



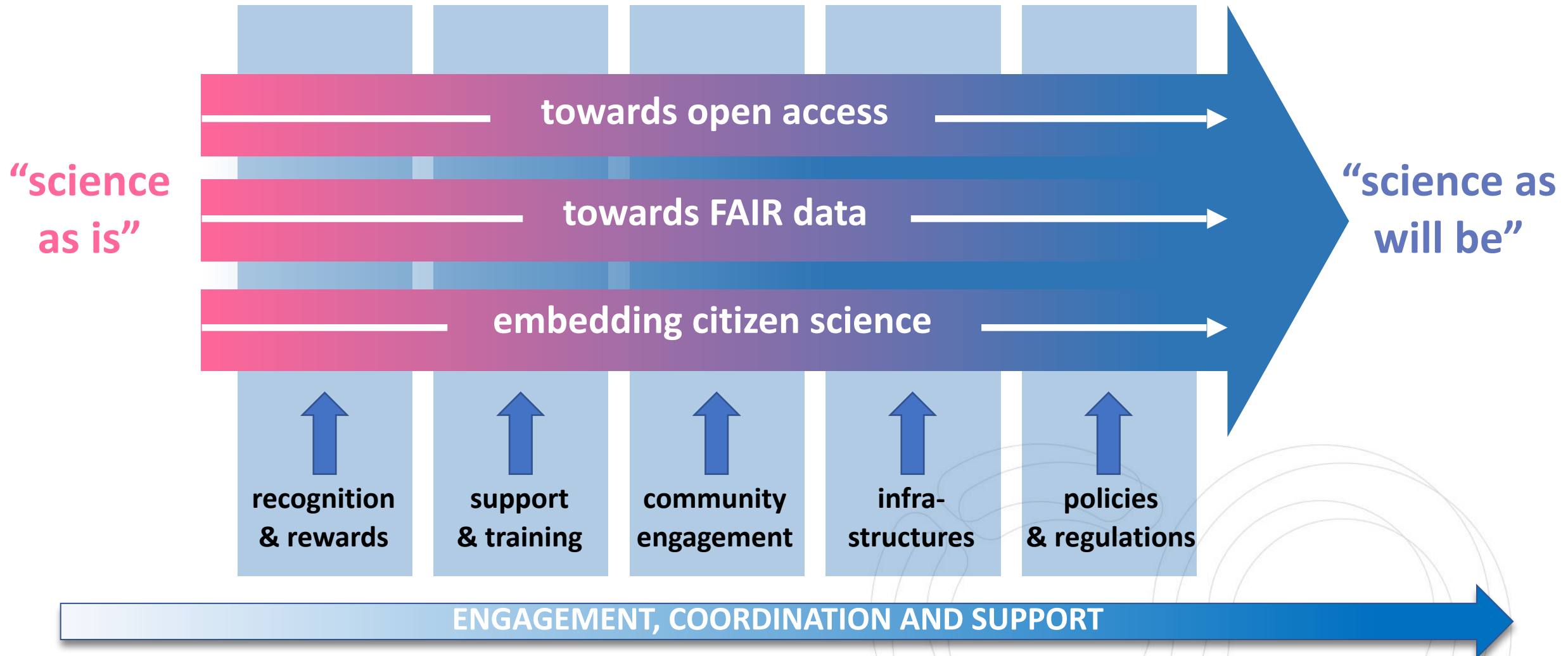
The Web of FAIR research data

Karel Luyben
President EOOSC
Association

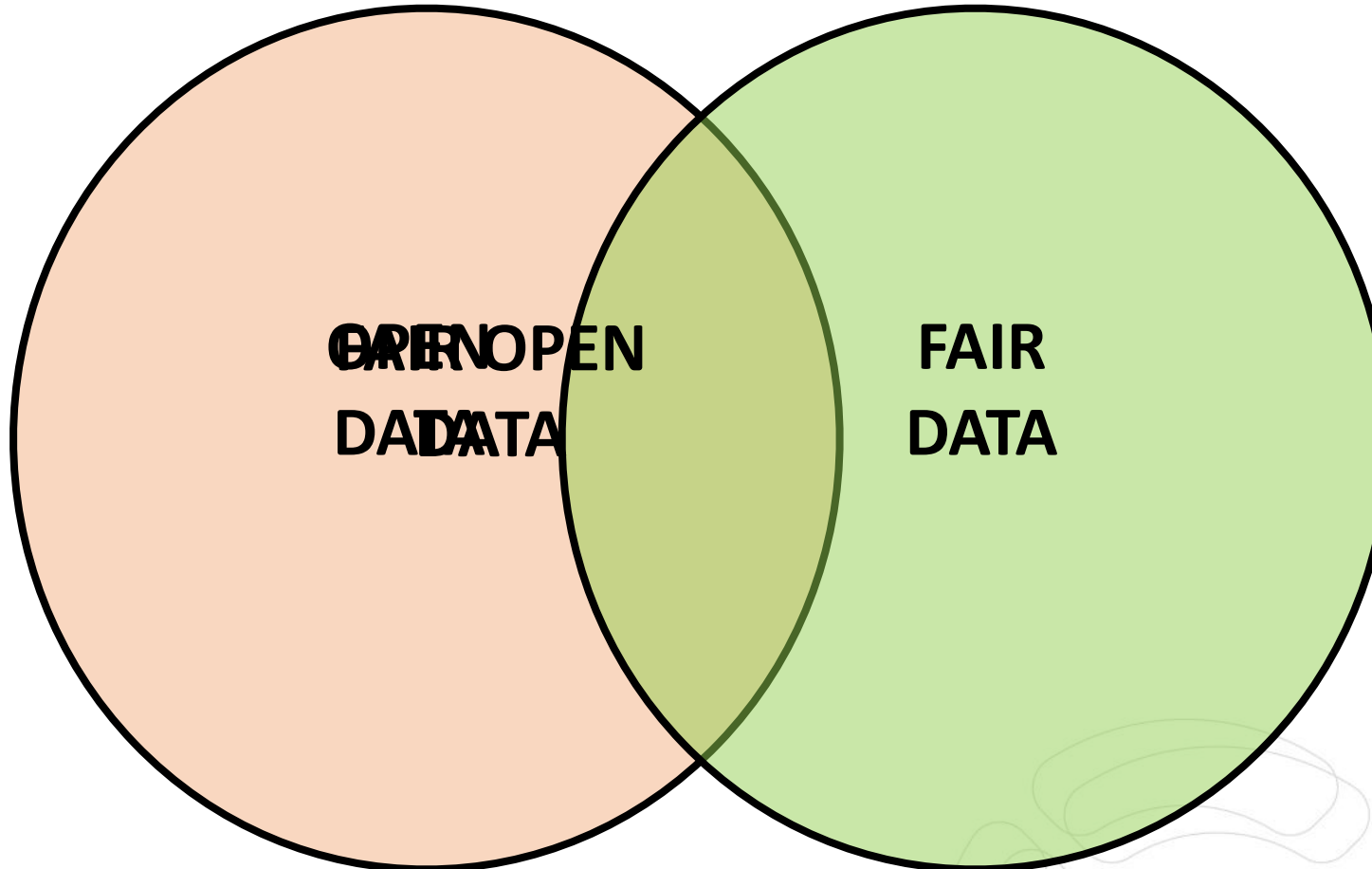
Polish Open Science Conference
2024
Krakow, 10-12 April 2024



Open Science



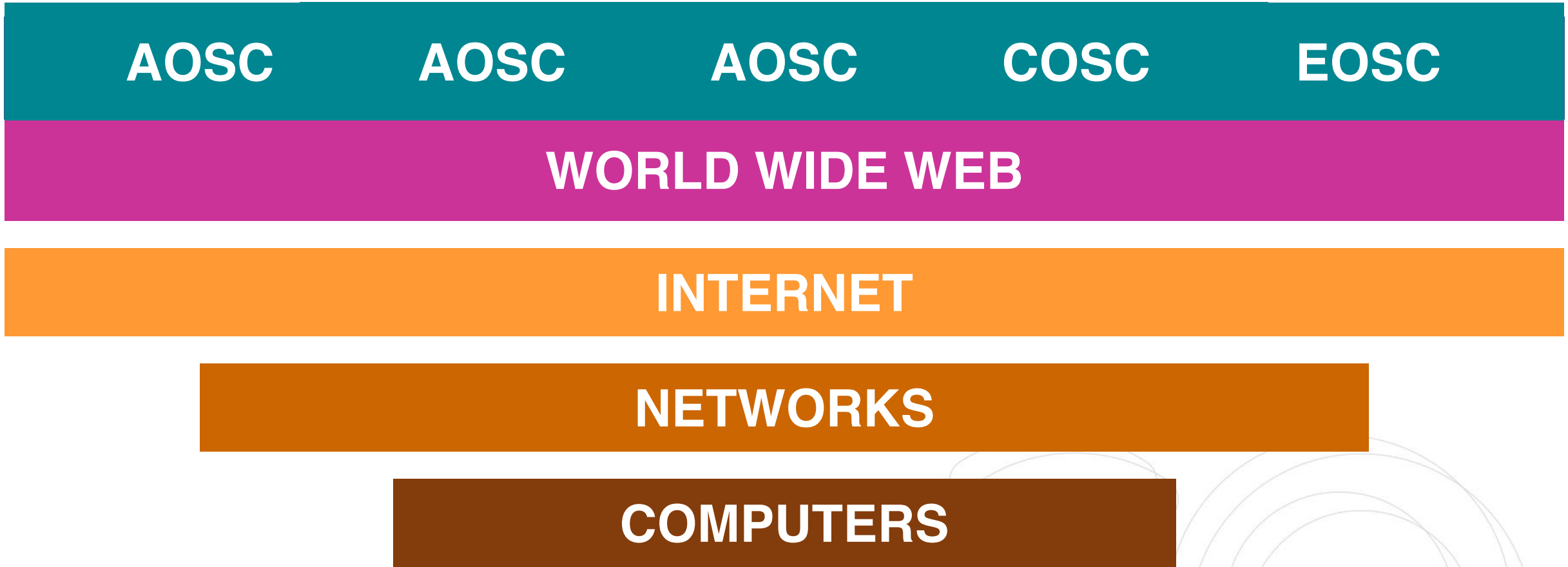
OPEN DATA and/or FAIR DATA



FAIR ≡
Findable
Accessible
Interoperable
Reusable

Towards “as FAIR as possible” and “as open as possible”

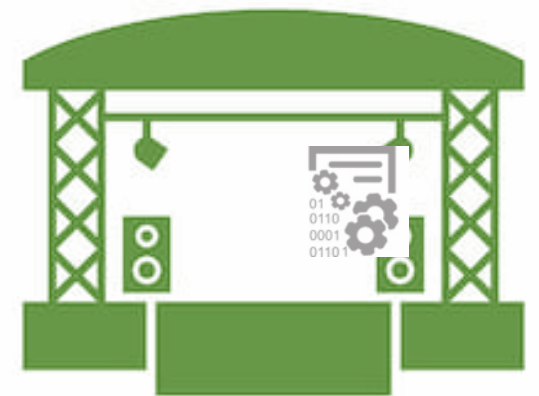
 **EOSC** – additionality to the web of FAIR data



[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

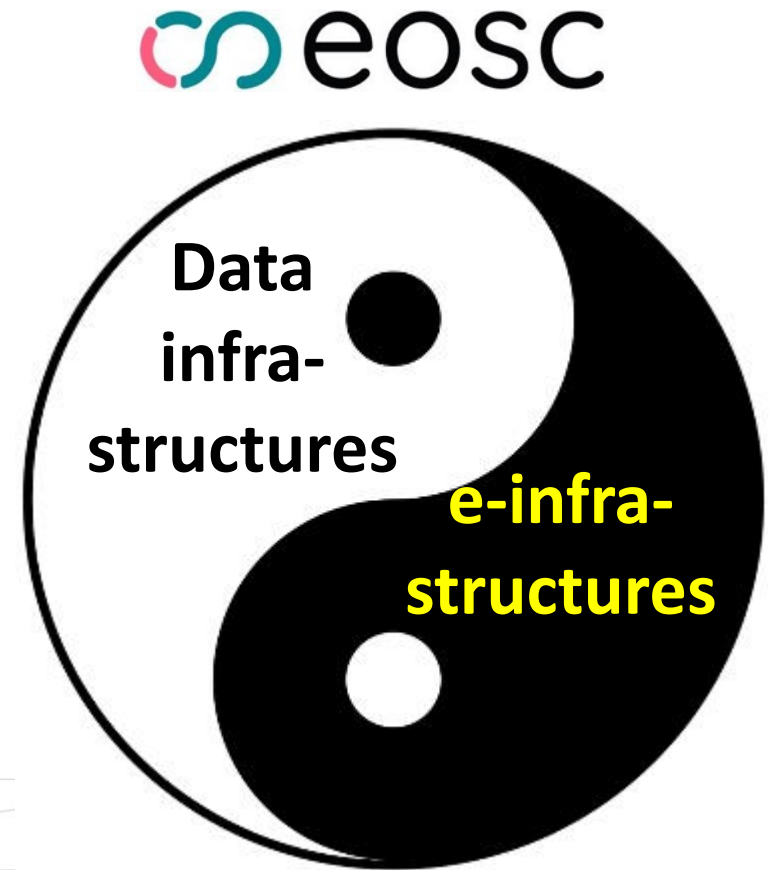


EOSC the twinning of data- and e-infrastructures

EOSC is to be an infrastructure and could be seen as the twinning of European data infrastructures and e-infrastructures.

The e-infrastructures offering the store, compute and connect services used by EOSC to offer the servicing of data and creating interoperability.

The combination forms the EOSC-ecosystem (yin / yang).



Ingredients for EOSC

- Hardware (available at many locations)
- Software (being put together for the first Node)
- Protocols for the federation
- EOSC Interoperability Board (EIB) as gatekeeper
- FAIR data as the most important content of the system
- Governance to make EOSC run smoothly



Guiding principles for EOSC

The **overarching** principle for developing EOSC is that research has to be at the centre of the EOSC initiative.

- **Multi-stakeholderism**

EOSC will succeed if and only if it follows a multi-stakeholder approach;

- **Openness**

EOSC will ensure research artefacts be ‘as open as possible, as restricted as necessary’;

- **FAIR principles**

EOSC research artefacts need to be findable, accessible, interoperable and reusable;

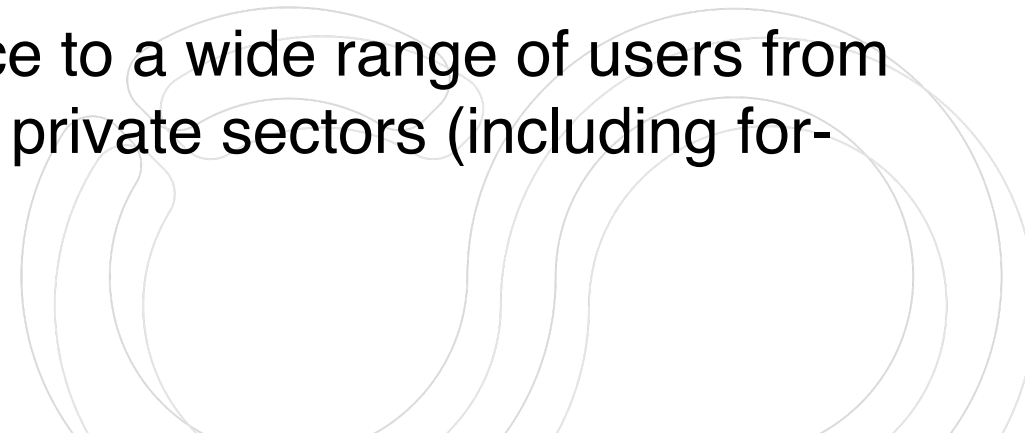
- **Federation of infrastructures**

EOSC will federate existing and upcoming data- and e-infrastructure;

- **Machine-actionable**

EOSC will ensure that machines can find, access and interoperate on data helping people in servicing the needs of European scientists.

Critical success factors for EOSC

- Research data produced by publicly funded research in Europe is FAIR as much as possible by design;
 - The EOSC Interoperability Framework supports a wide range of FAIR digital objects including data, software and other research artefacts;
 - EOSC is operational and provides a stable infrastructure, supporting researchers addressing societal challenges;
 - EOSC is populated with a valuable corpus of interoperable data;
 - The scope of EOSC is widened to serve the public and private sectors
 - EOSC is a valuable and valued resource to a wide range of users from the research and education, public and private sectors (including for-profit).
- 
- A decorative graphic in the bottom right corner consisting of several overlapping, light blue circles of varying sizes, creating a sense of depth and movement.

eossc **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**

Position of EOSC according to the European Commission

Taken from EC slides



“EOSC is the basis for a science, research and innovation data space that will bring together data resulting from research and deployment programmes and will be connected and articulated with the sectoral data spaces”

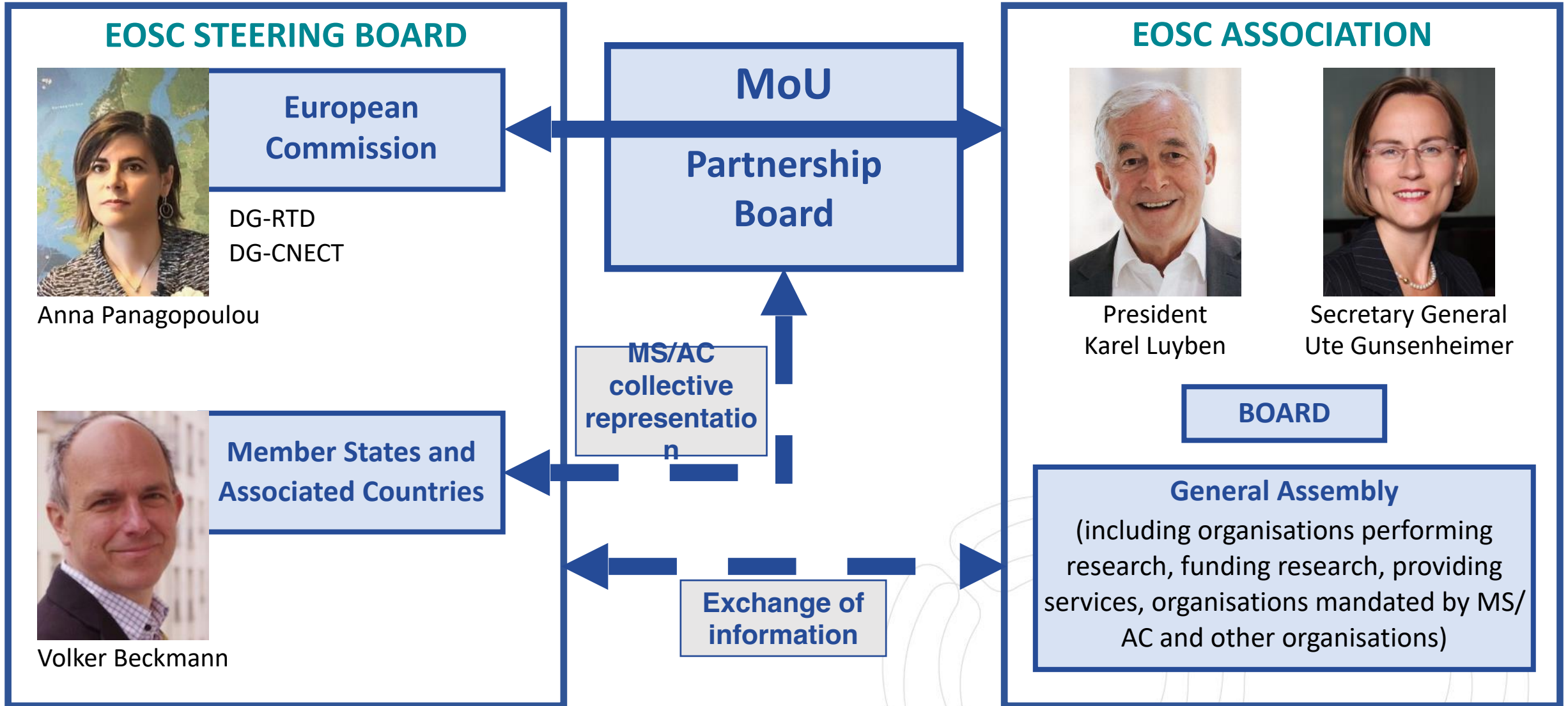
(European Data Strategy, COM(2020) 66 final)

Horizon Europe co-programmed EOSC Partnership

- Launched June 2021
- One billion euros commitment by the European Union and the EOSC Association



eosc European Open Science Cloud Partnership



Strategic Research and Innovation Agenda



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

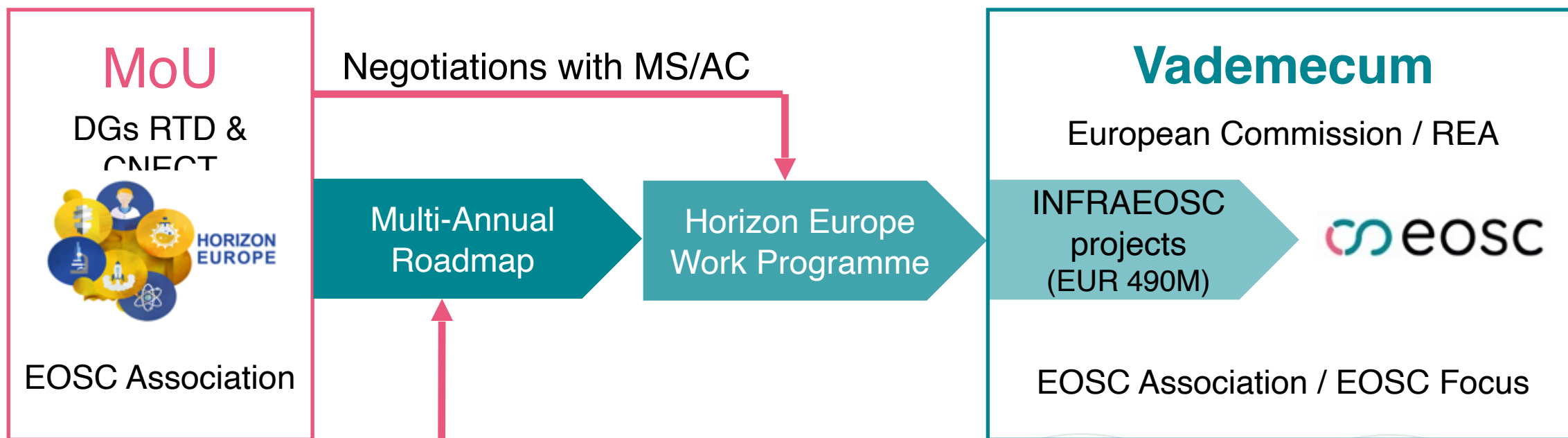
Multi-annual roadmap sets priority activities and outcomes grouped by three implementation levels – European, National, Institutional

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**

eosc Delivering EOSC – the Process

The critical role of the INFRAEOSC Projects



Community Consultation

Present action: MAR 2025 - 2027

Critical success factors for institutions

- Researchers are incentivised to perform Open Science;
- Researchers performing publicly funded research make relevant results available, as openly as possible;
- Research data produced by publicly funded research is FAIR as much as possible by design;
- Professional data stewards are available in research-performing organisations to support Open Science;
- EOSC becomes a valuable and valued resource to a wide range of users from research and education, public and private sectors



Looking at Research Performing Organisations (RPOs)



eosc Why Open Science and making data FAIR?

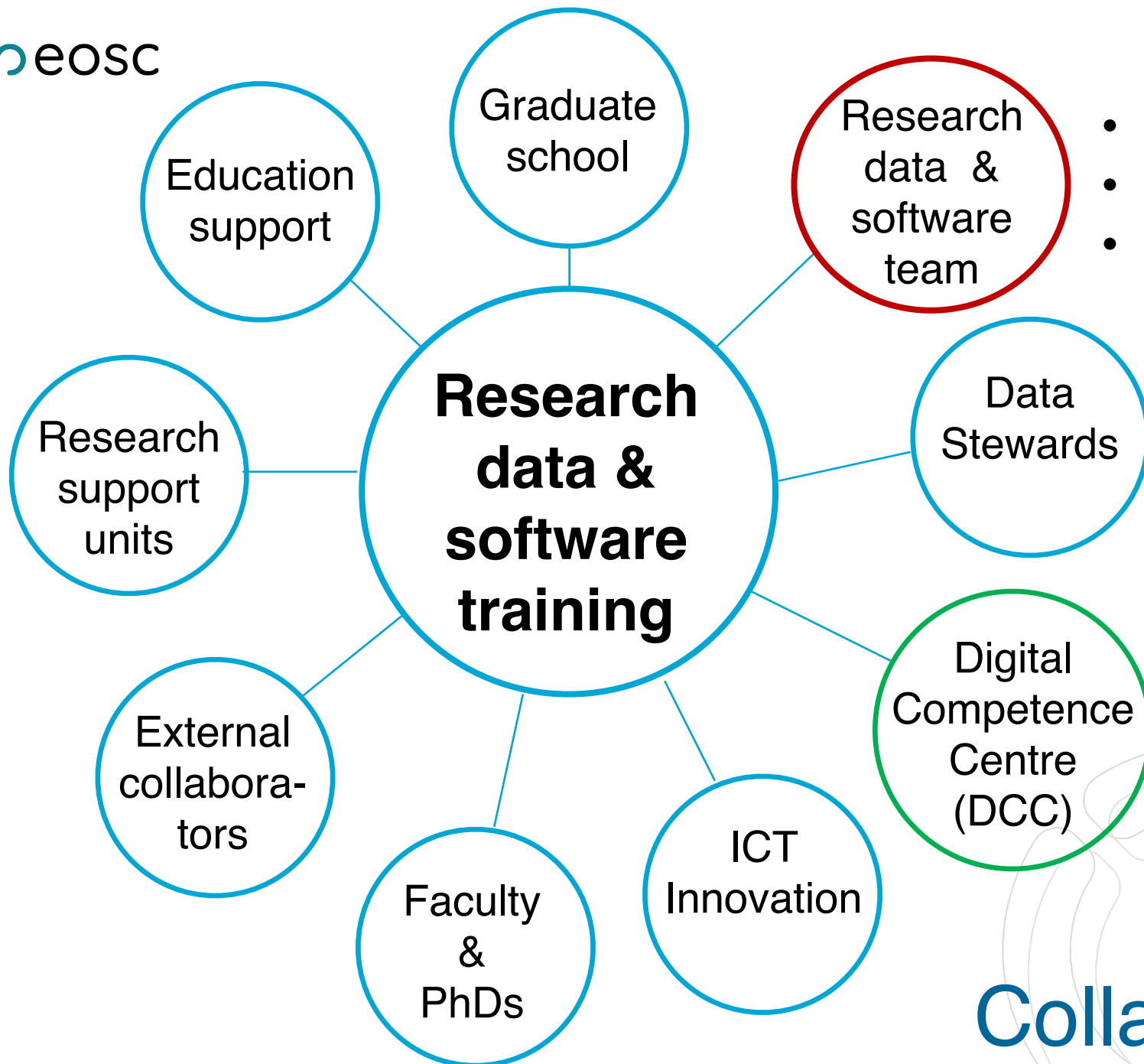
- Scientific integrity; public opinion on reliability of the scientific domain
- Democratising knowledge; equitability of data; positive image
- Diversity, equity, inclusivity (team science, academic career track)
- Quality in cooperation with partners (exchange of data / software)
- Part of the digital strategy (digitalisation of research and education)
- Helps in creating educational materials
- Open education helps the creation of new courses
- Not making data FAIR will cost a multitude of making them FAIR



eosc How Open Science and making data FAIR?

- Policy (top-down as well as bottom-up)
- Infrastructure (hardware / software; internal / external)
- Support (library and ICT – data stewards)
- Guidelines
- Training (FAIR etc.)
- Recognition
- Extra transition money





- **Creation**
- **Provision**
- **Coordination**

Convergence to FAIR ask for

- **Machine actionability**: the machine must (with the right request and search engine) be able to find the data; on the basis of the metadata be able to know 'what is possible with these data'; on the basis of the metadata be able to know 'what am I allowed to do with these data';
- **Metadata** that also determine if use and re-use is allowed with some or all of these data today and for what time in the future;
- Linking these FAIR data(sets) to **form a federation** of relevant existing and future data sources
- Combining these data sources with the services to create **EOSC**, and
- In interaction with other regions of the **world**

Equitability – challenges for RPOs

- Enable everyone to create FAIR-by design data (tailored to their field) through capacity-building, training, upskilling
- Recognize the value chain: FAIR data is generated, third parties benefit, third parties provide resources for FAIR data generation
- Make FAIR data infrastructure ‘as accessible as possible, as restricted as necessary’ for others to re-use
- Give credit and really “onboard” data “owners” and creators: define roles and responsibilities (credit and reward for all the players involved)
- Use the federating principles: **Ownership**, **Localisation** and **Regulatory compliance** (OLR) to let (as much as possible) the data stay where they are (data-visiting, not sharing)

eosc **Outlook to the future**



BACK- UP



FIP mini-questionnaire

use this form to make a first version of your FAIR Implementation Profile

Community description	Your answers
Name of Community	
Description of Community	
Supporting Links	
Research Domain	
Data Steward	
Date of FIP creation	

FAIR principle	FIP Question	FAIR Enabling Resource type	Your answers
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier service	
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier service	
F2	Which metadata schemas do you use for findability?	Metadata schema	
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata Data linking schema	
F4	In which search engines are your metadata records indexed?	Registry	
F4	In which search engines are your datasets indexed?	Registry	
A1.1	Which standardized communication protocol do you use for metadata records?	Communication protocol	
A1.1	Which standardized communication protocol do you use for datasets?	Communication protocol	
A1.2	Which authentication & authorisation technique do you use for metadata records?	Authentication & authorisation service	
A1.2	Which authentication & authorisation technique do you use for datasets?	Authentication & authorisation service	
A2	Which metadata longevity plan do you use?	Metadata longevity	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for metadata records?	Knowledge representation language	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for datasets?	Knowledge representation language	
I2	Which structured vocabularies do you use to annotate your metadata records?	Structured vocabularies	
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabularies	
I3	Which models, schema(s) do you use for your metadata records?	Metadata schema	
I3	Which models, schema(s) do you use for your datasets?	Data schema	
R1.1	Which usage license do you use for your metadata records?	Data usage license	
R1.1	Which usage license do you use for your datasets?	Data usage license	
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records?	Provenance model	
R1.2	Which metadata schemas do you use for describing the provenance of your datasets?	Provenance model	
R1.3	Who is the community, and what are their domain relevant community standards?	The FAIR implementation Profile	This FAIR implementation Profile

Link to this document (version 2.0): <http://bit.ly/FIPminiquestionnaire>
 GFF Interpretations of the FAIR Principles: <https://www.gfairfoundation/interpretation>
 FER Type Definitions: <https://csf.io/2/f9/ej>
 FER examples: <https://csf.io/3/an/2>