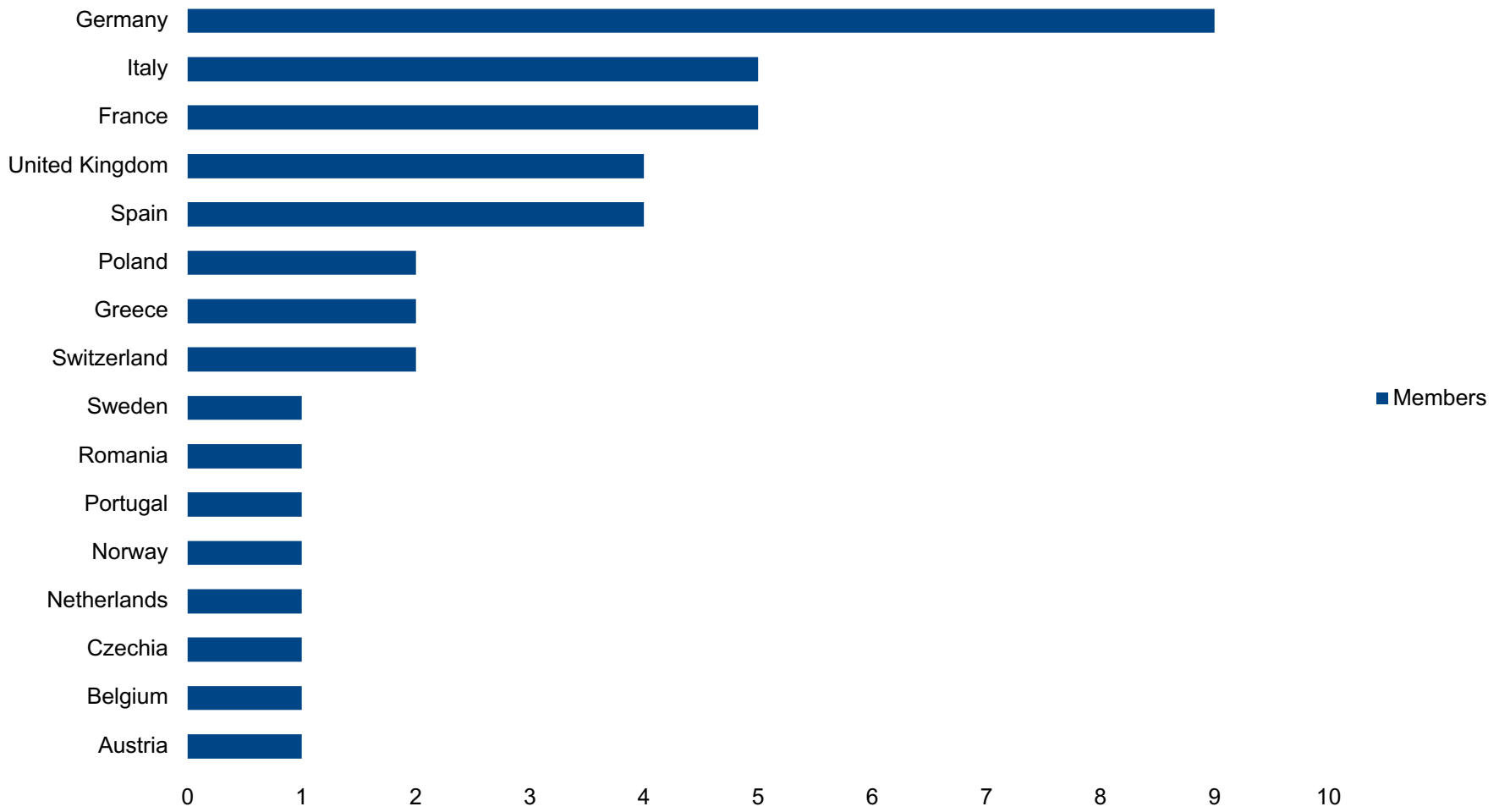
1. Broad dissemination of the EOSC SIRS report
2. Participation in existing standardisation efforts on metadata and identifiers for software source code
3. Federation and interconnection of the identified infrastructures address the interconnection between articles, data and software (extending the EOSC SIRS report)

1. Build a landscape of the approaches to sustaining the development of research software
2. Identify the blockers and limitations as well as the best practices
3. Formulate actionable recommendations based on the findings

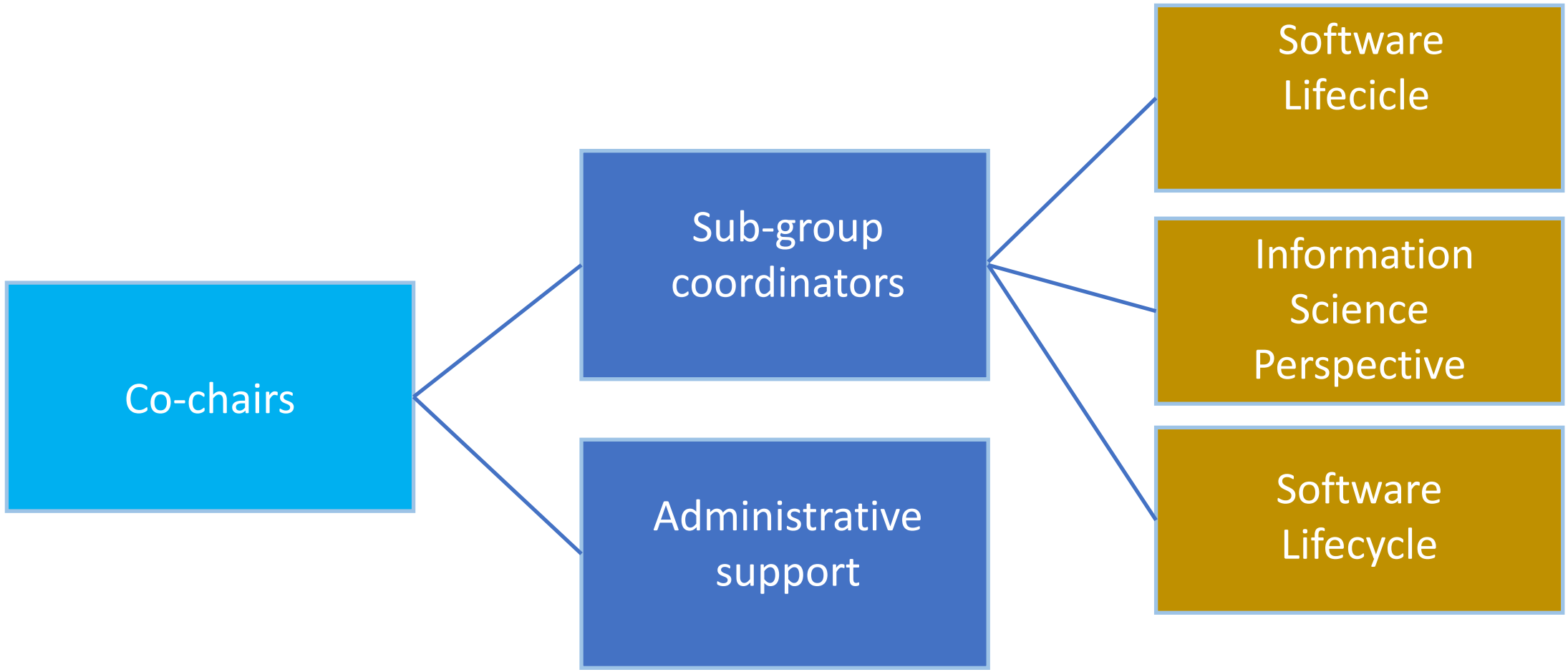
TF Members

41 members from 17 countries

Austria	1
Belgium	1
Czechia	1
France	5
Germany	9
Greece	2
Italy	5
Netherlands	1
Norway	1
Poland	2
Portugal	1
Romania	1
Spain	4
Sweden	1
Switzerland	2
United Kingdom	4



TF Organization



Web Page

https://www.eosc.eu/advisory-groups/infrastructures-quality-research-software

THANK YOU !