eosc

Institutional needs for FAIR research data and the route to SRIA 2.0
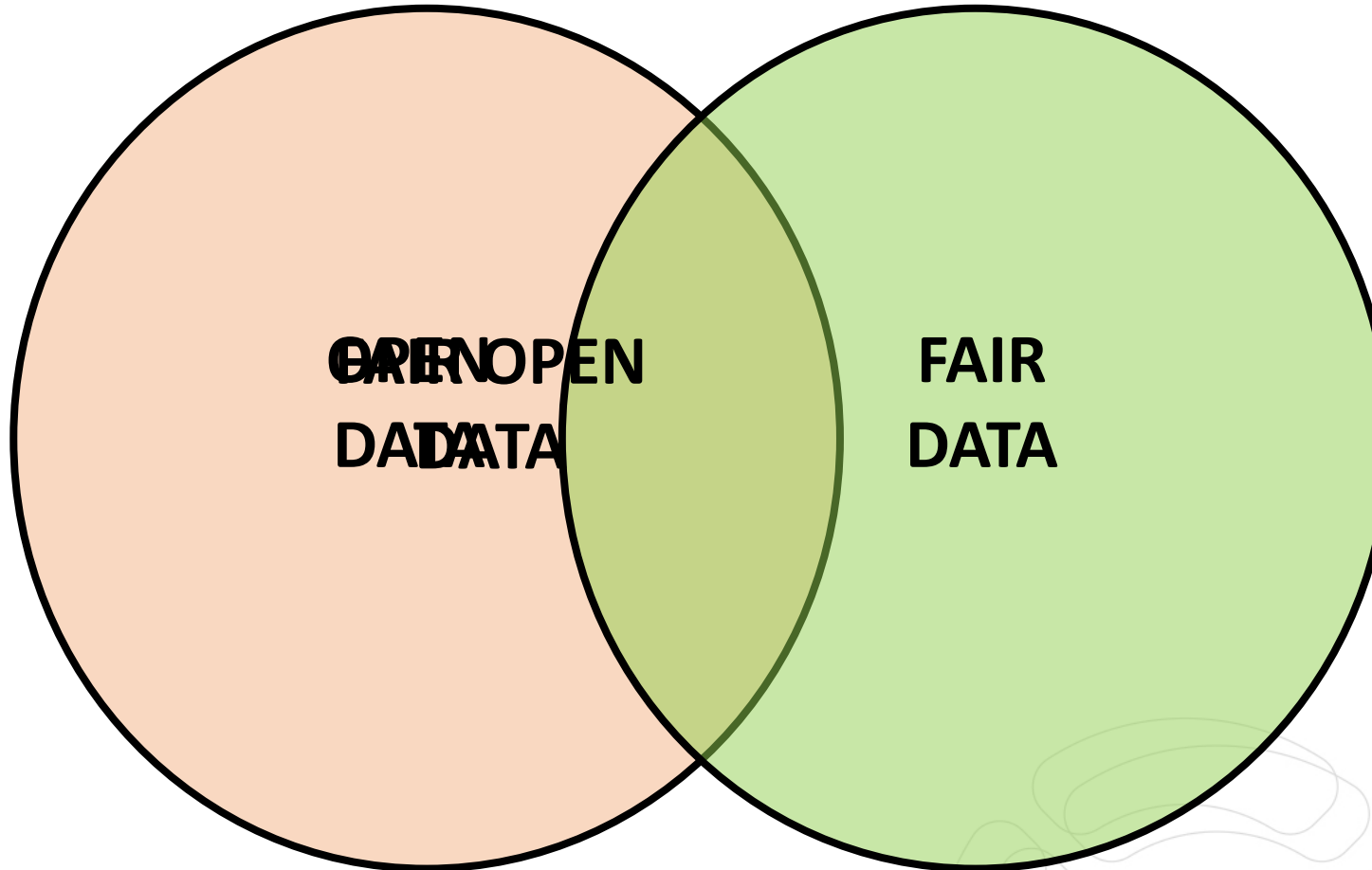
Karel Luyben
President EOSC Association

EOSC Winterschool, 29 January - 1 February 2024,
Thessaloniki, Greece

Why Open Science and making data FAIR?

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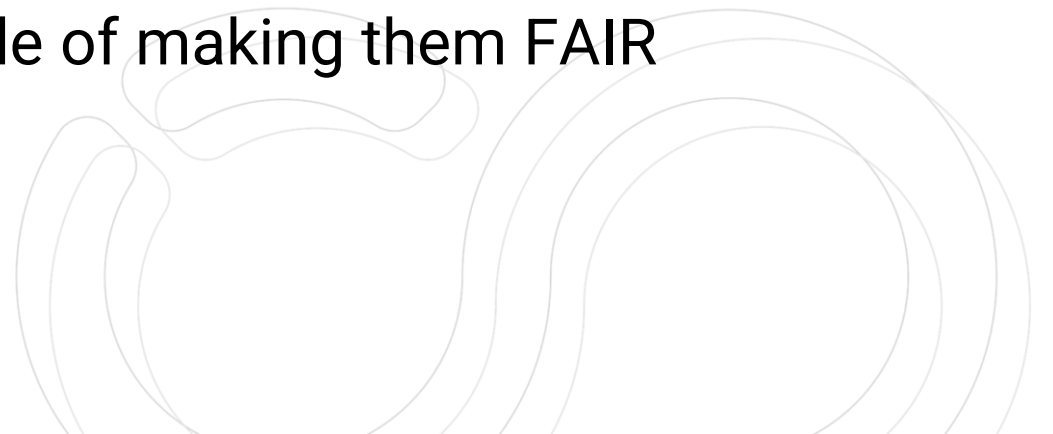
How Open Science and making data FAIR?

- 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

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

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)

- Include the Federated FAIR principles: **Ownership**, **Localisation** and **Regulatory compliance** (OLR) to let the data stay where they are (data-visiting, not sharing)

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*

Convergence to FAIR

- 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 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 allows to form a federation of relevant existing and future data sources

- For those contributing to and making use of EOSC, and

- In interaction with other regions of the world

EOSC – additionality to the web of FAIR data

| AOSC | AOSC | AOSC | COSC | EOSC |

WORLD WIDE WEB

INTERNET

NETWORKS

COMPUTERS
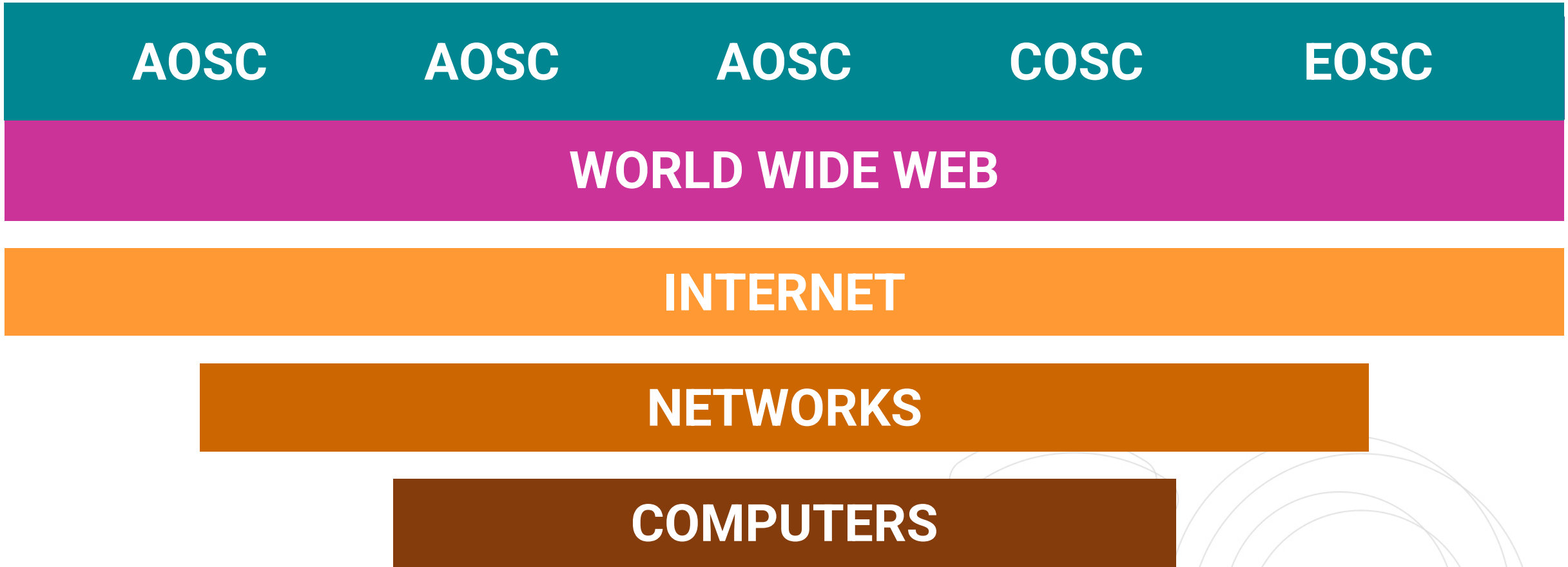
Modelled after: World Wide Web - Wikipedia

"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**

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

Protocols for EOSC

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...

from Peter Szegedi
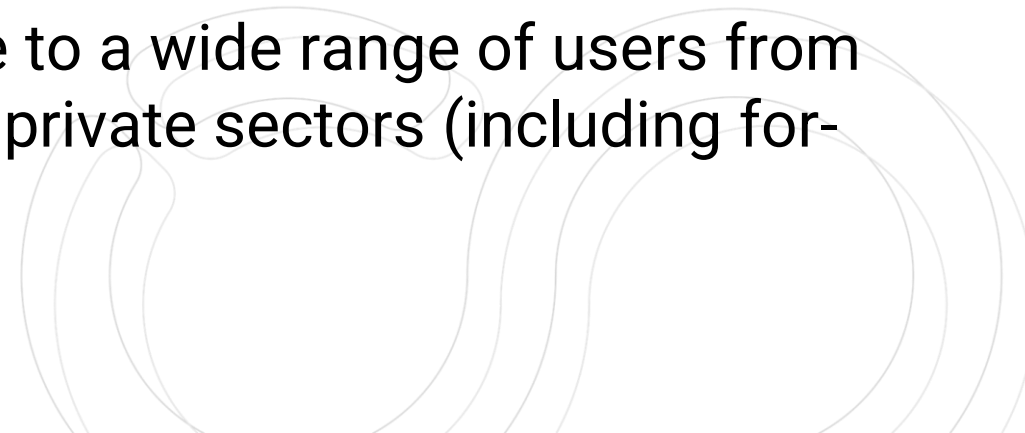
EOSC Interoperability Board

Taking the distributed and loosely coupled component character of EOSC as a reality. The major challenge is to functionally link, rather than physically integrate, all valuable resources (FAIR data and services).
The *access* to these services needs to be as easy as possible for both humans and machines and will likely be customised per research community
Therefore, an EIB is needed to ensure that all components of value to the EOSC ecosystem are findable through FAIR metadata (both data resources and services, such as algorithms, workflows, data orchestration software, and middleware components);
have a clear access protocol associated with these metadata; and become **machine actionable** and interoperable in the larger ecosystem, with optimal freedom to operate granted to both providers and users.

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).

Draft content of SRIA 2.0 (I)

- Introduction: goal;
                    history;
                    present;
                    future
- EOSC description: ingredients
- Federation: hardware;
                    software;
                    protocols
- FAIR data and metadata: search;
                             accessibility;
                             interoperability;
                             rules for reuse

eosc

Draft content of SRIA 2.0 (II)

- Services: core;
  horizontal;
  funding
- Governance: entity responsible for federation;
  involvement of countries
  involvement of stakeholders
- Full AI Ready:

AI readiness is needed

- AI for FAIR
  - How do we use AI to help structure and standardize the (meta)data?
  - How can AI help us collect new information (e.g., provenance, rights, etc.)

- FAIR for AI
  - AI wildly increases the demand for data and the need for computing power and storage
  - FAIR plus AI increases the risk of re-identification (of people and other protected entities)

- AI trained on un-FAIR data multiplies potential inaccuracies and inequities
  - Hides provenance trail
  - Looks really convincing