

European Open Science Cloud (EOSC)

Karel Luyben

Plan-E, Rome
15 April 2019



Open Science

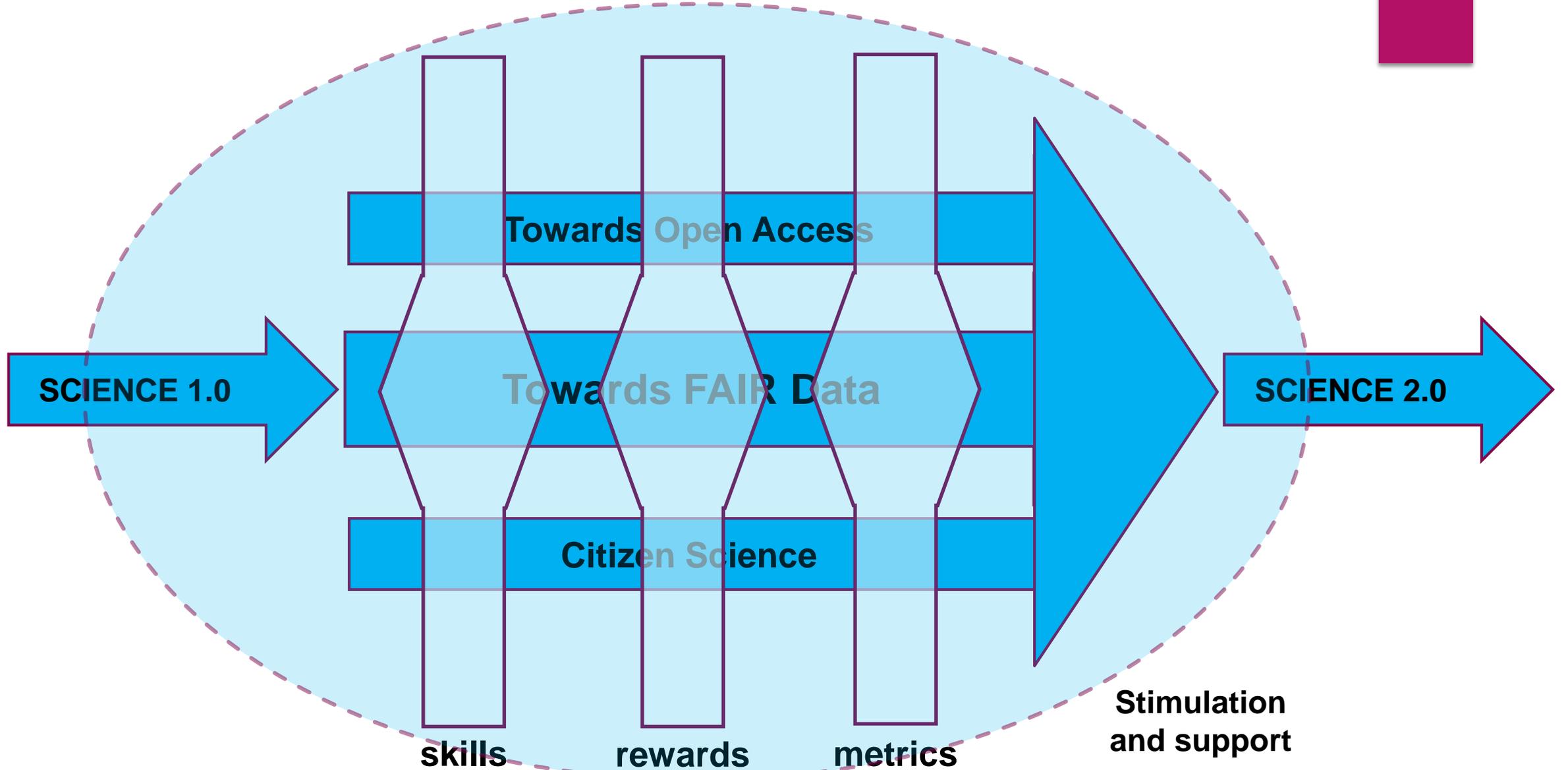


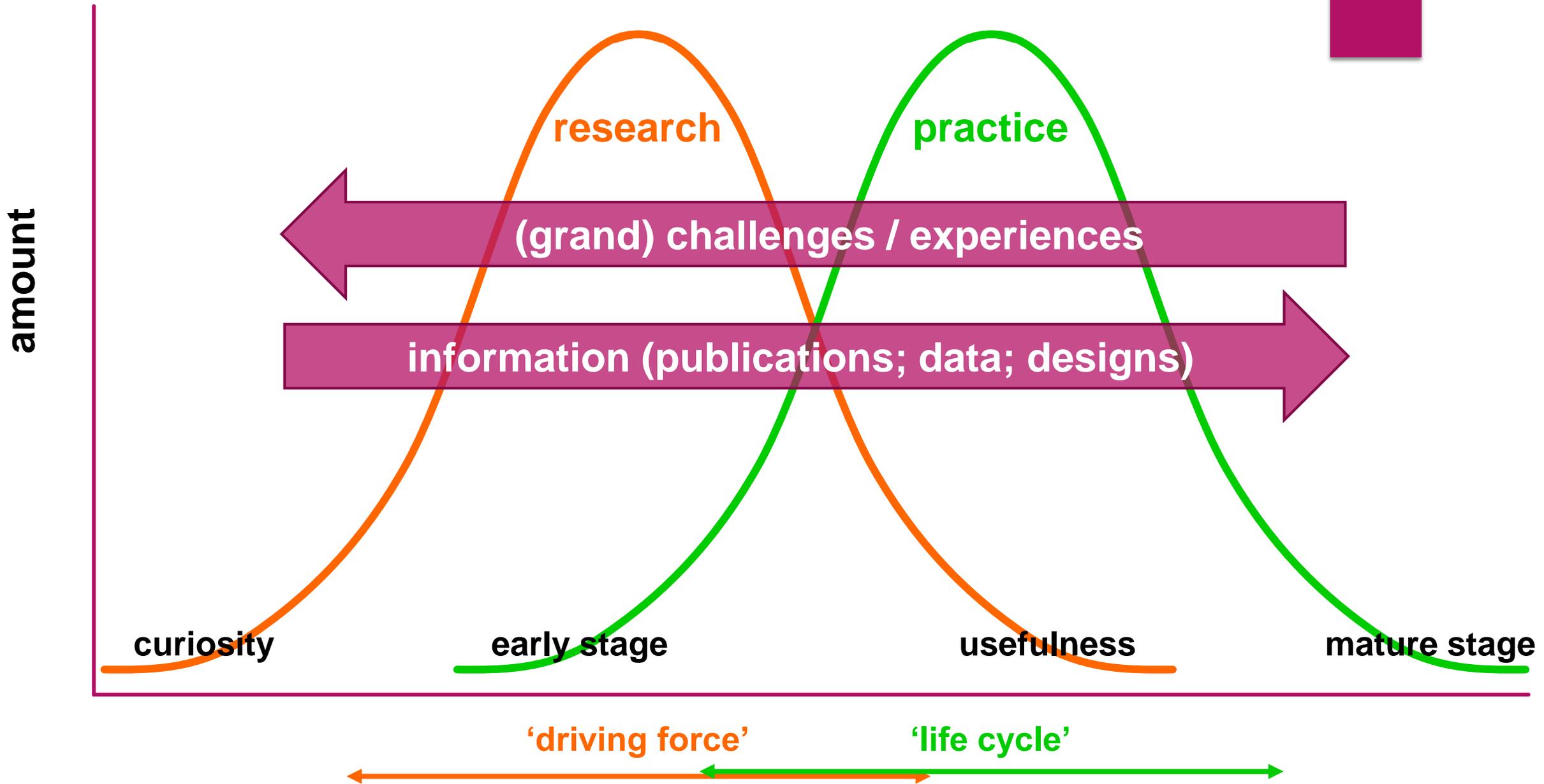
Open Science represents a **new approach** to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools.

Commissioner Carlos Moedas

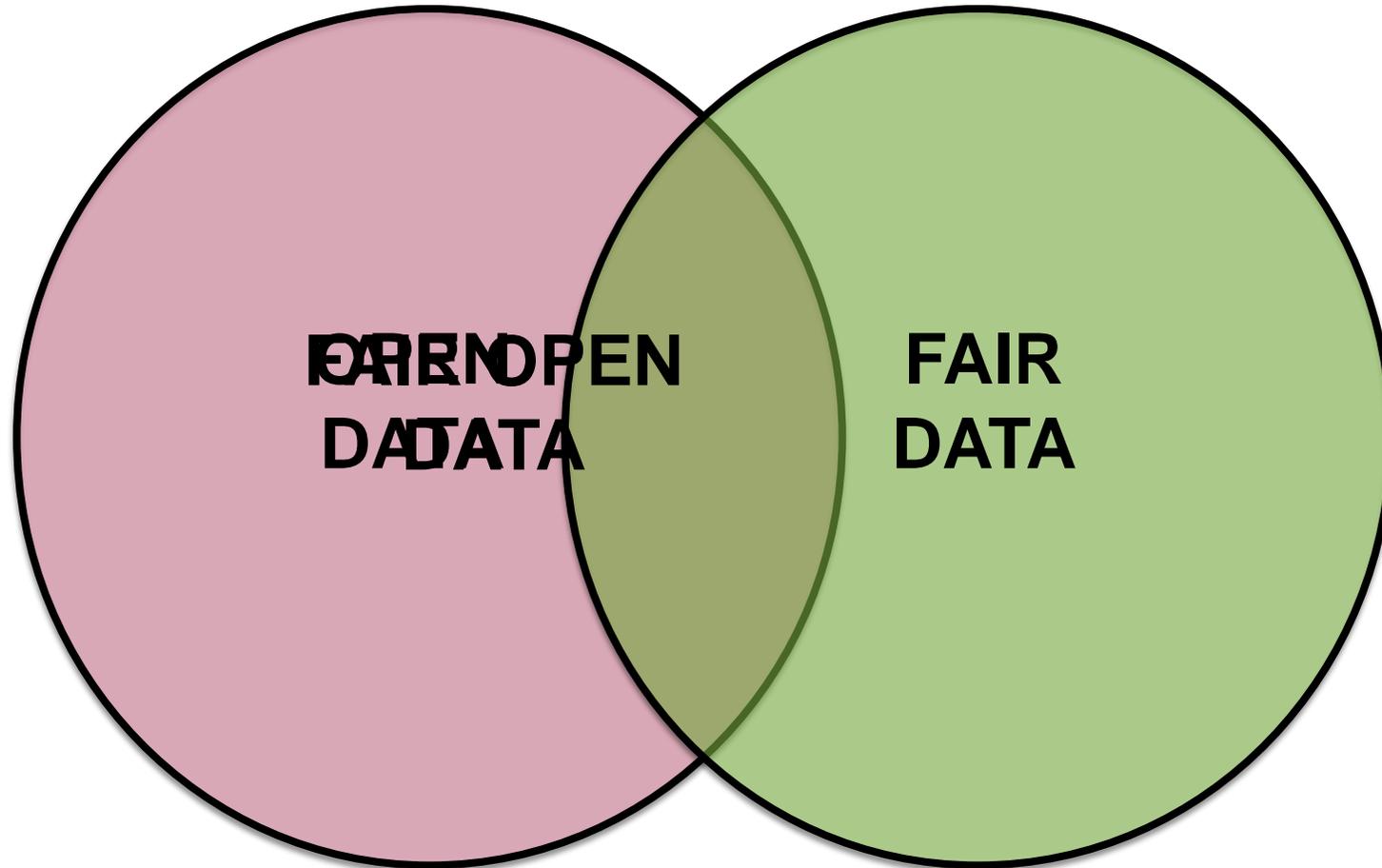
Open Innovation, Open Science,
Open to the World (2015)

OPEN SCIENCE





OPEN DATA OR FAIR DATA



FAIR = Findable; Accessible; Interoperable; Reusable

What FAIR is not ...

Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud DOI: 10.3233/ISU-170824

FAIR is **not** a standard, it's a guiding principle

FAIR is **not** only 'Semantic Web'

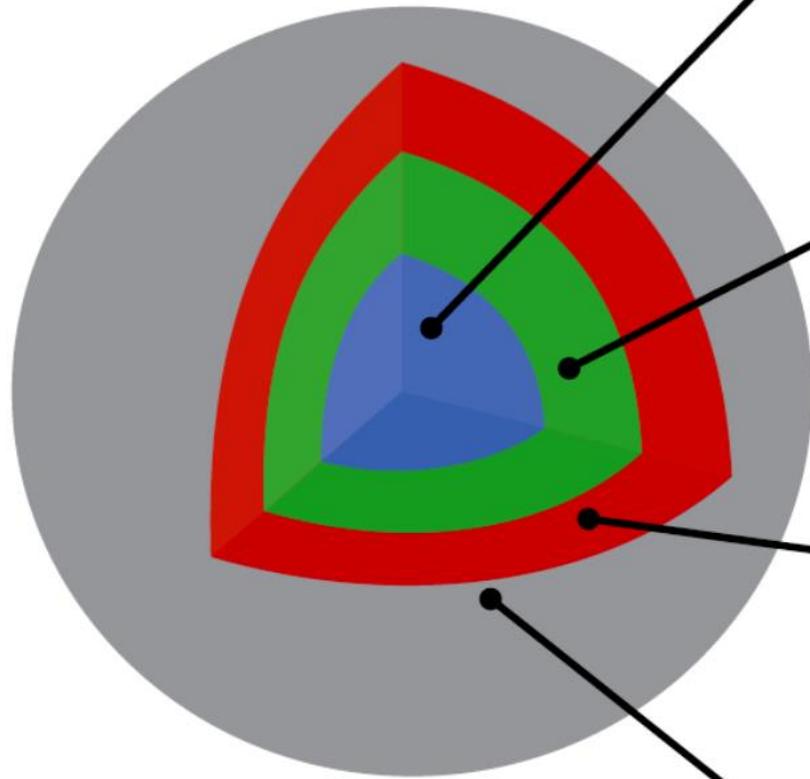
FAIR does **not** mean 'Open' or 'Free'

*Data are often Open but **not** FAIR*

*Data could be **non-Open** yet perfectly FAIR*

*Many data can **never** be Open*

FAIR principles do **not directly** prescribe data quality, trustworthiness, ethics or responsibilities.



DIGITAL OBJECT

Data, code and other research outputs

At its most basic level, data or code is a bitstream or binary sequence. For this to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and documentation. These layers of meaning enrich the object and enable reuse.

IDENTIFIERS

Persistent and unique (PIDs)

Digital Objects should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and supports citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCIDs), projects (RAIDs), funders and associated research resources (RRIDs).

STANDARDS & CODE

Open, documented formats

Digital Objects should be represented in common and ideally open file formats. This enables others to reuse them as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code use to process and analyse the data.

METADATA

Contextual documentation

In order for Digital Objects to be assessable and reusable, they should be accompanied by sufficient metadata and documentation. Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the objects were created. To enable the broadest reuse, they should be accompanied by a plurality of relevant attributes and a clear and accessible usage license.

European Data Market



Data workers

6.16 million in 2016

→ 10.43 million by 2020



Data companies

255,000 in 2016

→ 359,050 by 2020



Data economy value

Almost € 300 billion in 2016 → € 739 billion by 2020



European
Commission

Source: European Data Market study

Cost-Benefit analysis for FAIR research data



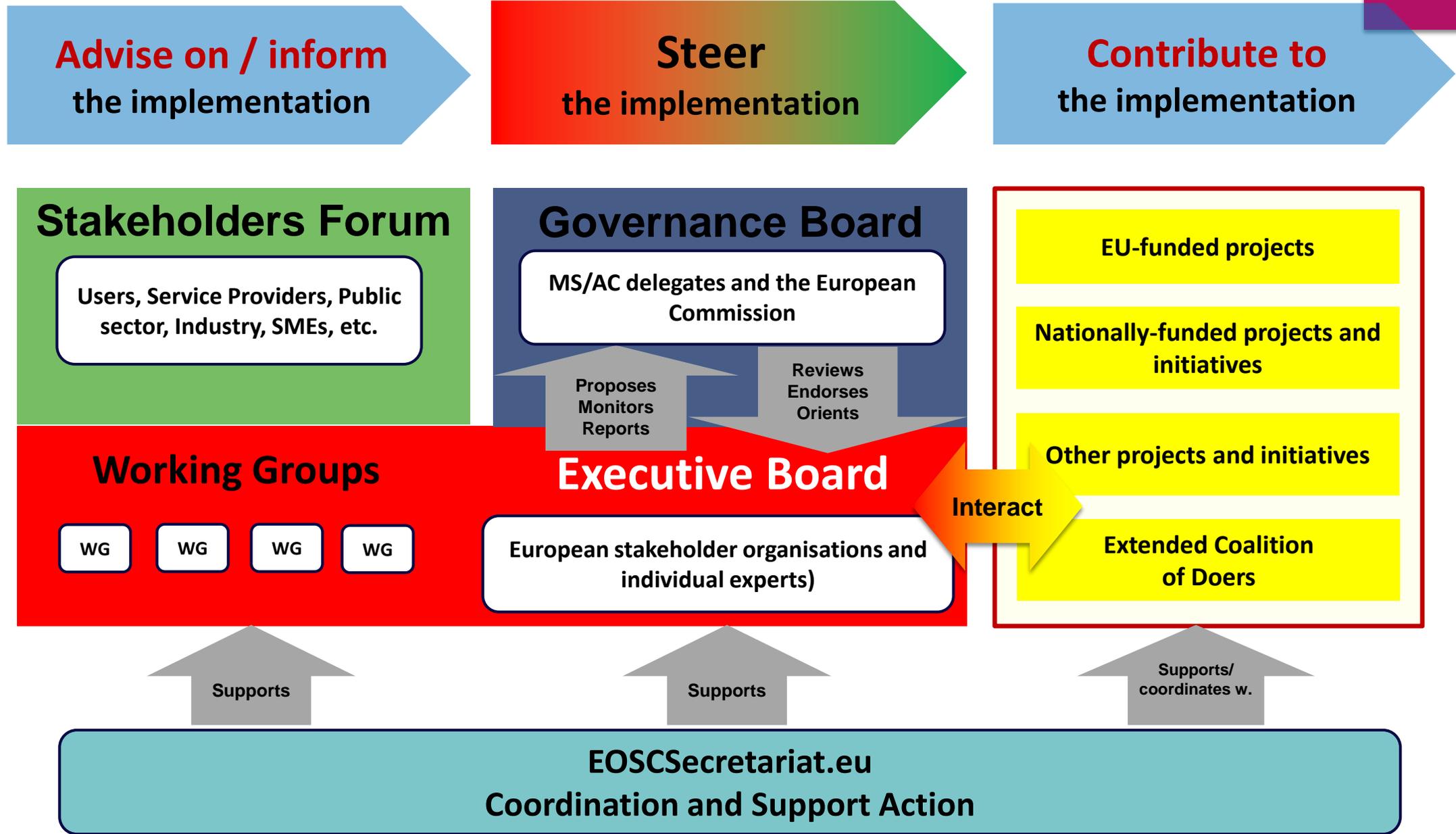
To drive the implementation of the FAIR principles in Europe, the European Commission together with a number of pioneering European research stakeholders is taking measures to raise the awareness about costs and benefits of FAIR data, and is encouraging funding bodies to set guidelines or support the development of an infrastructure for publishing FAIR data. In a study which preceded this report, the cost of not having FAIR data for the EU-28 has been estimated at **EUR 10,2 bn** per year, and this is bound to grow unless action is taken.

The EOSC is projected to become a reality by 2020 and will be Europe's virtual environment for all researchers to store, manage, analyse and re-use data for research, innovation and educational purposes.

European Open Science Cloud
New Research & Innovation Opportunities



Governance of EOSC 2019-2020



EB-Working Groups EOOSC

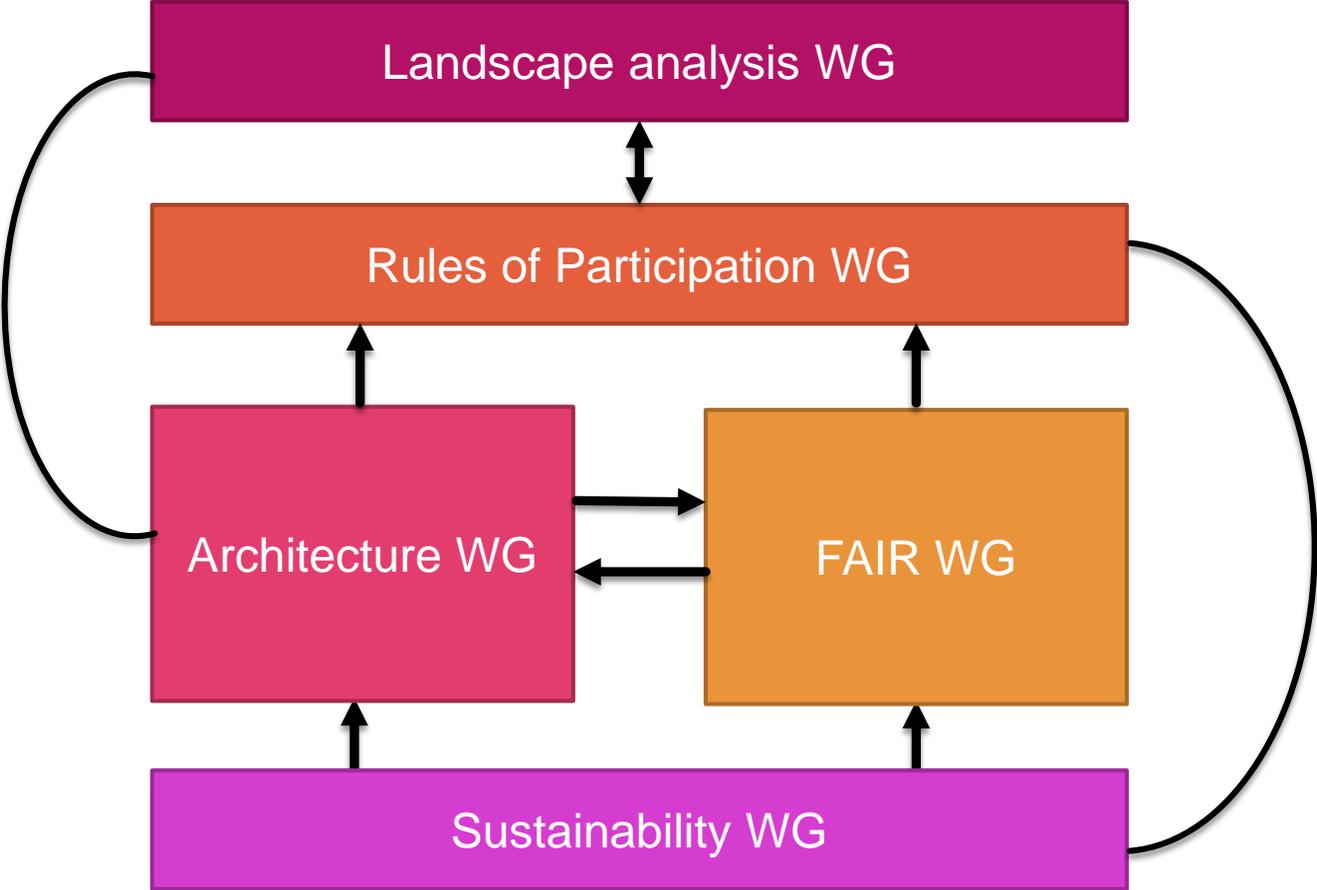
- **FAIR**
- **Architecture**
- **Rules of participation**
- **Landscape**
- **Sustainability**

What are our timeframes and contingencies?

Working Group	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Landscape			Mapping					
Rules of Participation				Initial RoP				New RoP
Architecture								
FAIR		Work Plan			Work Plan			
Sustainability				Initial Plan				Final Plan

- They all run till end 2020!
- Further map the deliverables and contingencies
- Define overall plan which maps out what has to come when

Relation between the EOSC WGs?





BACK-UP

Managing group interactions



Landscape WG

- **RoP:** align with MS initiatives
- **FAIR:** MS structures are FAIR
- **Sustainability:** match MS preparedness

Rules of Participation WG

- **Landscape:** align with MS initiatives
- **Sustainability:** compatible legal entity
- **Architecture :** technology facilitates contributions
- **FAIR:** services support FAIR and are FAIR

Architecture WG

- **RoP:** tech architecture facilitates contributions
- **FAIR:** tech architecture meets FAIR requirements

FAIR WG

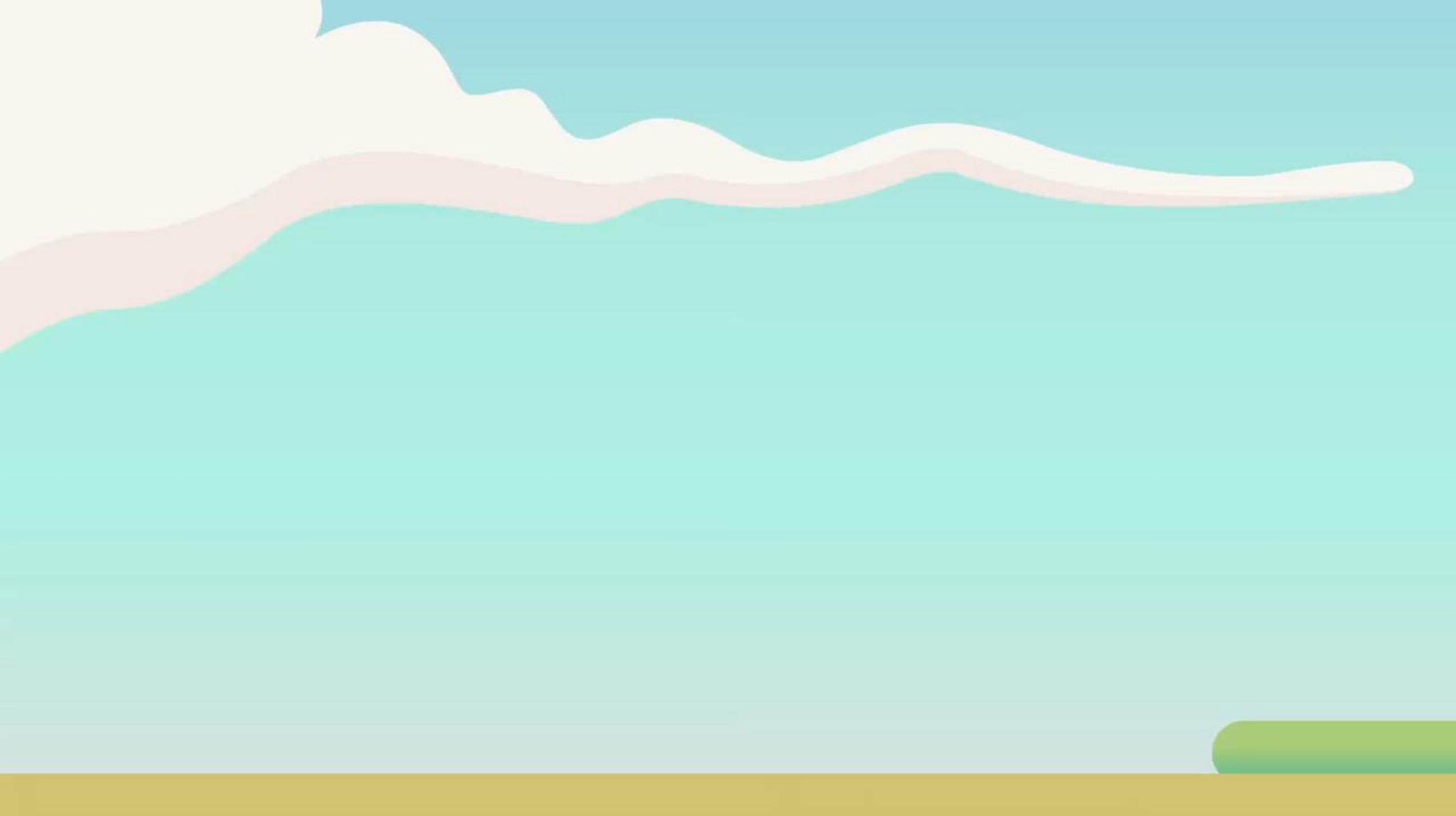
- **Landscape:** MS structures are FAIR
- **RoP:** services support FAIR and are FAIR
- **Architecture :** technology facilitates contributions

Sustainability WG

- **Landscape:** match / be compatible with MS preparedness
- **RoP:** compatible legal entity

The image features a light gray background with two thick, dark purple horizontal lines. The top line is positioned near the top of the frame, and the bottom line is near the bottom. A thin, light gray diagonal line starts from the top left corner and extends towards the top purple line. Another thin, light gray diagonal line starts from the bottom right corner and extends towards the bottom purple line. The text 'Personal Health Train' is centered between these two purple lines.

Personal Health Train



FAIR DATA PRINCIPLES

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a **searchable** resource

To be Accessible:

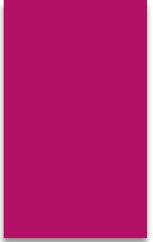
- A1. (meta)data are retrievable by their identifier using a **standardized communications protocol**
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a **formal, accessible**, shared, and broadly applicable language for knowledge representation
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards



EU-reports on datamarket

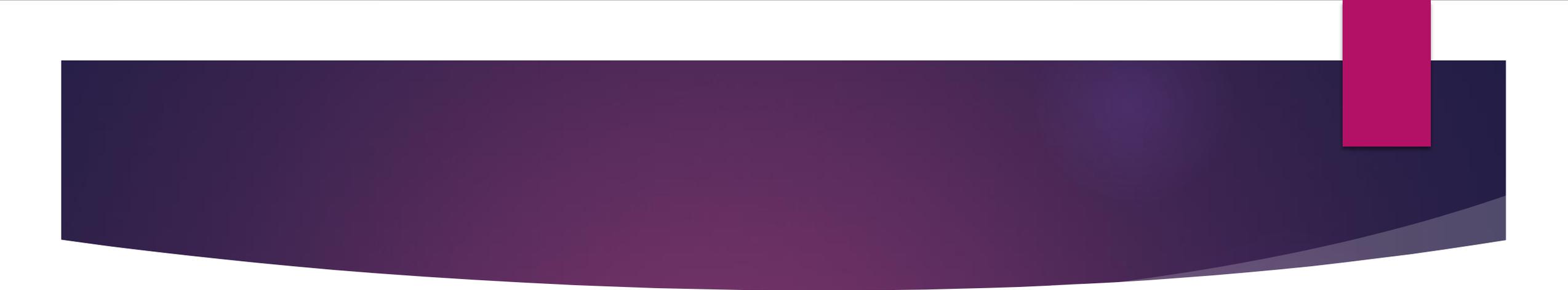
http://datalandscape.eu/sites/default/files/report/EDM_D2.1_1stReport-FactsFigures_revised_21.03.2018.pdf

<https://ec.europa.eu/digital-single-market/en/news/recommendation-access-and-preservation-scientific-information>

http://datalandscape.eu/sites/default/files/report/Opening_Up_Scientific_Data_for_Innovation-Story_2_FinalNewFormat.pdf

http://datalandscape.eu/sites/default/files/report/EDM_D2.2_First_Report_on_Policy_Conclusions_20.04.2018.pdf

<https://ec.europa.eu/digital-single-market/en/guidance-private-sector-data-sharing>



All authors must make their data and software (i.e. excluding, if relevant, data owned by third-parties, etc) appearing in their open access publications FAIR (Findable, Accessible, Interoperable and Reusable). To this end, a key requirement is deposition in a trusted repository that adheres to FAIR principles. In addition, all publications must include a statement of FAIR compliance for the source data underpinning their claims and the licence for its reuse.