Le CNRS dans la construction de la fédération EOSC
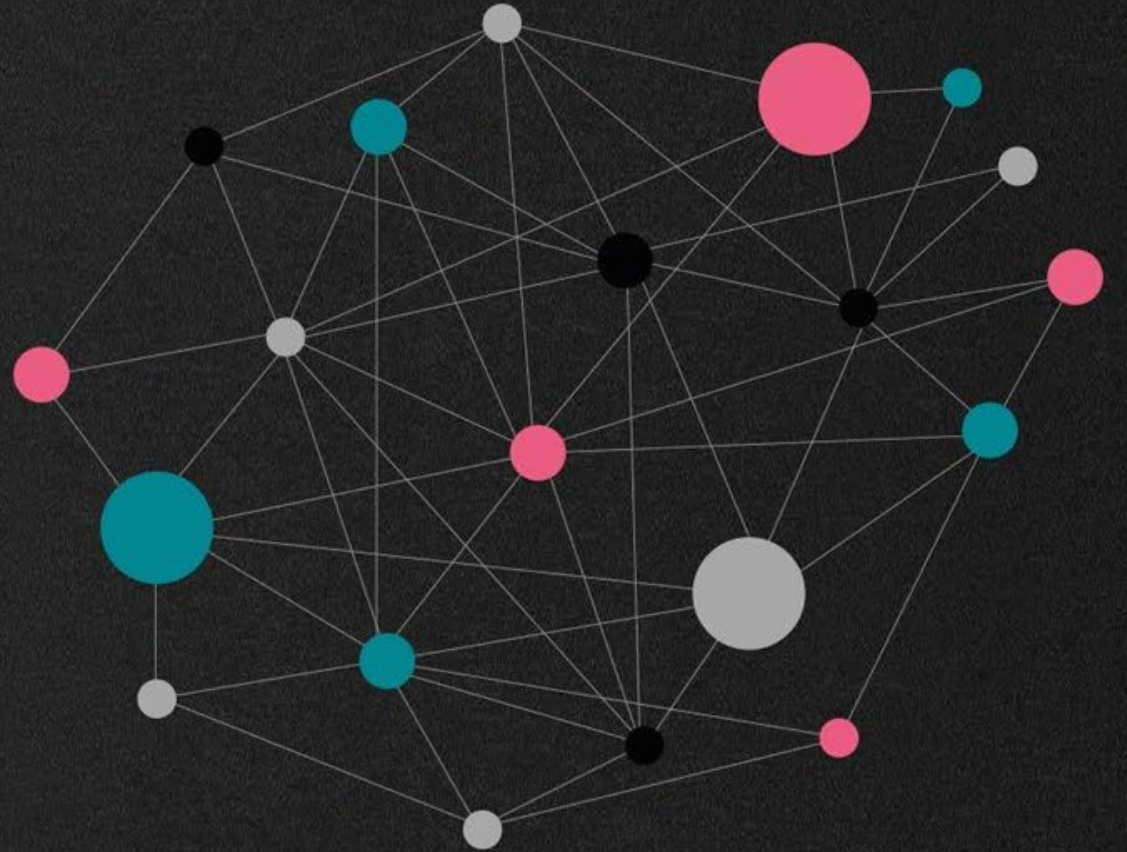
Suzanne Dumouchel,
DDOR, EOSC-A Board
27 Novembre 2024

1

Questionnaire

Open: 12 Jun 2024 - 18 Aug 2024 (late submissions accepted until 31 Aug 2024)

1. Do you intend to offer your institutional resources through a future EOSC Node of the EOSC Federation?

2. Do you wish to join the build-up phase with the intention to build a potential future EOSC Node?

3. Are you able to identify the legal entity that would assume legal responsibility for the potential EOSC Node?

4. Please list the resources your organisation intends to make available via the node. For each entry give the Technology Readiness Level (TRL 1-9), any restrictions that apply to use of the resource, and describe their added value to the EOSC Federation.

5. Are all these resources owned/operated by your organisation?

6. Please estimate when you would be able to start offering resources to the EOSC Federation?

7. What support/competences/training will be offered to the users of the resources made available via the potential node?

8. What is the added value that you see for you and for your users in contributing to the EOSC Federation?

9. Please add any additional remarks you would like to make about the EOSC Federation

10. Upload supporting documents (optional)

Overview of questionnaire submissions

Open: 12 Jun 2024 - 18 Aug 2024 (late submissions accepted until 31 Aug 2024)

121 Submissions

25 Countries plus many international organisations

52 Nodes with hundreds of repositories and services

5 Node types: thematic, national, institutional, regional, and other (such as the EU Node)

42 Nodes are ready to be deployed before July 2025

eosc

41 Research Performing Organisations

19 universities

25 service providers

21 EIRO Forum / ERIC / ESFRI*

5 national authorities, funders, or coordination bodies

What is the added value that you see for you and for your users in contributing to the EOSC Federation?

Summary of responses

- Increased visibility and reach

- Collaboration and interoperability

- Support for Open Science and FAIR principles

- Access to broader resources and expertise

- Influence on policy and standardisation

- Sustainability and economic benefits

Goals of the build-up phase of the EOSC Federation

Users

By **Q4 202**

... access t

... search a
enhance

... use an

... access
from di

... be **cont**
of the E

Federation

By **Q4 202**

... consist

... have es
clear de

... have de
organisa

... have es
Access a

... have es

... have de
monitor

Communities and stakeholders

By **Q4 2025** communities/stakeholders interested to contribute to the EOSC Federation will be provided with ...

... detailed **practical information** about the key aspects and components of the EOSC Federation, such the governance, operations, policies, procedures, technical specifications and standards (included in the EOSC Federation Handbook)

... **guidance** on the procedures for enrolment of Nodes and onboarding of resources, as well as the estimated costs and effort

... demonstration of multi-node end-to-end **use cases** that showcase the added-value of joining the EOSC Federation

... continuous opportunities to be **consulted** in the stages of development of the build-up phase

Identification of potential candidate Nodes
Implementation of sequencing criteria

Interest to become an EOSC Node

✓ Respondent expresses intention to build an EOSC Node and can identify legal entity

Maturity and **diversity of resources** that can be offered through the EOSC Federation

✓ Respondent can offer a diverse set of resources/services proven

✓ Services / resources offered are in mature state, priority to those with track record of operating them in production grade environment

✓ Respondent is referred by other respondents as representative candidate Node to onboard resources

Inclusivity and **representation** of research communities and different types of organisations

✓ Inclusion of organisations from the thematic domains of the **five science clusters,** as well as of **national scope**

✓ Inclusion of **RPOs, service providers, research infrastructures, national organisations and institutions**

✓ Broad **geographic representation**

✓ Potential for **coordination** amongst groups of respondents

✓ Take account of participation in **EOSC EU Node consortium**

2-stage dialogues with potential candidate Nodes

1st stage – confirm commitment to join build-up phase

- Objectives
  - o Present key **goals/deliverables** of the build-up phase & **timeline** for delivery
  - o Discuss the **expectations of candidates** in joining the Federation
  - o Clarify and confirm expected levels of **commitment** and required resources
  - o Encourage community coordination to identify **representative organisations** to take part in the build-up phase
  - o **Identify** the potential candidate Nodes to participate in the 2nd stage
- Participants
  - o all potential candidate Nodes; the Tripartite Group
- Timing
  - o 4 meetings; **December 2024 to January 2025**

2-stage dialogues with potential candidate Nodes

2ⁿᵈ stage – technical and organisational deep-dives (one-on-one)

- Objectives
  - Clarify the **organisational capacities** to achieve the goals of the build-up phase
  - Clarify the **technical requirements** to achieve the goals of the build-up phase
  - Discuss the **technical challenges the candidates aim to tackle** by joining the Federation
  - Identify a list of around **7-10 candidate Nodes** to participate in the first wave of enrolment
- Participants
  - potential candidate Nodes (reduced list); EOSC EU Node, EOSC Tripartite Group
- Timing
  - Around 4-8 meetings*, each w/ 2-5 organisations
  - **January 2025 to early February 2025**

* Will be determined by the number of organisations continuing to the 2ⁿᵈ stage.

2-stage dialogues with potential candidate Nodes
1st stage: 4 meetings during December - January

Dialogue #1
*
BBMRI ERIC
Elixir Hub
EMBL
Euro-BioImaging ERIC
Instruct ERIC

Dialogue #2
**
ACTRIS ERIC
CNR (Blue-Cloud)
ESRF (PANOSC)
CNRS-LAPP (ESCAPE)
CERN
ARIADNE RI
CLARIN ERIC

Dialogue #3
*
DE   **NFDI**
FR   **CNRS (Data Terra)**
HU   **HUN REN**
IT   **Foundation ICSC**
LU   **Lux NDS**
PL   **NCN**
SK   **CVTI SR**
UA   **BITP**

Dialogue #4
*
AT   **ACOnet / EOSC Sup. Of.**
DK   **DeiC**
FI   **CSC – IT Center for Science**
HR   **SRCE**
NL   **SURF**
NO   **NRIS**
SE   **SND**
SI   **ARNES**
      EUDAT

* Max 2 to continue to the 2nd stage
** Max 7 to continue to the 2nd stage

EOSC EU Node officially launched on 22 October

As the first building block in the EOSC Federation, the EOSC EU Node makes a clear contribution to the adoption of Open Science principles across Europe.

Services offered by EOSC EU Node

- **Bulk Data Transfer:** Move data effortlessly to data-intensive execution environments

- **Large File Transfer:** Streamline large file transfers online with added security and integrity

- **Virtual Machines:** Design and conduct experiments with flexibility while ensuring reproducibility

- **Cloud Container Platform:** Deploy cloud-native containerised applications that can easily scale

- **Interactive Notebooks:** Create and share documents with real-time code execution

- **File Sync & Share:** Enable automatic file syncing and secure sharing across locations and teams

EOSC Federation Handbook

Progress update by the EOSC Association Board

Why a Handbook?

The "How To" manual for building the EOSC Federation

1 Most Deliverables from EOSC projects are not "How To's"

2 EOSC Federation and Node concepts (2023) not well defined

**3** Goal to build the 1st Federation in 2025 is very ambitious

4 Need a clear manual (+ guides) on how to build the Federation for Nodes to commit to joining the Federation

Outline of EOSC Federation Handbook

- Chapter 1 – Purpose
- Chapter 2 – Governance
- Chapter 3 – Operations

- Chapter 4 – Architecture
- Chapter 5 – Scientific Resources
- Chapter 6 – Policies

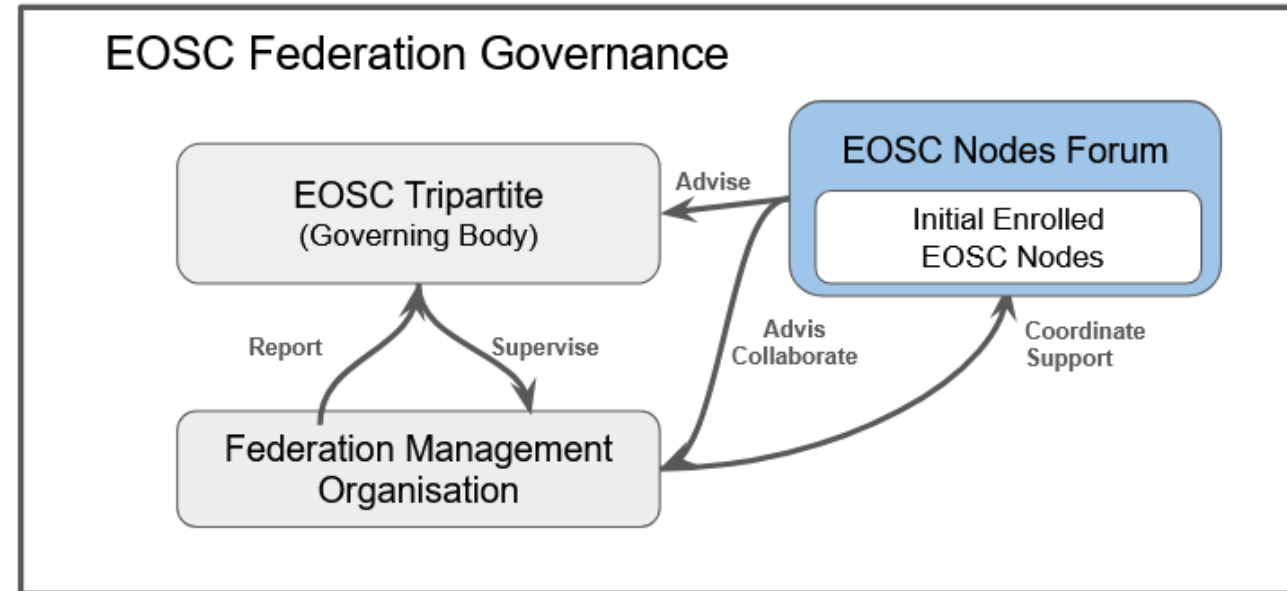## Purpose of 1st version of Handbook
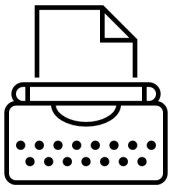
1 How to build the EOSC Federation during the period 2025 to 2027

2 What will the EOSC Federation enable and provide for researchers

3 How will the success of the EOSC Federation be measured

Chapter 2 - Governance

- Governance has been the trickiest chapter to write because:
  - Long-term view depends on newly formed Commission e.g. FP10 or not
  - Additional funding is coming from the European Commission
  - Resources are coming from the Nodes

- Clearly the **Tripartite Governance** is in charge of governance today and determines the roadmap and timeline

- However, the Federation needs a dedicated organisation to operate and manage it on a daily basis – **Federation Management Organisation**

EOSC Federation Governance

EOSC Tripartite (Governing Body)

EOSC Nodes Forum

Initial Enrolled EOSC Nodes

Advise

Report

Supervise

Advis Collaborate

Coordinate Support

Federation Management Organisation

Chapter 3 - Operations

"Operations management (OM) is the administration of business practices to create the highest level of efficiency possible within an organisation." [Investopedia]

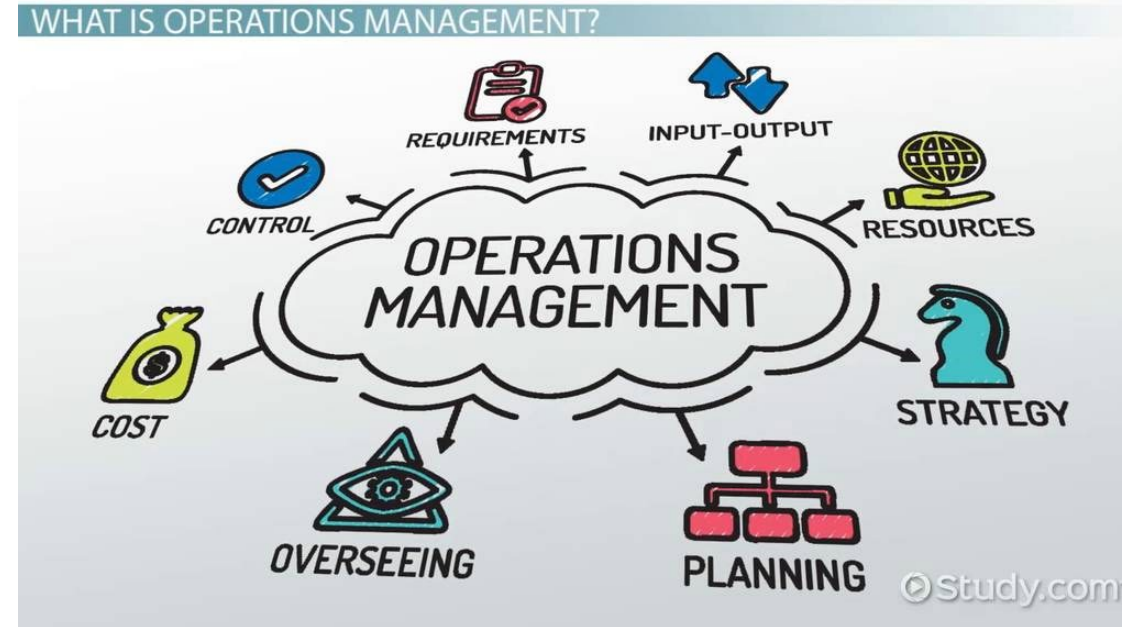This chapter describes what is needed to operate the Federation :

Day-to-day operations

Regular **management** activities

How to **Apply to become a Node**

EOSC Node roles identified

- o **Coordinator**
- o **Operation Manager**
- o **Technical Coordinator**
- o **Security Officer**
- o **Scientific Officer** (for Thematic Nodes)

Feedback = all sections are relevant or highly relevant

Describing the EOSC Node architecture will cover:

EOSC Node Architecture
EOSC Federation Interoperability Framework
EOSC Federating Capabilities
EOSC EU Node Federating Capabilities and Exchange Services
Technical Operations
Compute and Storage Resources
Cybersecurity
Links to External Entities
Training

18 Pages so far !

Interoperability

Chapter 5 - Scientific resources

- EOSC is supposed to be the Web of Scientific FAIR Data
  - ☐ it should therefore federate [**eventually all**] Scientific FAIR Data in Europe
- The chapter will describe what Scientific Resources can be found in EOSC Nodes:
  - o Data Repositories
  - o Metadata standards
  - o Training material for FAIR
  - o Scientific productivity tools
  - o Fostering domain specific communities
- It will describe how to connect them and make them visible in the EOSC Federation
- Challenge: how to attract trustworthy scientific resources of each domain

2.5 Pages so far…

Chapter 6 - Policies

Feedback = all sections are relevant or highly relevant

- Rules of Participation for Nodes
  - Must be Open, FAIR, follow standards e.g. cybersecurity
  - Includes requirements to help sustain EOSC Federation

- Access Policy
  - Defines who can access EOSC Resources

- Acceptable Usage Policy
  - Ethical, Legal and Social Implications (ELSI) Policy
  - Cybersecurity, Privacy and Data Protection Policy
  - Intellectual Property (IP) Policy
  - Code of Conduct

11 Pages so far ...

La communauté impliquée
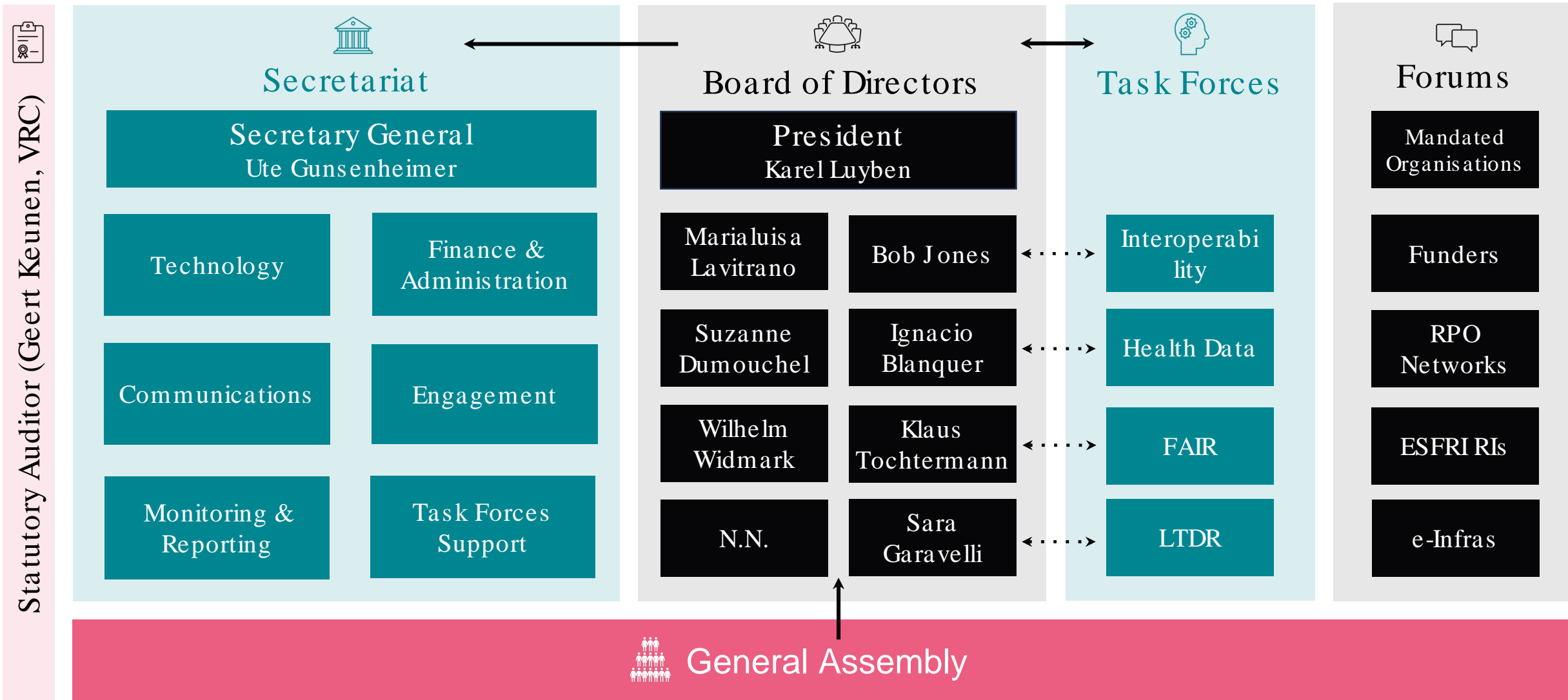
EOSC Association Task Forces

	FAIR Metrics and Digital Objects	Health Data	EOSC Technical and Semantic Interoperability	Long-Term Data Retention
Co-chairs	Elli Papadopoulou, Athena RC; and Mark Wilkinson, UPM	Lene Krøl Andersen, Computerome, DTU; and Petr Holub, BBMRI-ERIC & Masaryk University	Diego Scardaci, EGI Foundation; Jiří Marek, Masaryk University; Christos Kanellopoulos, GÉANT	Jenny O'Neill, HEAnet; Jacques Flores, Utrecht University
Members	51	44	60	37
Board liason	Klaus Tochtermann	Ignacio Blanquer	Bob Jones	Sara Garavelli

EOSC Opportunity Area Expert Groups

OA Expert Groups are an **important mechanism for collaboration** on technical and related matters within the EOSC Partnership

All OA Expert Groups have developed work plans and **meet roughly every four weeks**

EOSC Symposium 2024 featured a **panel discussion** dedicated to the collaboration within OA Expert Groups

Winter School 2024, Thessaloniki

OA Expert Group: Open Scholarly Communication

OA Expert Group: User and Resource Environments

OA Expert Group: Research Software

OA Expert Group: Persistent Identifiers

OA Expert Group: Metadata, Ontologies and Interoperability

OA Expert Group: FAIR Assessment and Alignment

OA Expert Group: Skills, Training, Rewards, Recognition and Upscaling

Et le CNRS?

Une approche consultative

- Deux groupes de travail au CNRS autour d'EOSC:
  - ASOP: réunit des directeurs d'infrastructures et des membres du COPIL DDOR
  - AT: réunit des membres du CNRS investis dans EOSC (projets EU, task forces, etc.).
- Deux workshops en 2024 pour lancer la réflexion sur les noeuds EOSC (invitation élargie)
  - 24 avril en ligne 2h
  - 7 février en présentiel 3h30

Avec diffusion des [CR](#) pour validation interne par le groupe.

- Questionnaire sur les noeuds EOSC lancé pendant l'été. Coordination interne pour rassembler les réponses et collecter les intérêts dans un [document synthétique](#) envoyé au DGDS
- Les réponses ont été analysées par l'association EOSC (slides à la fin de la présentation) et permettent de lancer les prochaines étapes.

Deux rôles possibles pour le CNRS dans les futurs noeuds EOSC

Porteur d'un noeud

- Data-Terra
- HAL+ (qui pourrait initier un noeud "archives ouvertes" porté par le CNRS qui inviterait ensuite d'autres organisations à rejoindre ce noeud).
- ESCAPE

Contribue à un noeud

- Le nœud Diamant,
- ELIXIR,
- RDG,
- CERN x2,
- EGI.

CDS: les deux options sont ouvertes

Giovanni Lamanna

ESCAPE sustainability

Since 1ˢᵗ Feb 2023 ESCAPE is an Open Collaboration

- ESCAPE has a Collaboration Agreement signed by Directors of all the partner RIs
  - The agreement came into effect at the end of ESCAPE H2020 project
  - Recognises many synergies: communities, technical, coordination, political, funding …
  - Common facilities (data centres, networks, etc.)
  - OPEN: RI's will join because they see a value in collaborating
  - Representing a single/integrated community towards EOSC and EC

Giovanni LAMANNA (CNRS-LAPP Director)
Director of the ESCAPE Open
Collaboration

The ESCAPE Open Collaboration Agreement (CA)

This Agreement
between
The Organisations listed in Annexe 1
Hereinafter referred to as 'Parties',
defines the ESCAPE Open Collaboration, the subject of this Collaboration Agreement.
Hereinafter referred to as "ESCAPE".

1. DEFINITIONS

Background IPR — means any IPR controlled or owned by any Party prior to the entry into force of this CA or IPR generated by any of the Parties independently of this CA and controlled or owned by that Party;

Confidential Information — has the meaning given to it in Section 7 of this CA;

Creating Party — has the meaning given to it in Section 8 of this CA;

EOSC — means European Open Science Cloud;

Foreground IPR — means any IPR arising from tasks as described in the Work Plan and carried out in the course of this Agreement by any of the Parties;

IPR — means any and all intellectual property rights anywhere in the world whether registered, are registerable or otherwise, including but not limited to patents, trademarks, registered designs, domain names, applications for any of the foregoing, trade or business names, copyright and rights in the nature of copyright, design rights, rights in databases, moral rights and know-how;

Party in Default — has the meaning given to it in Section 14 of this CA;

Work Plan — means the tasks to be performed during the course of this Agreement [further detail in Annexe 2 if required].

1 — 21ˢᵗ April 2022

ESCAPE work programme aimed at building a domain-based implemetation of EOSC

Virtual Observatory:
Extend the VO *FAIR* standards, methods and to a broader scientific context; prepare the VO to interface the large data volumes of next facilities.

Science Platforms:
Flexible science platforms to enable the open data analysis tailored by and for each facility as well as a global one for transversal workflows.

Citizen Science:
Open gateway for citizen science on ESCAPE data archives and ESFRI community

Data Lake:
Build a scalable, federated, data infrastructure as the basis of open science for the ESFRI projects within ESCAPE.

Software Repository:
Repository of "scientific software" as a major component of the "data" to be curated in EOSC.

EOSC Nodes

- "(CNRS-LAPP) ESCAPE" is one Candidate EOSC Node (first wave) invited to participate in first-dialogue

- The ESCAPE collaboration answered the ESOC questionnaire in July 2024
  - Proposing a thematic 'Science Cluster Node' to operate its resources within the EOSC Federation; asses the EU Nodes services' quality; deploy training and science camps; operate a competence centre and a virtual institute for software; manage cascading grants.
  - Any service that could be offered by the ESCAPE EOSC Node in the future will be the result of distributed commitments of all its partners, of a group of them and of national institutes.
  - ESCAPE considers relevant the inclusion of the five Science Clusters in the EOSC Federation

- Giovanni LAMANNA replied to the questionnaire on behalf of the ESCAPE international collaboration.
  - CNRS-LAPP being his affiliation and deploying the management team of ESCAPE, appears as the reference legal entity in the questionnaire. This does not imply that CNRS is responsible for all ESCAPE current and future resources, including those prospected within the remits of a future ESCAPE Node.

DATA-TERRA... un nœud EOSC thématique autour du Système Terre

Le nœud EOSC Data-Terra sera en mesure d'offrir des services de **découverte**, de **catalogage**, d'**accès**, de **traitement**, d'**analyse** et de **visualisation dynamique** de données (spatiales et in situ) ainsi que de **services** thématiques provenant des composants du système terrestre (atmosphère, surfaces terrestres, biodiversité, océan, terre solide) et de leurs interactions.

Le **CNRS** est l'entité légale qui porte le projet au nom de DATA-TERRA et de l'ensemble de ses tutelles et partenaires.

Data Terra s'appuie sur ces centres de données et de services pour fournir des données et des services aux scientifiques

1. **Accès aux données** — Liens avec les producteurs de la donnée — Accéder à des services dédiés
2. **Production régulière de données** — Améliorer la qualité des méta données
3. **Analyse et traitement à la demande** — Faciliter le croisement des observations
4. **Services d'aide aux utilisateurs** — Partage logiciels, Plateforme d'analyse, Évaluation modèle

2

Un offre diversifiée et structurée sur le long-terme

5 pôles thématiques de données et services ainsi que un pôle d'imagerie spatiale (TRL 9)

« Développer et opérer une infrastructure distribuée et intégrée de données et de services pour l'observation et la modélisation du système Terre, de la biodiversité et de l'environnement »

Un portail de la découverte et service d'accès (TRL 7 > TRL 9)

« The first interdomain digital architecture for integrated use of environmental data »

Virtual Research Environments / Analytics services, tools et trainings (i.e. Galaxy sub-domain pour le système Terre); cloud-based services thématiques (TRL 7 > TRL 9)

EaSy Data — Data Repositories (TRL 8)

EarthPortal — Semantic Artefacts Repository (TRL 7)

3

Un projet en adéquation avec les prérequis de la Fédération EOSC

- Legal status

- Large-scale, quality service provision

- Capacity to onboard third-party services

- Capacity to contribute to EOSC core capabilities

- Compliance with EOSC federation rules and standards

- Effective monitoring

- Community engagement

- Sustainability (5 years)

HAL+ permet aux chercheurs et aux institutions d'archiver et de diffuser leurs publications en accès ouvert

- Soutien du CNRS depuis 2000

- Label d'infrastructure de recherche

- 3 plateformes
  - HAL, archive ouverte nationale
  - Episciences, publication de revues en OA diamant
  - Sciencesconf, gestion d'évènements scientifiques

- Focus sur **HAL**
  - Inscription dans les PNSO
  - 1,4M documents scientifiques en texte intégral
  - 152 portails HAL
  - 70% des consultations web en provenance de l'étranger

1

Réseaux, partenariats et collaborations existant

France

- IR : HumaNum, OpenEdition
- Recherche data gouv
- Agences de financement : ANR, ANSES, ADEME
- Centres de calcul : cc in2p3 et CINES
- Référents HAL (bib. universitaires, labos)

Europe

- Projets : EOSC Pillar, OpenAire Nexus, SoFAIR, FAIR4EOSC, EOSC Data Commons
- OpenAire
- DARIAH

Monde

- Confederation of Open Access Repositories (COAR)
- Interopérabilité : arXiv, REPEC, bioRxiv, medRxiv, Zenodo, Software Heritage

HAL+, un nœud de services de publications en accès ouvert intégré au sein de l'EOSC

Valeur ajoutée de HAL+

- Entrepôt de publications scientifiques
- Infrastructure opérationnelle (TRL9)
- Contenu multidisciplinaire
- Contenu multilingue
- Compatibilité avec les principes FAIR
- Adéquation avec les prérequis pour devenir un nœud EOSC

Positionnement de HAL dans la fédération EOSC

Fournisseur de données	Fournisseur de services	Animation réseau Sensibilisation
• Partage et ouverture de données FAIR • Métadonnées • Publications • Liens des publications avec les jeux de données et les logiciels • Site web et API	• Publication de revues en accès ouvert via Episciences	• Bonnes pratiques et partage d'expertise avec les AO européennes • Liens entre AO et services de *peer reviewing* • Mise en relation des publications, données, logiciels

2

38

Pour conclure

Le CNRS est largement impliqué dans la construction de la fédération:
- En tant que membre de l'Asso EOSC avec la participation de collègues dans les initiatives l'association
- En participant et coordonnant des projets EU
- En étant positionné dans le board
- En étant impliqué dans le Collège EOSC-France

Pour garantir une fédération EOSC qui réponde aux enjeux du CNRS mais aussi aux besoins et attentes des communautés scientifiques du CNRS.

Des questions?

Merci de votre attention!

Suzanne Dumouchel
CNRS / EOSC-A
suzanne.dumouchel@cnrs.fr