



# Turning FAIR into Reality

## Report and Action Plan

Simon Hodson

Executive Director, CODATA

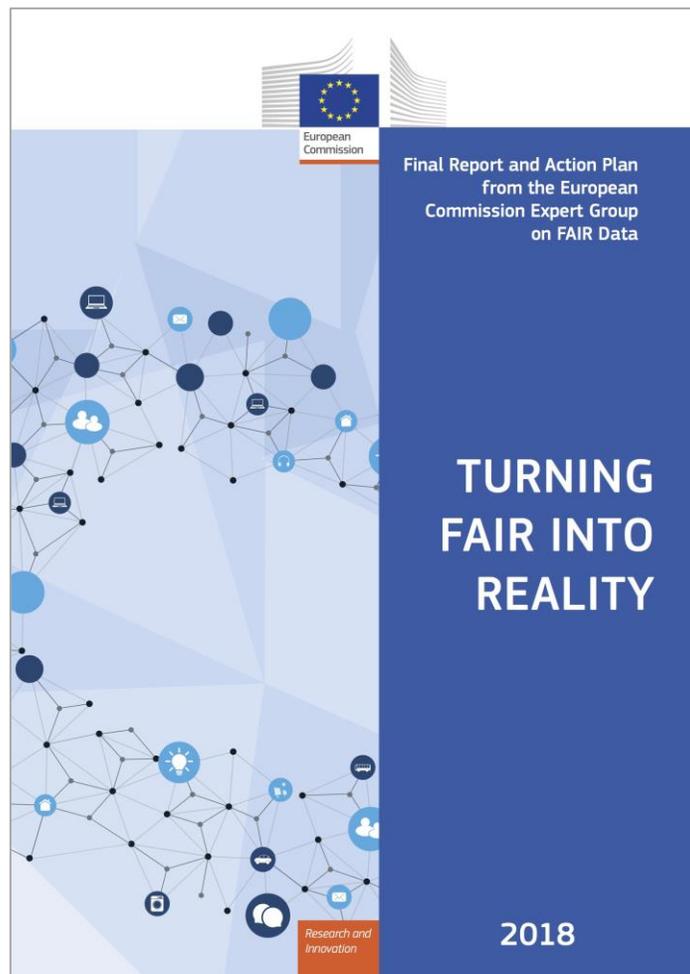
Chair of EC Expert Group on FAIR data

[simon@codata.org](mailto:simon@codata.org)

Twitter: [@simonhodson99](https://twitter.com/simonhodson99)

Research and  
Innovation





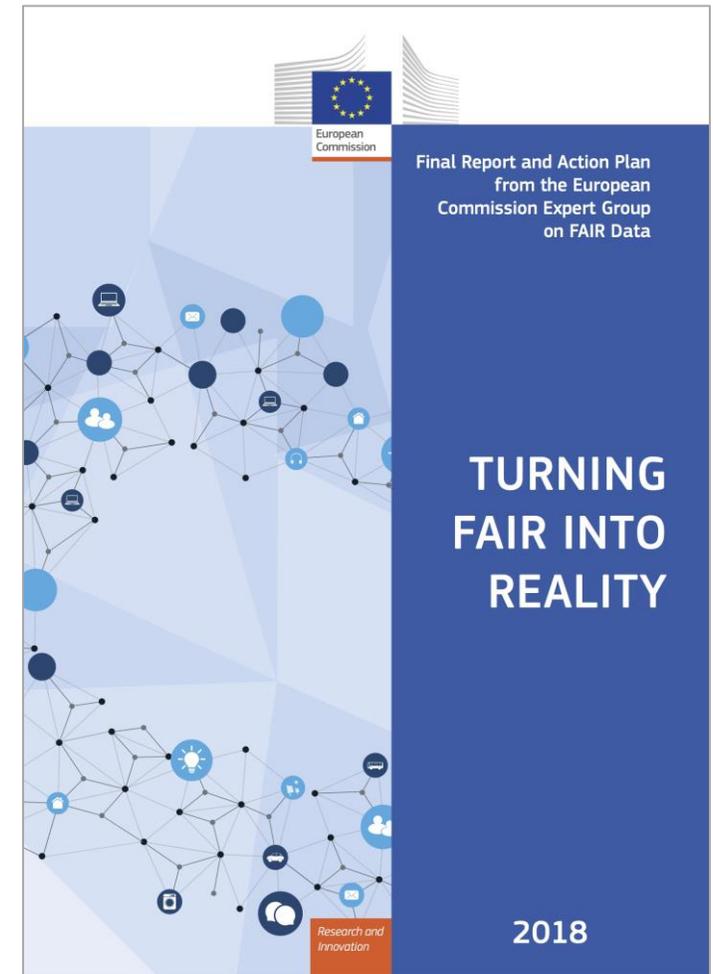
**Report and Action Plan:** Take a holistic approach to lay out what needs to be done to make FAIR a reality, **in general and for EOSC**

**Recommendations and Actions:** 27 clear recommendations, structured by these topics, are supported by precise actions for stakeholders.

**Ambitious and pragmatic:** thorough discussion and a comprehensive Action Plan. **If these things are done by institutions, nation states and in the context of EOSC we will move a lot closer to achieving the vision of Open Science and a network of FAIR data, digital objects and services.**



- To develop **recommendations** on what needs to be done to turn the FAIR data principles into reality (EC, member states, international level).
- Develop the **FAIR Data Action Plan**, by proposing a list of concrete actions as part of its Final Report.
- **Extensive consultation** on report framework and on interim report published early June 2018.
  - Consultations: <https://github.com/FAIR-Data-EG>
  - Interim Report: <https://doi.org/10.5281/zenodo.1285272>
- **Launch and disseminate FAIR Data Action Plan** and support Commission communication in November 2018
- <http://tinyurl.com/FAIR-EG>



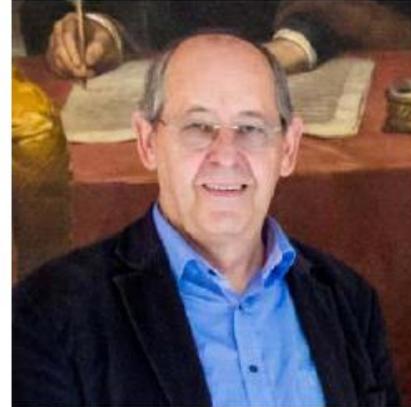
# Expert Group Members



Simon Hodson, CODATA  
Chair of FAIR Data EG



Rūta Petrauskaitė, Vytautas  
Magnus University



Peter Wittenburg, Max Planck  
Computing & Data Facility



Sarah Jones, Digital Curation  
Centre (DCC), Rapporteur



Daniel Mitchen, Data  
Science Institute,  
University of Virginia



Françoise Genova,  
Observatoire Astronomique  
de Strasbourg



Leif Laaksonen, CSC-  
IT Centre for Science



Natalie Harrower,  
Digital Repository of  
Ireland – year 2 only



Sandra Collins,  
National Library of  
Ireland – year 1 only

# Structure of the Report



## Addresses the following key areas:

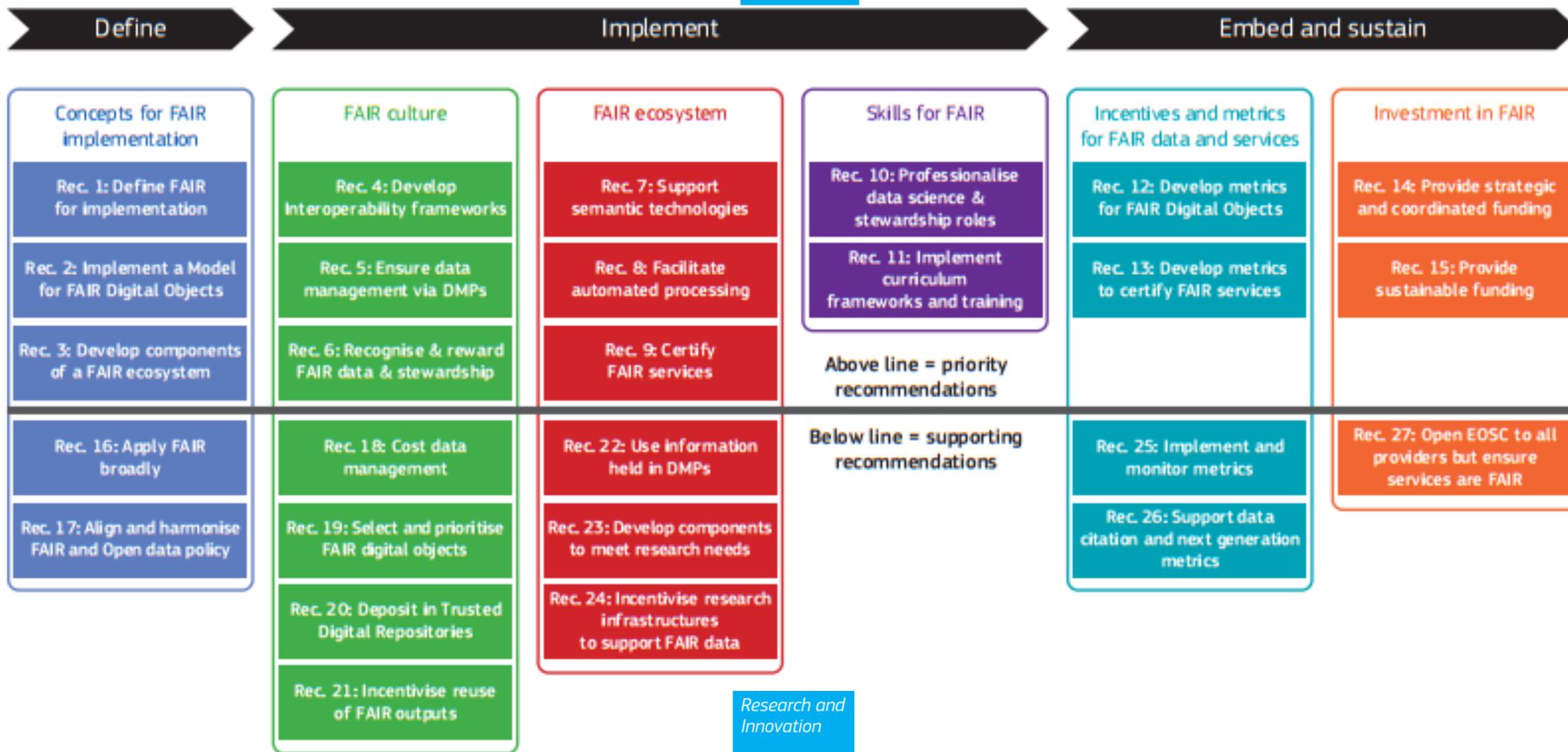
1. Concepts for FAIR
2. Creating a FAIR culture
3. Creating a technical ecosystem for FAIR
4. Skills and capacity building
5. Incentives and metrics
6. Investment and sustainability



# FAIR Action Plan



European  
Commission



Research and  
Innovation



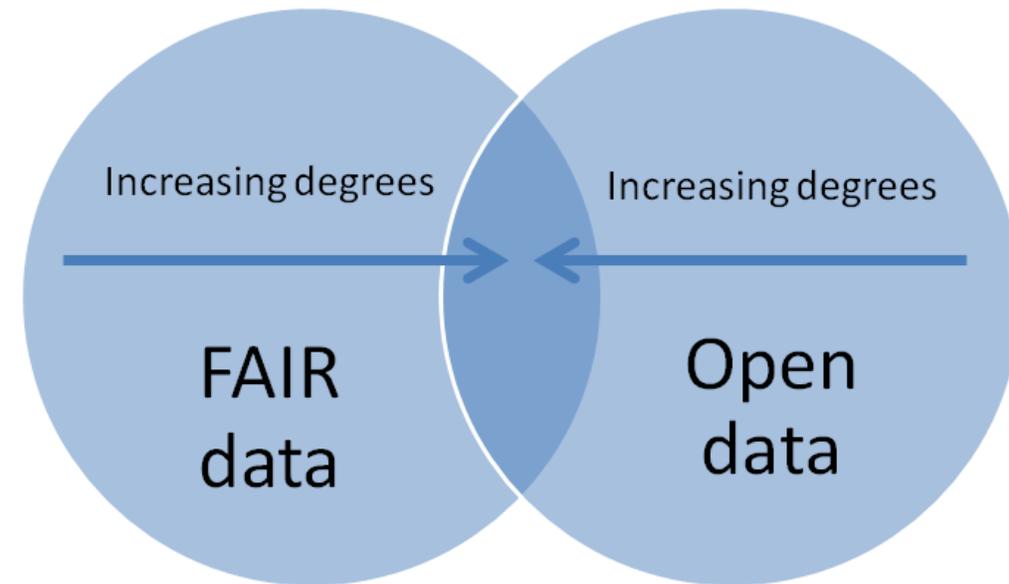
## Key Concepts and Enabling Actions

1. FAIR and Open and other supporting policies
2. Model for FAIR Digital Objects
3. FAIR ecosystem
4. Interoperability frameworks
5. Skills
6. Metrics and Incentives
7. Investment and Sustainability

# FAIR and Open



- Concepts of FAIR and Open should not be conflated. Data can be FAIR or Open, both or neither
- Even internal or restricted data will benefit from being FAIR, and there are legitimate reasons for restriction which vary by discipline
- **The greatest potential reuse and value comes when data are both FAIR and Open.**
- Align and harmonise FAIR and Open data policy: **as Open as possible, as closed as necessary**
- Open by default for publicly funded data.
- Allow disciplines to develop their agreements and find optimal balance, short embargo periods can be appropriate.



# Additional supporting concepts and policies



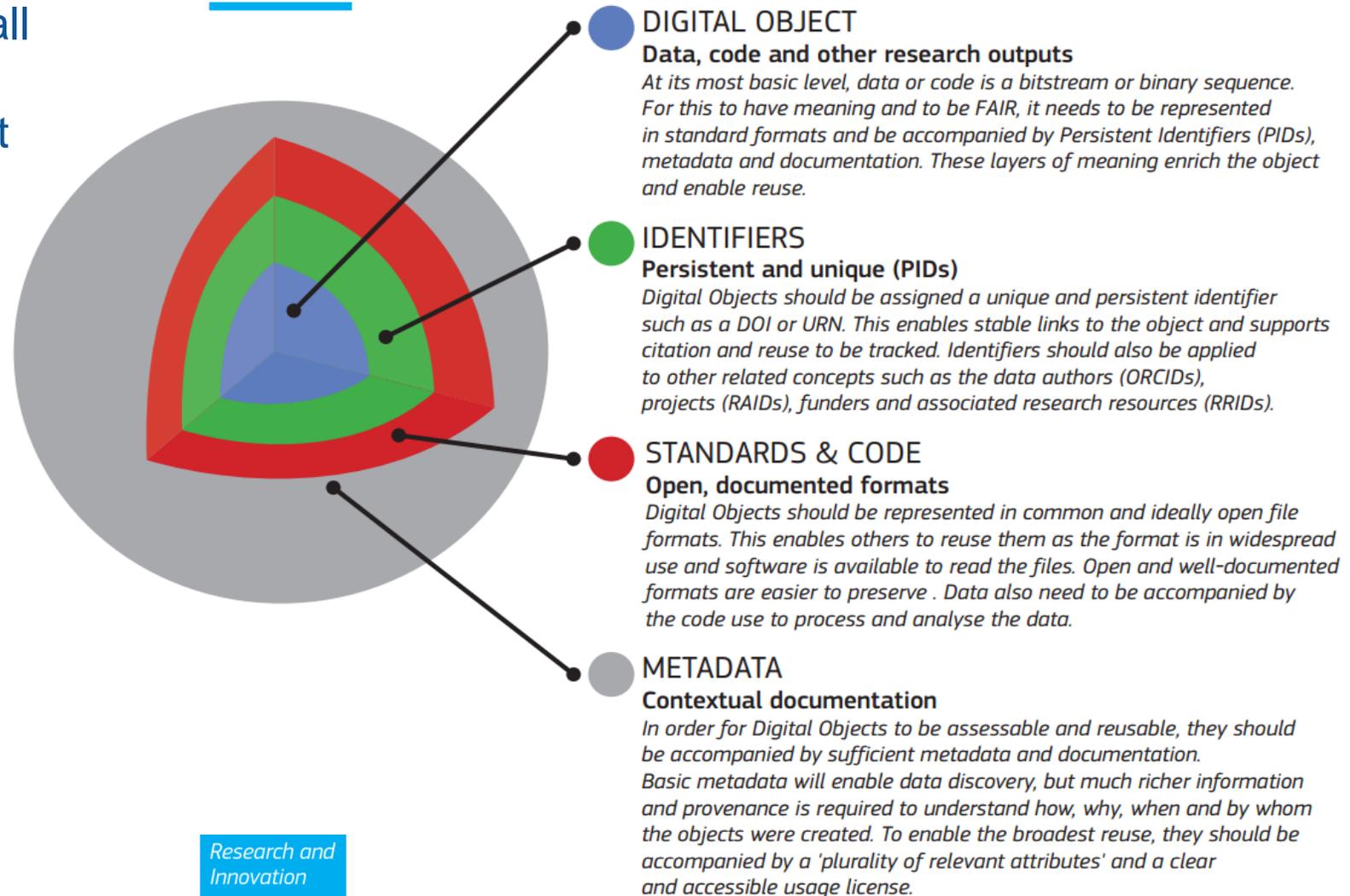
Making FAIR a reality depends on additional concepts that are implied by the principles, including:

- The timeliness of sharing
- Data appraisal and selection
- Long-term preservation and stewardship
- Assessability – to assess quality, accuracy, reliability
- Legal interoperability – licenses, automated

# FAIR Digital Objects



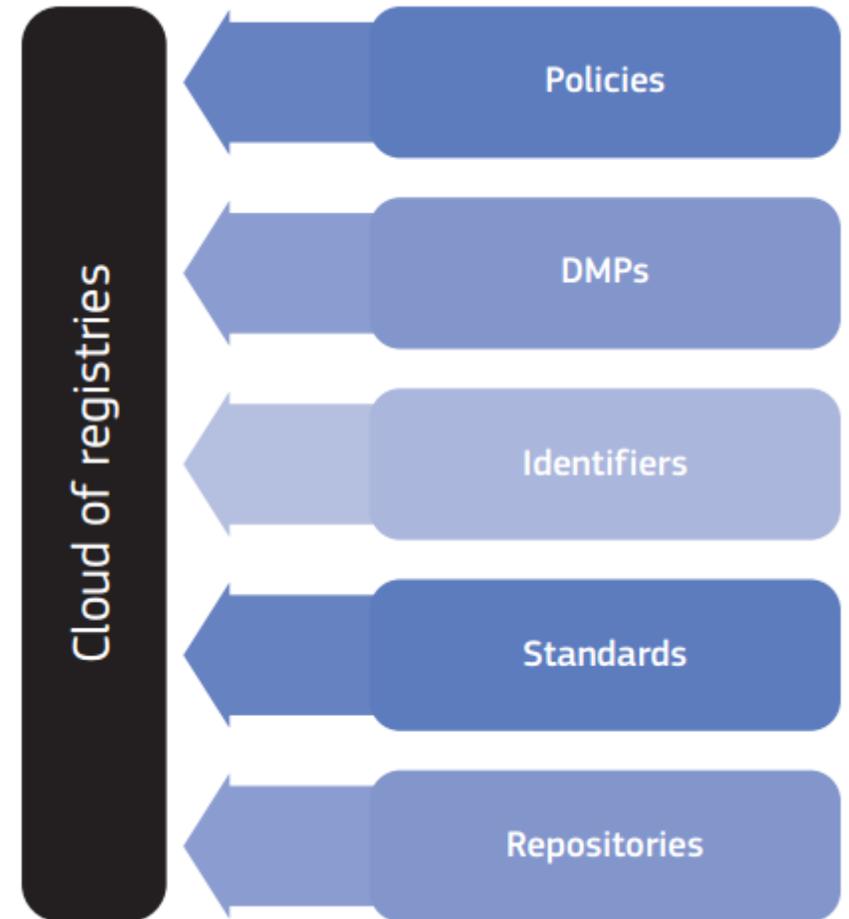
- FAIR should be applied broadly to all objects (including metadata, identifiers, software and DMPs) that are essential to the practice of research
- Universal use of PIDs
- Use of common formats
- Data accompanied by code
- Rich metadata
- Clear licensing



# The FAIR Ecosystem



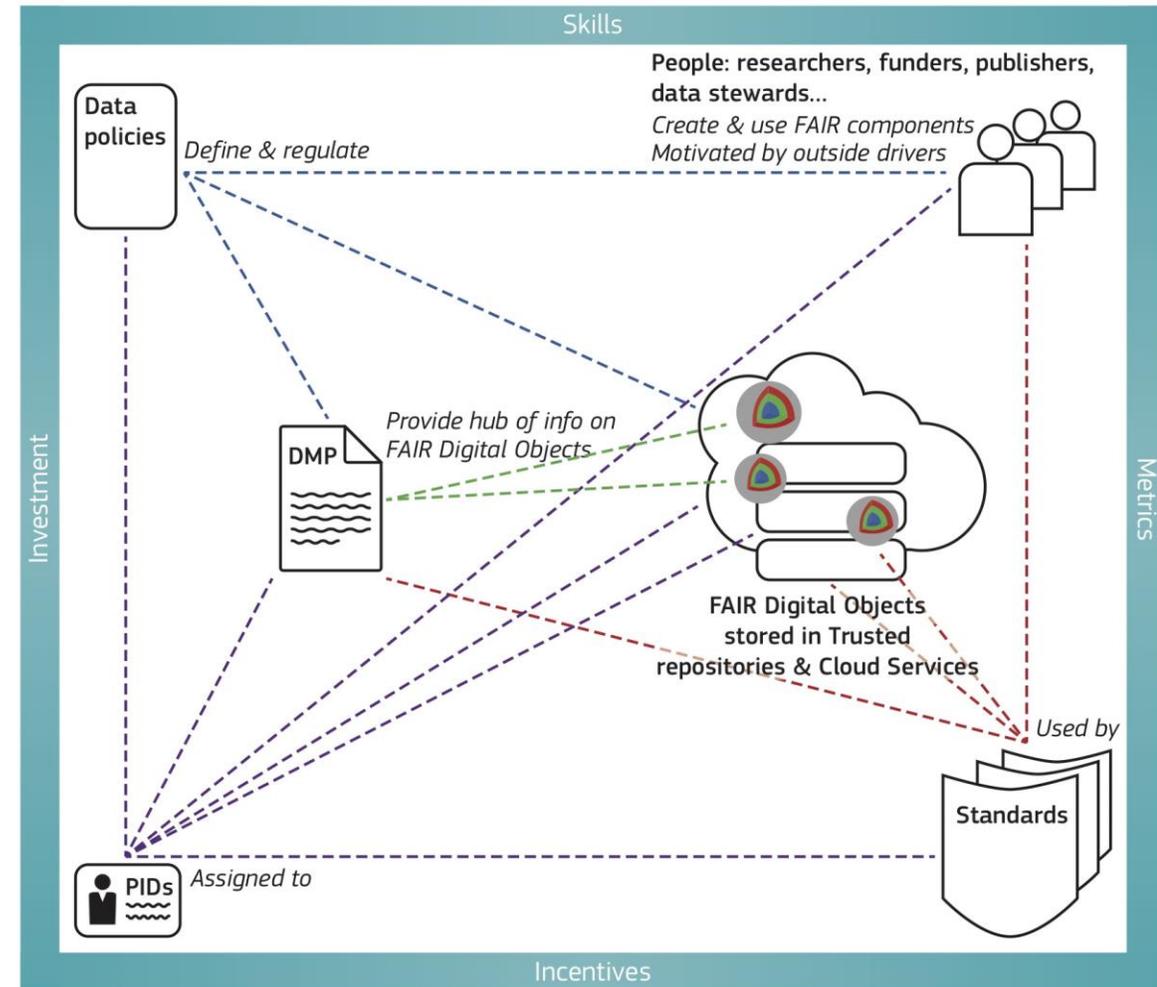
- The realisation of FAIR relies on an ecosystem of components
- Essential are:
  - Policies
  - Data Management Plans
  - Identifiers
  - Standards
  - Repositories
- Registries to catalogue each component of the ecosystem, and automated workflows between them.
- Begin by enhancing existing registries; build those for DMPs and IDs



# The FAIR Ecosystem



- **Ecosystem and its components should work for humans and machines**
- Need to clearly define infrastructure components essential in specific contexts and fields
- Semantic technologies are essential for interoperability and need to be developed, expanded and applied.
- Automated processing should be supported and facilitated by FAIR components.





## A core element of research projects

- DMPs should cover all research outputs
- DMPs should be living documents
- DMPs should be tailored to disciplinary needs
- DMPs should be machine-actionable – use information in them!
- Harmonisation of DMP requirements across funders and organisations



*DMP acting as a hub of information on FAIR digital objects, connecting to the wider elements of the ecosystem*



- Data services must be encouraged and supported to obtain certification, as frameworks to assess FAIR services emerge.
- Existing community-endorsed methods to assess data services, in particular CoreTrustSeal (CTS) for trusted digital repositories, **should be used as a starting point.**
- Many aspects of FAIR apply to services (findability, accessibility, use of standards...) but also important to assess:
  - Appropriate policy is in place
  - Robustness of business processes
  - Expertise of current staff
  - Value proposition / business model
  - Succession plans
  - Trustworthiness



# Interoperability Frameworks



- **FAIR and Open Science imply significant change in some research communities.**
- Some research communities have already made great progress to develop interoperability frameworks.
- Urgent need to develop interoperability frameworks for FAIR sharing within disciplines and for interdisciplinary research
- Research communities need to be supported and encouraged to to develop interoperability frameworks that define their practices for data sharing, data formats, metadata standards, tools and infrastructure.
- **To support interdisciplinary research**, these interoperability frameworks should be articulated in common ways and adopt global standards where relevant.

Semantic technologies

Metadata specifications

Data formats

Shared infrastructures

Community agreements

# International Virtual Observatory Alliance Case Study



Astronomy has been a pioneer of open data sharing, and remains at the forefront. Jointly using data from different instruments or gathered at different times is at the core of the discipline's science process.

The discipline established the International Virtual Observatory Alliance (IVOA <http://www.ivoa.net>) in 2002 to develop its interoperability framework at the international level. It is fully operational and continuously updated to deal with evolving requirements.

It progressively developed the standards necessary to Find, Access and Interoperate data, which have been taken up by archives of space and ground-based telescopes and major disciplinary data centers.

The first step was the definition of a standard for observational data called Flexible Image Transport System (FITS) in 1979. This includes data and metadata, allowing data Reuse.

The VO is an interoperability layer to be implemented by data providers on top of their data holdings. It is a global, open and inclusive framework: anyone can "publish" a data resource in the VO, and anyone can develop and share a VO-enabled tool to access and process data found in the VO.

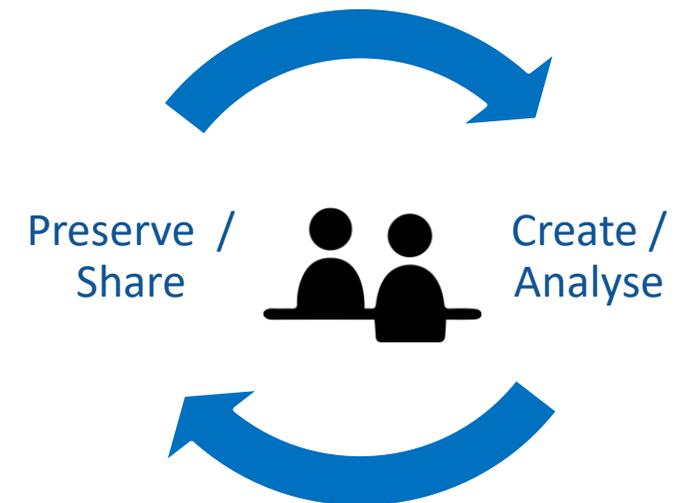
# Key Drivers Needed for Change



Research and Innovation



- Two skills sets needed to support FAIR data:
  - **data science skills** (the ability to handle, process and analyse data)
  - **data stewardships** (skills to ensure data are properly managed, shared and preserved throughout the research lifecycle and in subsequent storage)
- Cohorts with combinations of these skills urgently need to be developed
- Coordinate, systematise and accelerate the pedagogy
- **Need for a major programme of train-the-trainer activity**
- Support formal and informal learning, CPD
- Ensure researchers have foundational data skills





- A set of metrics for FAIR Digital Objects should be developed and implemented, starting from the basic common core of descriptive metadata, PIDs and access.
- Build on existing work in this space – new RDA Working Group
- Certification schemes are needed to assess all components of the ecosystem as FAIR services.
- FAIR services require additional metrics that are not just based on F.A.I.R.
- Existing certification schemes (CTS and OAIS and related should be used as a starting point).



# From metrics to incentives



- Use metrics to measure practice but beware misuse and unintended consequences.
- Generate genuine incentives to promote FAIR practices – career progression for data sharing and curation, recognise all outputs of research, include in recruitment and project evaluation processes.
- Recognise diversity of research roles, give credit to data management and sharing.
- Evidence of past FAIR practice should be included in assessments of research contribution.
- Implement ‘next-generation’ metrics.
- Automate reporting as far as possible.



- Considerable evidence of the ROI for Open Science and FAIR.
- Provide strategic and coordinated funding to maintain the components of the FAIR ecosystem.
- Component / service providers need to demonstrate value proposition and robust business model.
- Ensure funding is sustainable, understand the dynamics of the income streams and funding models. No unfunded mandates.
- Open EOSC to all providers, but ensure services are FAIR

OECD publishing

## BUSINESS MODELS FOR SUSTAINABLE RESEARCH DATA REPOSITORIES

OECD SCIENCE, TECHNOLOGY  
AND INNOVATION  
POLICY PAPERS  
December 2017 No. 47

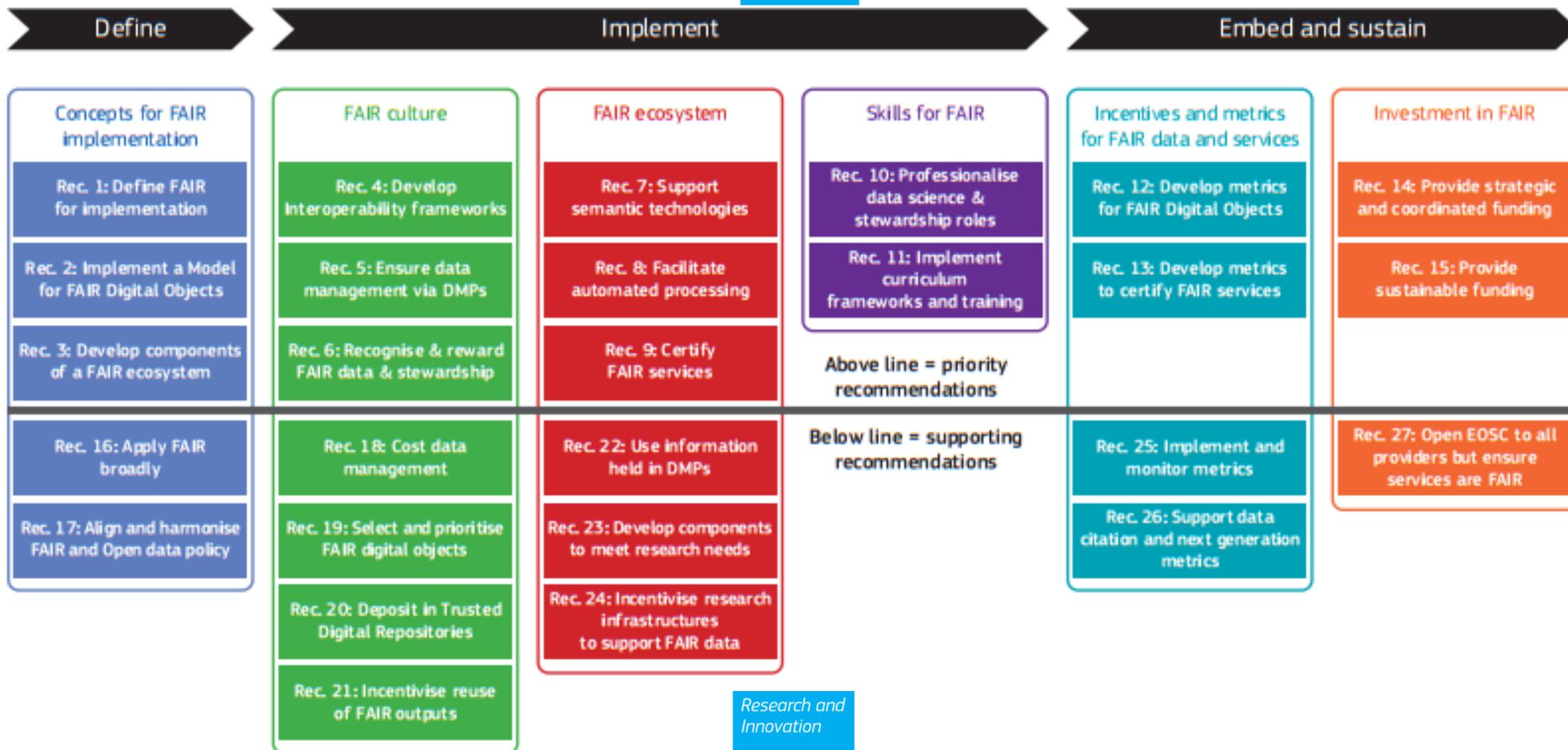


**Business models and sustainability:**  
<https://doi.org/10.1787/302b12bb-en>

# FAIR Action Plan



European  
Commission



Research and  
Innovation



## Rec. 1: Define FAIR for implementation

To make FAIR data a reality it is necessary to incorporate and emphasise concepts that are implicit in the FAIR principles, namely: data selection, long-term stewardship, assessability, legal interoperability and the timeliness of sharing.

Action 1.1: Additional concepts and policies should be refined that make explicit that data selection, long-term stewardship, assessability, legal interoperability and timeliness of sharing are necessary for the implementation of FAIR.

**Stakeholders:** Coordination fora; Research communities; Data service providers.

Action 1.2: The term FAIR is widely-used and effective so should not be extended with additional letters.

**Stakeholders:** Research communities; Data service providers.

Action 1.3: The relationship between FAIR and Open should be clarified and well-articulated as the concepts are often wrongly conflated. FAIR does not mean Open. However, in the context of the EOSC and global drive towards Open Science, making FAIR data a reality should be supported by policies requiring appropriate Openness and protection, which can be expressed as 'as Open as possible, as closed as necessary'.

**Stakeholders:** Policymakers; Research communities.

**Related recommendations:** Rec. 2: Implement a model for FAIR Digital Objects; Rec. 4: Develop interoperability frameworks for FAIR sharing; Rec. 17: Align and harmonise FAIR and Open data policy.

- The Expert Group has developed an **overarching** FAIR Action Plan
- Intended that this will be a document that can be applied in EOSC, by member states, by research communities and disciplines.
- Hope is that this will inspire the definition of more detailed FAIR Action Plans at research community and Member State level
- What are the priority actions in your area and for which stakeholders?

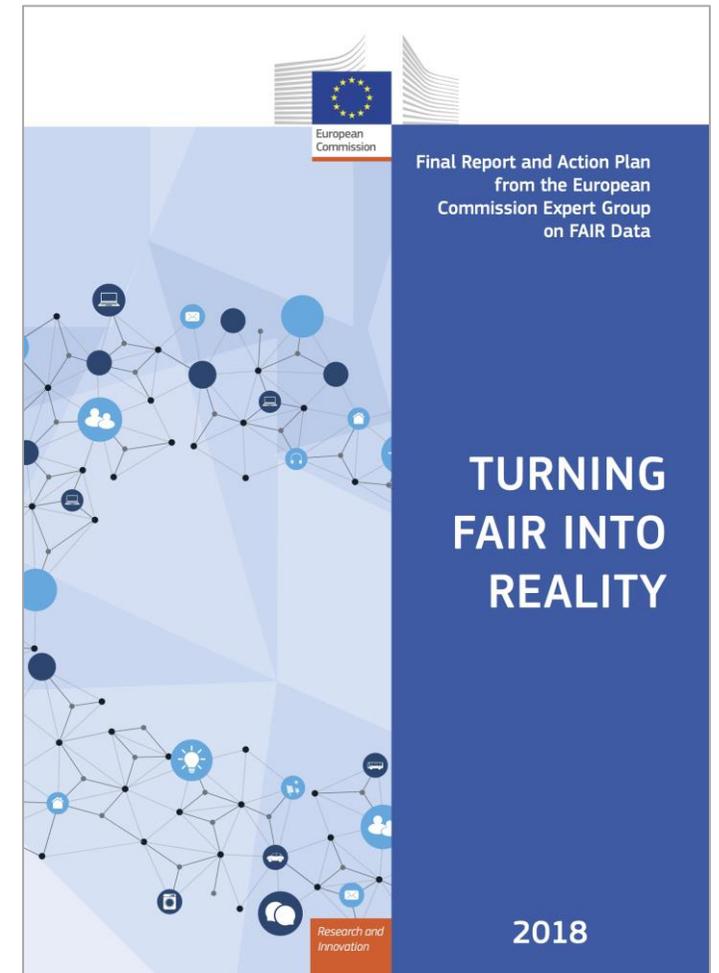


# Thank you for your attention!



## Questions?

- Simon Hodson, Chair, [simon@codata.org](mailto:simon@codata.org), @simonhodson99
- Sarah Jones, Rapporteur, [sarah.jones@glasgow.ac.uk](mailto:sarah.jones@glasgow.ac.uk), @sjDCC
- <http://tinyurl.com/FAIR-EG>



Simon Hodson, CODATA  
Chair of FAIR Data EG



Rūta Petrauskaitė, Vytautas  
Magnus University



Peter Wittenburg, Max Planck  
Computing & Data Facility



Sarah Jones, Digital Curation  
Centre (DCC), Rapporteur



Daniel Mietchen, Data  
Science Institute,  
University of Virginia



Françoise Genova,  
Observatoire Astronomique  
de Strasbourg



Leif Laaksonen, CSC-  
IT Centre for Science



Natalie Harrower,  
Digital Repository of  
Ireland – year 2 only



Sandra Collins,  
National Library of  
Ireland – year 1 only

Research and  
Innovation

Turning FAIR into Reality: Report and Action Plan

<https://doi.org/10.2777/1524>