The Web of FAIR Research Data

Karel Luyben
President EOSC Association
Rector Magnificus Emeritus TU Delft
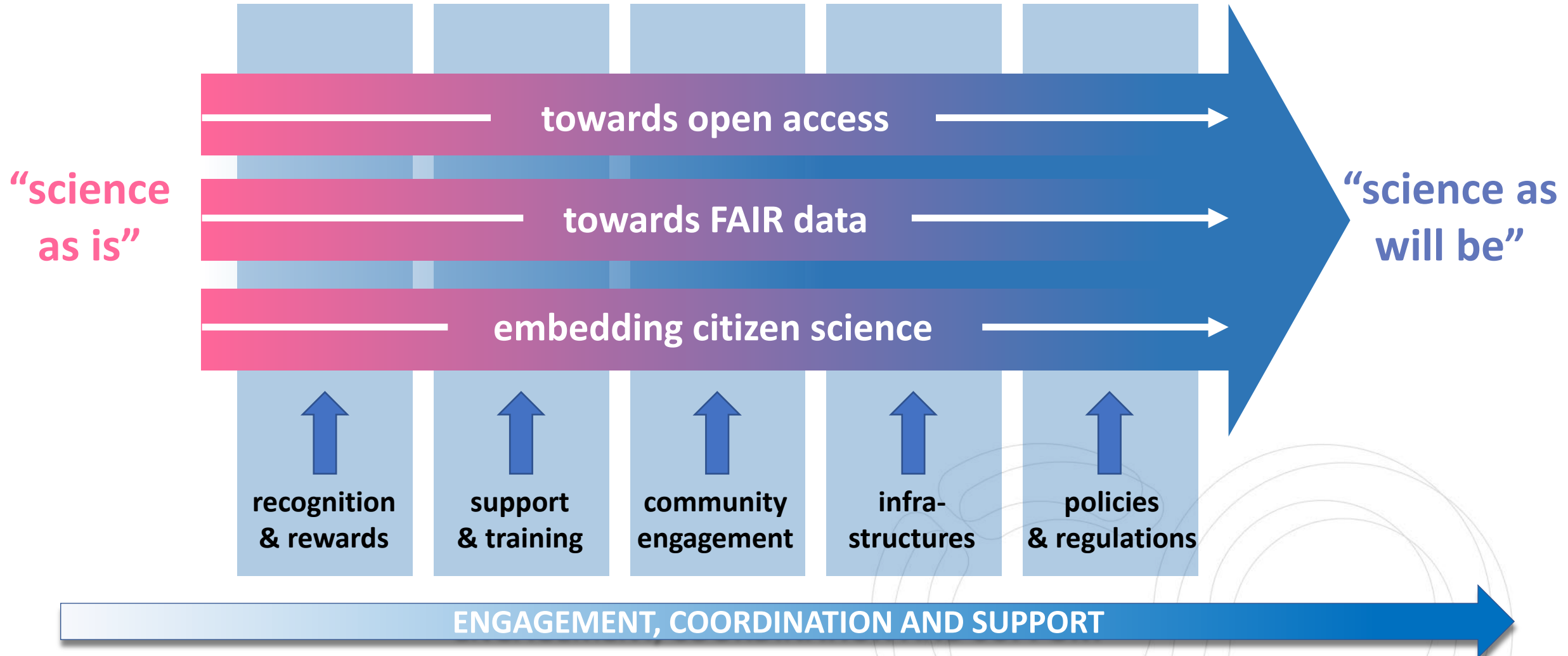
ETH Zurich, 18 January 2024

Elements of Open Science (UNESCO 2021b)

CESAER and Open Science

Task Force **'Openness of Science and Technology'** works on:

Open science: The focus is on Research Data Management (RDM), advancement of the European Open Science Cloud (EOSC), FAIR Data and Open access.

Knowledge safety and security: Taking no measure would be naive. We need to prevent too many and the wrong measures which would reduce scientific integrity, academic freedom and institutional autonomy

Citizen science: This task force explores effective two-way engagement (outreach and engagement) between our Members and society at large.

In 2019, TU Delft library and the Faculty Data Stewards drafted and started implementing a '*Vision for Research Data and Software Management training*' based on seven principles (five are listed here)

https://doi.org/10.5281/zenodo.3516874

TUDelft

1. Data and software management training should be built upon the existing courses were possible
2. New courses are also necessary to deal with the fast-changing demand in research data and software management
3. Researchers should receive the proper incentives to take part and contribute to the training.
4. The library and graduate schools should continuously engage in consultation processes with PhD students and researchers
5. All material produced for the courses described in this vision will be published and made available under a Creative Commons – Attribution License (CC-BY 4.0)

Training program – Research Data & Software Management
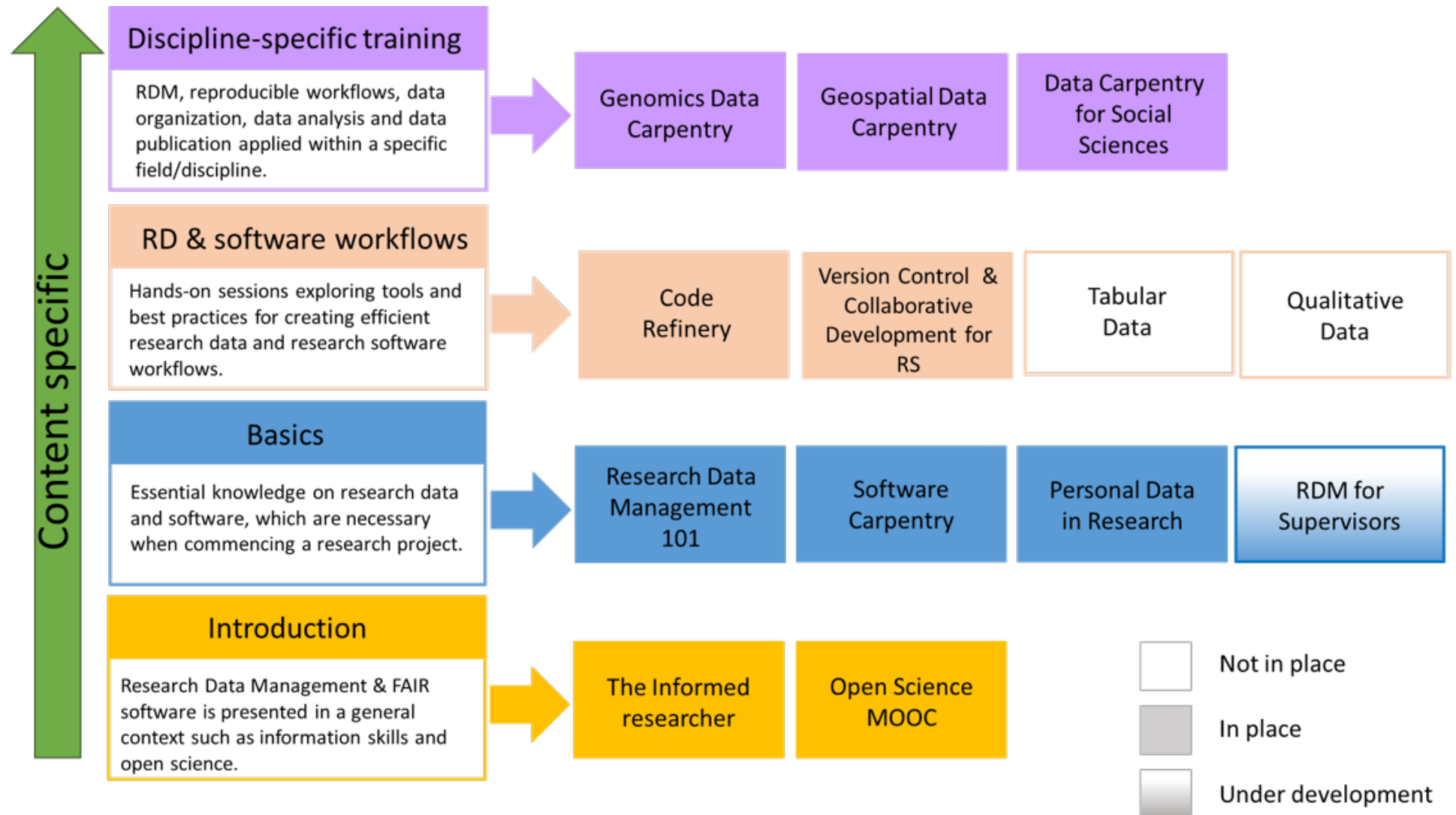
TU Delft (example)
Digital Competence Center (DCC)

Steering committee: ICT Innovation (coordinator) and TUD Library

Members: 4 data managers (Library), 6 Research Software Engineers (ICT-Innovation) and 1 HPC data steward (Delft High Performance Computing)

Mission: help researchers produce FAIR data, improve research software, and apply suitable computing practices to increase the efficiency of the research process.

Services:

- Supporting research projects with hands-on **co-development** of solutions (calls) - https://tudelftdcc.shinyapps.io/dccapp/

- **Training** on programming skills via Carpentries, Code Refinery and custom workshops

- **Consulting** on (FAIR) data management and software development

TU Delft Digital Competence Centre - YouTube

Organisational units/offices/services already in place before the establishment of the TU Delft DCC

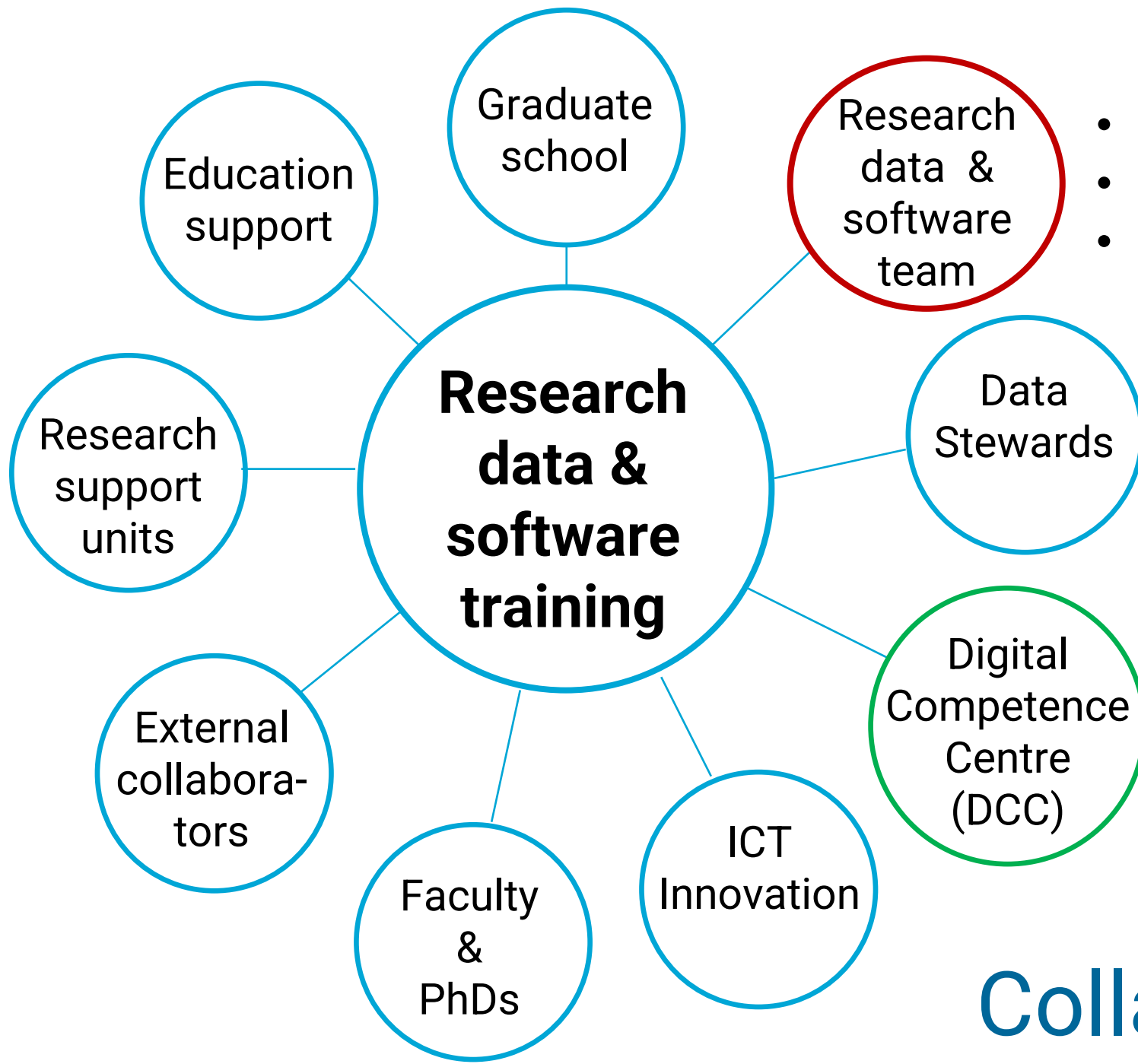## TUD Library - Research Data & Software Team

- Policies and workflows
- Collaboration across the university
- Data stewardship coordination
- Training provision/coordination
- Tools

4TU.ResearchData

- Repository for research data and software
- Tools and workflow integration
- Capacity building for data stewards (RDNL)

Faculty Data Stewards

- Advice and coordination at faculty level
- Advocacy and training
- Faculty policy and workflow development

TU Delft

Research Data Management 101 (basics)

- RDM101 is a blended three-week course aimed at first-year PhD candidates (one meeting per week)

- After taking this course, participants should be able to:
  - Realise the importance of **good data management** for research
  - Identify different **data types** relevant for their projects
  - Recognise the **relevant regulations, policies, and legal requirements** for their data
  - List the main components of the **FAIR data** principles
  - Connect the FAIR data principles to their **own research workflow**
  - Use what they learned to design an efficient **RDM strategy**

Faculty Data Stewards

They contribute to central training according to the skills needed by researchers at their faculty and/or to their interests

They teach and help in software carpentry, data carpentry for social sciences, genomic data carpentry, CodeRefinery, Version control and collaborative software development.

They were the initiators of the software carpentry workshops and recommended to become members of The Carpentries.

Seven of them are certified instructors of The Carpentries (soon to be eight)

They encouraged the library to organise Data Carpentry for Social Sciences workshops (an LDEV collaboration)
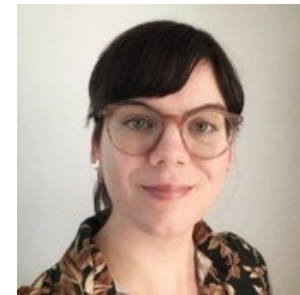
Digital Competence center (DCC)

They contribute to central training related to Research Software in voluntary basis. It is not well defined how much of their time should be dedicated to the central training.

They teach and/or help in software carpentry, CodeRefinery and Version control and collaborative software development

They were the initiators of the FAIR4RS mentoring programme in connection with the CodeRefinery workshops

Six of them are certified instructors of The Carpentries

They have recently updated the module on 'Making your research software FAIR' within the OS MOOC
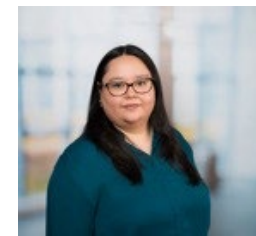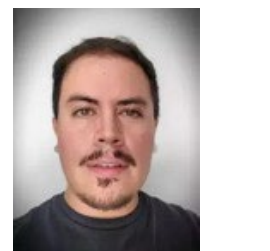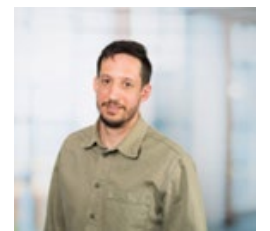
Research Data Management and digital skills trainers
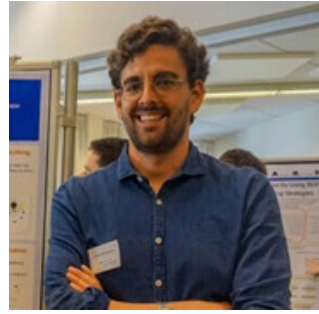
Trainers

Training assistant

Educational advisor

Coordinator

TUDelft
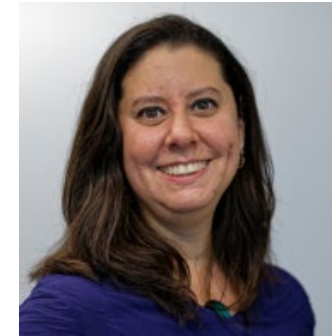
What is EOSC and what is not?

EOSC – additionality to the web of FAIR data

EUROPEAN OPEN SCIENCE CLOUD

WORLD WIDE WEB

INTERNET

NETWORKS

COMPUTERS

Modelled after: World Wide Web - Wikipedia

"A web of scientific insight"

- **Web of FAIR Data and related Services**

- **Federation of relevant existing and future data sources**

- **Virtual space where science producers and consumers come together**

- **An open-ended range of content and services**

- **Based on the FAIR principles**

- **Meeting all European data requirements**

- **In interaction with other regions of the world**
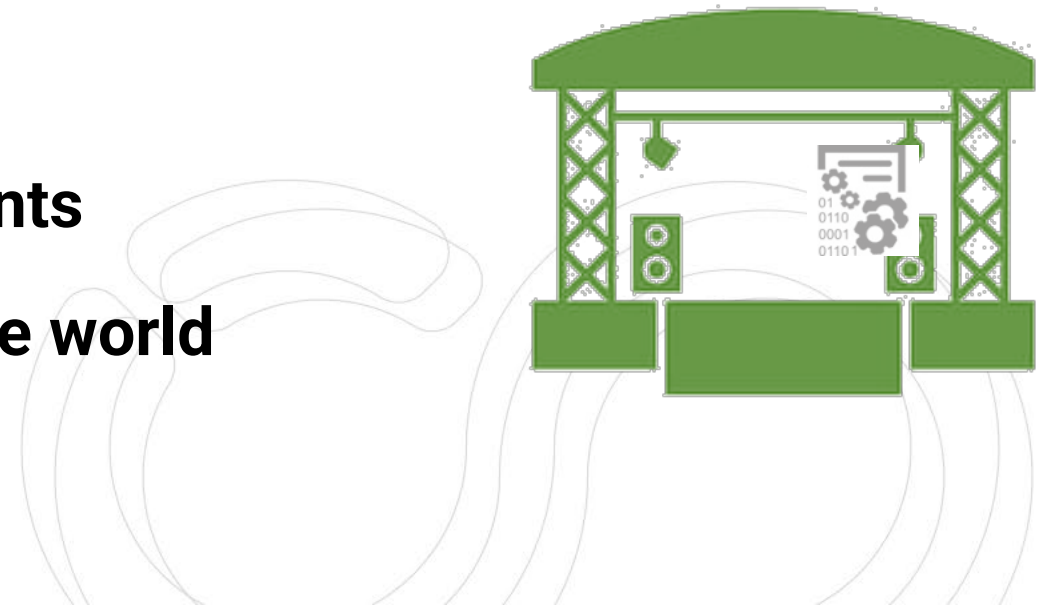
Ten things EOSC is not …

… a cloud infrastructure

… synonymous to Open Science

… the EOSC Portal

… the EOSC Association

… a new research data repository or data management system

… owning the data it will help to provide access to

… a new pan-European e-infrastructure

… substitution of any existing data- or (e-)infrastructures

… directly for individual researchers, but it is for research

… in competition with anything or anybody

Horizon Europe co-programmed EOSC Partnership

- Launched June 2021

- One billion euros commitment by the European Union and the EOSC Association

European Open Science Cloud Partnership

EOSC STEERING BOARD

European Commission

Anna Panagopoulou
DG-RTD
DG-CNECT

Member States and Associated Countries

Volker Beckmann

MoU

Partnership Board

MS/AC collective representation

Exchange of information

EOSC ASSOCIATION

President
Karel Luyben

Secretary General
Ute Gunsenheimer

BOARD

General Assembly

(including organisations performing research, funding research, providing services, organisations mandated by MS/AC and other organisations)

SRIA and MAR

Strategic Research and Innovation Agenda

Strategic Research and
Innovation Agenda (SRIA)
of the
European Open Science
Cloud (EOSC)

Version 1.0
21 June 2021

Independent
Expert
Report

EOSC Executive Board

General Objectives

GO1: Open Science practices and skills are rewarded and taught, becoming the 'new normal'

GO2: Standards, tools and services allow researchers to find, access, reuse and combine results

GO3: Sustainable and federated infrastructures enable open sharing of scientific results

Multi-annual roadmap sets priority activities and outcomes grouped by three implementation levels – European, National, Institutional
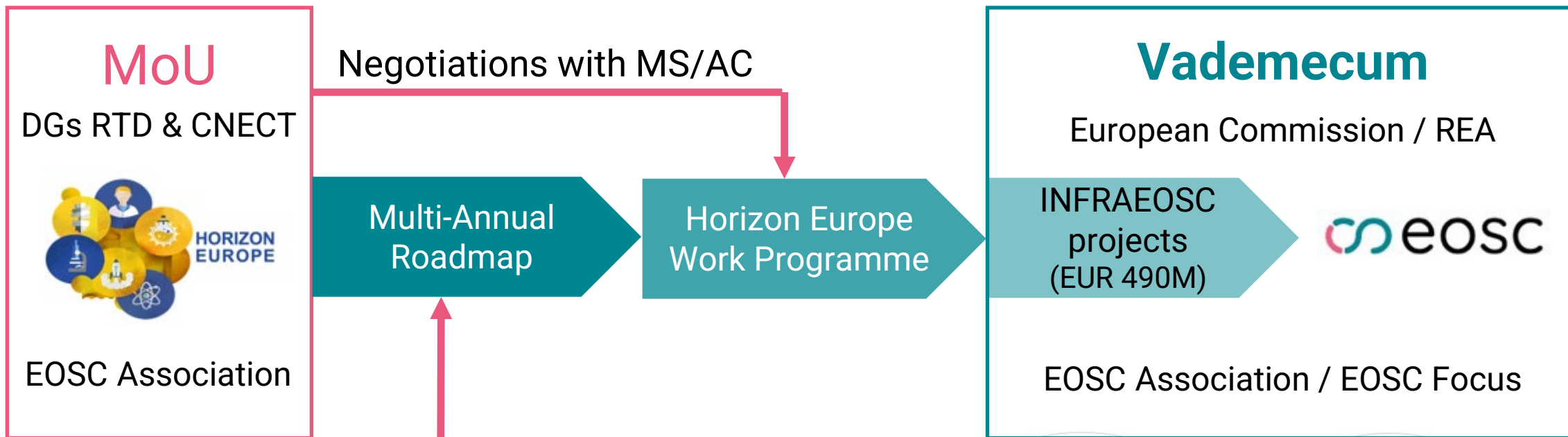
Three phases of MAR

- 2021-2022: development towards a functional federation of infrastructure
- 2023-2024: expansion to production that generates added value
- **2025-2027: expansion to develop impact from Open Science**

Delivering EOSC – the Process

The critical role of the INFRAEOSC Projects

MoU
DGs RTD & CNECT

HORIZON EUROPE

EOSC Association

Negotiations with MS/AC

Multi-Annual Roadmap

Horizon Europe Work Programme

Community Consultation

Vademecum

European Commission / REA

INFRAEOSC projects (EUR 490M)

eosc

EOSC Association / EOSC Focus

Present action: MAR 2025 - 2027

The EOSC Federation and the role of Nodes

Main points in the position paper

https://eosc.eu/wp-content/uploads/2023/11/20231112-Short-paper-on-the-EOSC-Federation-draft-v3.pdf

EOSC Federation

- is a federation of distributed systems ('system of systems')
- enables collaboration to achieve common goals and users to access additional resources beyond their usual environment
- has policies and rules defined by the legal entity governing the EOSC Federation
- consists of multiple "Nodes"
- can be scaled by adding more Nodes

EOSC Nodes

- contain resources adding value to the Federation
- act as the legal representative that can interact with EOSC's governance structure
- offer interfaces that comply with the EOSC Interoperability Framework
- control their own operations and resources and ensure that policies are followed within the Node
- may vary in their local policies, the resources to which they provide access and the infrastructure on which they are built

Initiatives (European, national, regional, institutional or thematic) may join the EOSC Federation and become EOSC Nodes when they meet the requirements listed above
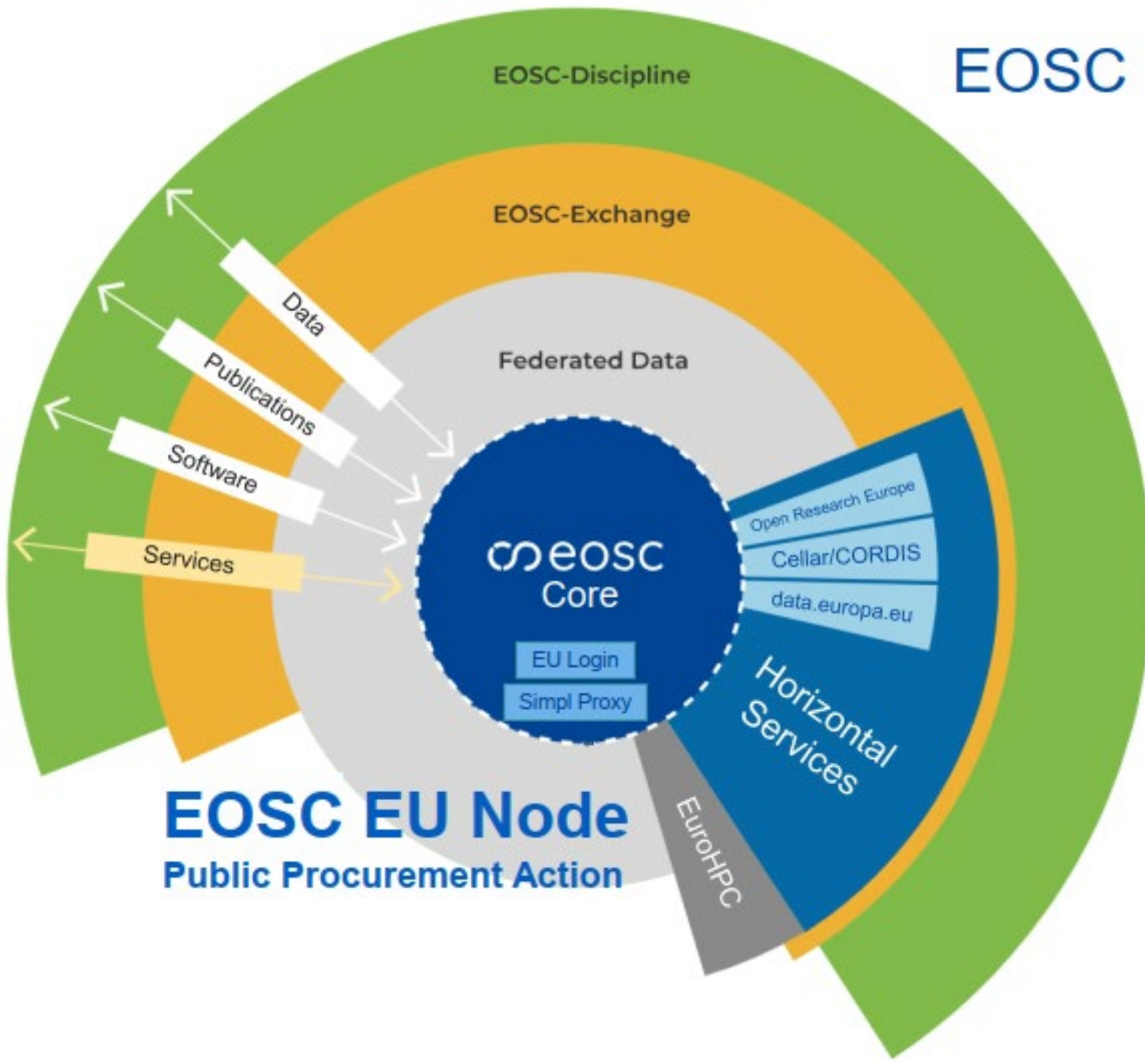
EOSC EU Node
Value Proposition and Architecture

DG CNECT Unit C.1 Open Science and Digital Modelling

Peter Szegedi

EOSC EU Node Value Proposition

- **Facilitate** the creation of the *"Web of FAIR data and interoperable services"* (aka. EOSC Federation) under the Open Science Policy

- **Put** a *"seed in the ground"* by operationalizing the first recognised EOSC Node at the European level for the initial 3 years

- **Offer** *"core services"* for scientific research infrastructures to federate (single-sign-on, catalogues, knowledge graph, application workflow, monitoring, accounting, helpdesk) and common *"horizontal services"* for end-users to benefit from (compute, containers, data transfer, notebooks, file sharing, open research data)

- **Define** the *pathway and blueprint* (EOSC Interoperability Framework) for other potential EOSC Node operators to join the federation
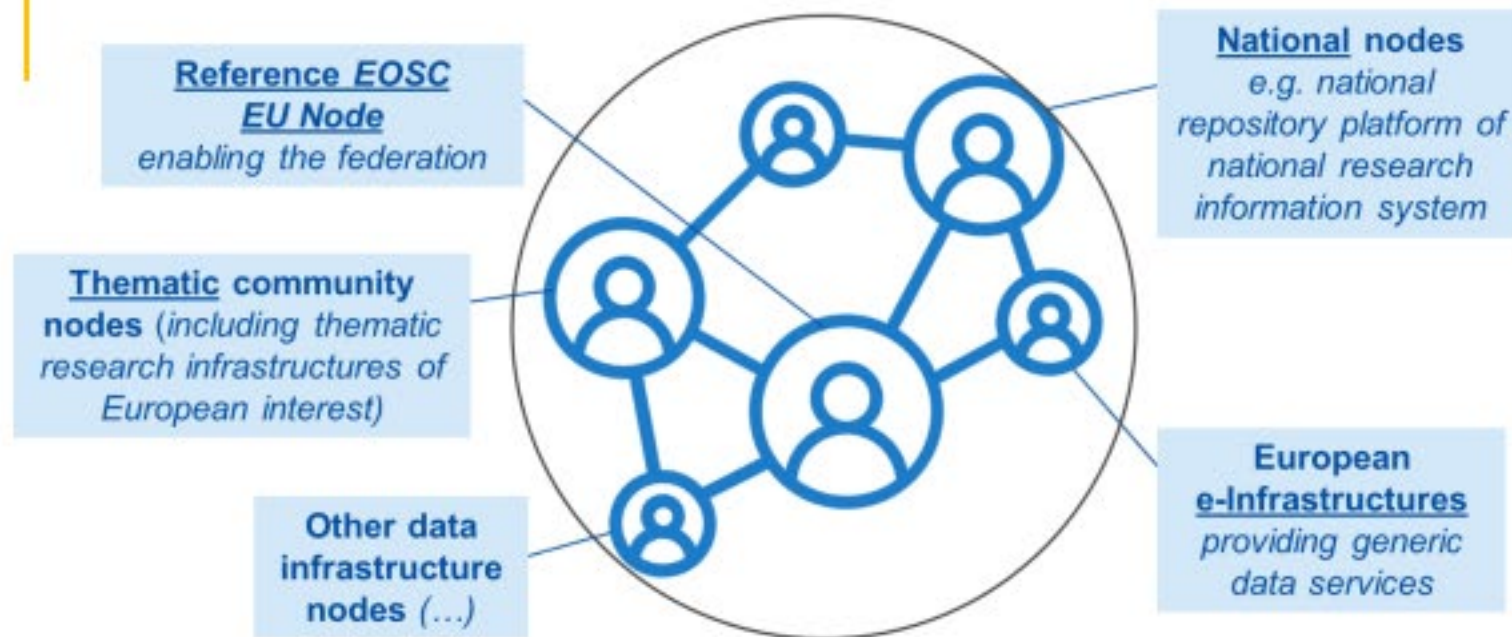
EOSC EU Node characteristics and features

- **European level multi-disciplinary and multi-national** scientific data/service portfolio for all research users (eduGAIN) and citizen scientists (EU Login/eIDAS)

- **For now, owned by the EC** and governed by the EOSC Tripartite Governance (EC, EOSC-A, MS/AC) Future ownership is under discussion.

- **Operated and maintained 24/7** by contracted third-parties (result of the EOSC Procurement Action) in production

- **SIMPL Agent proxy** to connect to other industrial Data Spaces

- **EuroHPC resources** may be offered to the EOSC Federation

- **Open concept: National, regional and/or thematic** service providers as well as autonomous EOSC Nodes can connect to the federation (established interoperability frameworks and policies)

Nodes

Federated system of systems

European Commission

Peter Szegedi (DG-CNECT)

About the EOSC Federation and EOSC Nodes

National nodes e.g. national repository platform of national research information system

Reference EOSC EU Node enabling the federation

Thematic community nodes (including thematic research infrastructures of European interest)

Other data infrastructure nodes (...)

European e-Infrastructures providing generic data services

EOSC policies and standards**: A baseline should be defined to ensure that each node can have a minimum working set of features and supports a minimum set of policies. It is important to mandate compliance with protocols and standards, but to give freedom to each node on how to support them.

EOSC Federation*: Open and trusted federation of collaborative, autonomous infrastructures applying agreed, consensus-based EOSC policies and rules of participation, combined into a system of systems to enable European researchers to store, share, process, analyse, and reuse research digital objects (e.g. data, publications and software)

EOSC Node*: Data infrastructure system of variable nature (national, regional, institutional or thematic) with consensus-based policies, transparent ownership and clear responsibility, connected to the EOSC Federation to share information and resources within the EOSC community and to leverage common services

* Source: "EOSC operations and evolution post-2027" supporting document by the EOSC-SB Policy subgroup (November 2023)
** Source: GEANT and NREN's position on EOSC Nodes (October 2023)

The connections in the Federation

Peter Szegedi (DG-CNECT)
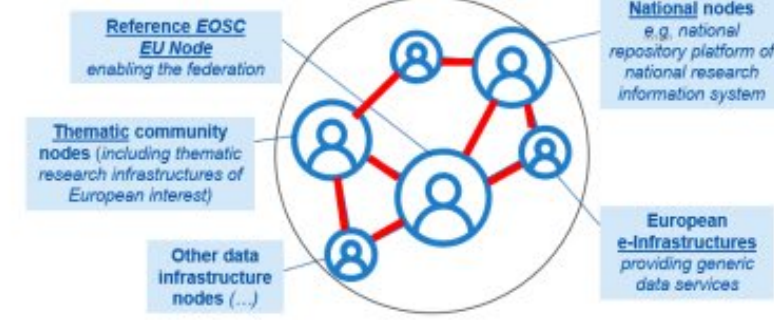
To address the EOSC puzzle

And cope with a vast quantity of infrastructure components of various scales and scopes, centralised or distributed, being generic, domain-specific or cross-disciplinary at institutional, national or European level.

- Minimal standards to be applied across the Federation (EOSC Interoperability Framework).
- Common policies and Rules of Participation applicable across the Federation
- A model to govern, enforce and evolve those policies.
- Trade-off between inclusiveness and quality of EOSC resources ?

The Use Policy affect all users (registered or anonymous) and defines what is acceptable use of EOSC. For example: lawful, respect of intellectual property, respect of security rules…

The Access Policy primarily regulates the use of resources (computing, data, network) and security rules for registered users.

The Rules of Participation affect all resource providers / Node operators. Define the minimal standards for any 3rd party service (ie: assets, support) provided within the Federation. Deeply related to onboarding.

Acceptable Use Policy

Access Policy

Rules of Participation (RoP)
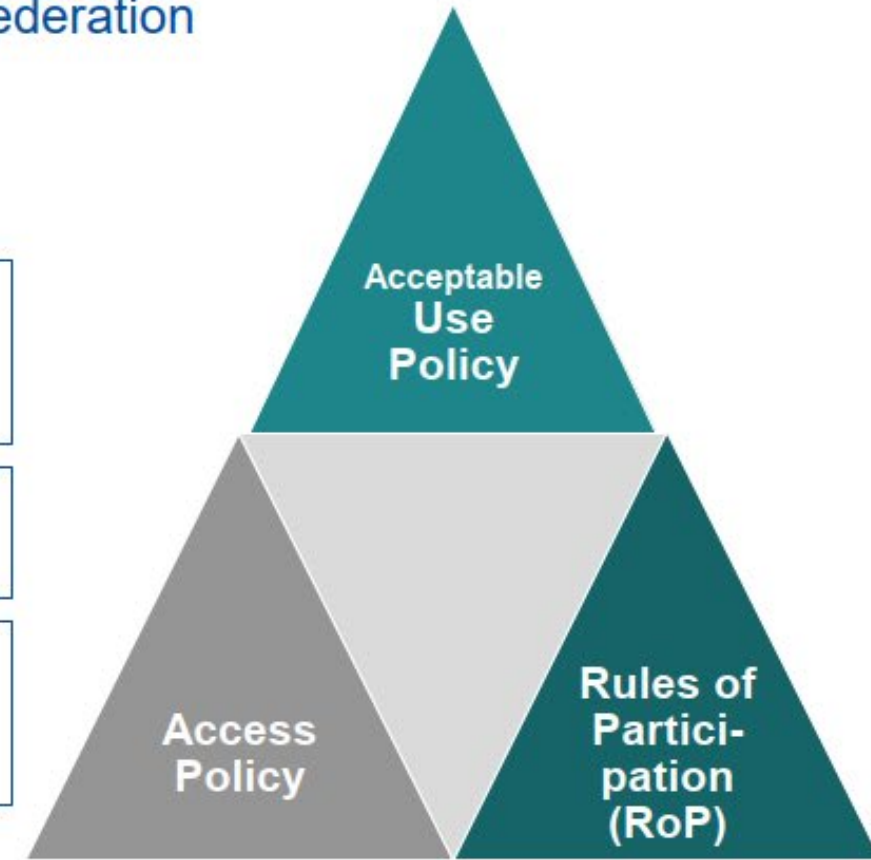
Further defining the EOSC Federation

- EOSC-A has drafted a document defining the EOSC Federation and its building blocks, EOSC Nodes

- EOSC-A will continue to lead on shaping the definition of the EOSC Federation and the role of nodes

- It is the ambition for 2024 to evolve and extend the paper into a formalised handbook of operations that provides policies, guidelines, recommendations and related procedures for the EOSC Federation.

- A key dependency to this evolution, and indeed that of the nodes concept itself, is the definition of access policies at both the EOSC Node and Federation levels

- Key to this evolution is the agreement on the future EOSC governance model (post 2027)

Some conclusions

- The awareness of the need for data and digital skills is emerging but needs to be stimulated.
- These are new and partly discipline specific skills, so basis training (Minimal Viable Skillset) as well-advanced knowledge (data stewards) are important.
- Collaboration of all data/software support units in the provision of training allows scalability and sustainability of training. It is relevant that training provision is part of the roles.
- The data stewards from organisations can jointly form a 'Data Competence Centre' (local or thematic)
- Linking 'Data Competence Centres' in countries and across countries strengthens the back-bone
- Creating in Europe a federated structure of nodes containing data and/or services will be the next step in realising EOSC

THANK YOU