Data Archiving and Networked Services

DANS

Emergence of minimal European standards for Open Science from a heterogeneous landscape

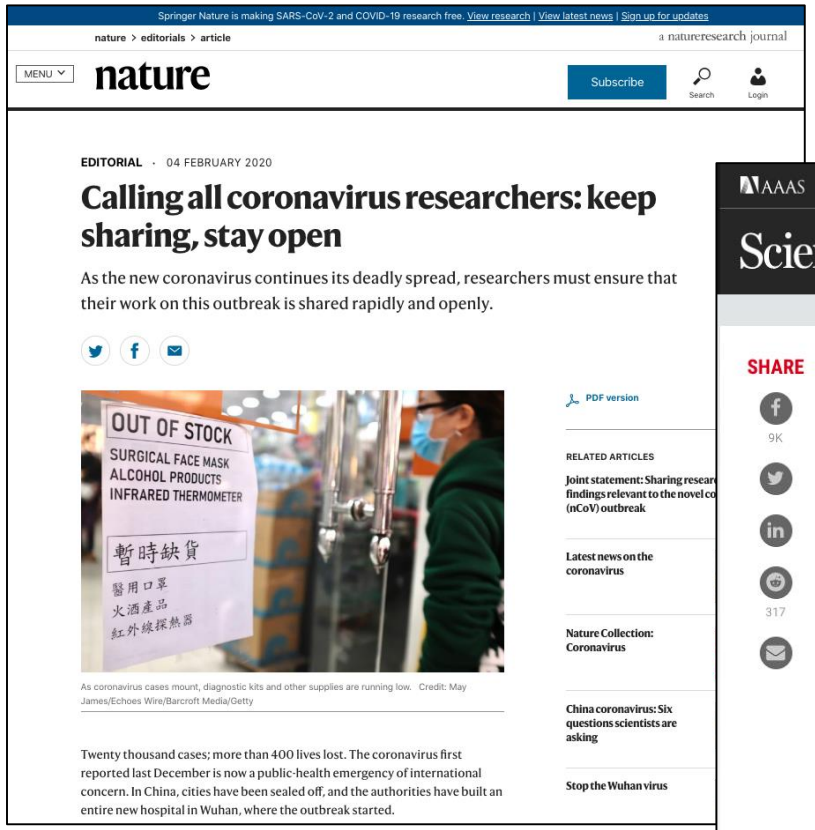
Data as a first class Open Science citizen

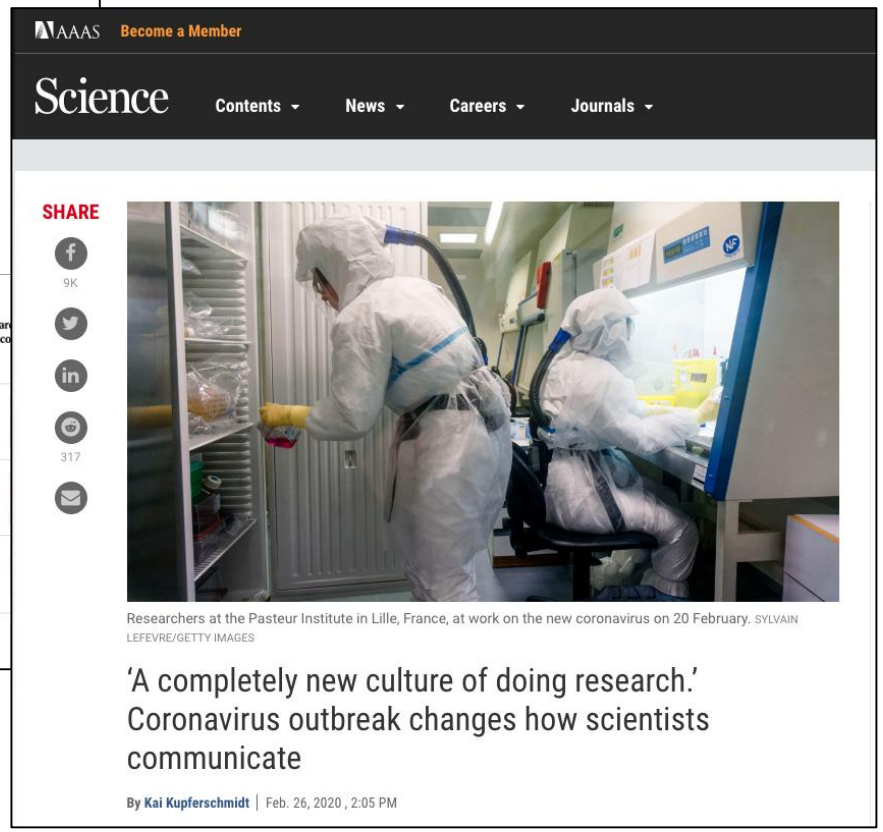
Mustapha Mokrane

@MokraneMA

6 March 2020, CHIST-ERA Workshop on Open Science in Transnational Research, Bern

Open Data for Open Science





https://doi.org/10.1038/nature.2015.17694

Open Data for Open Science

Growing recognition of the value of research data

→ Strong trend of Open Science supported by data policies

Replication (crisis)

→ Data sharing and preservation are becoming part of scientific integrity

→ E.g. Reproducibility project: Cancer Biology research

Reuse of data is recognized as a scientific and economical imperative

→ Research funders are increasingly mandating science to be “open”

→ E.g.: NIH, NSF, Wellcome, European Commission and many others…

Good Open Data citizenship!

Three Examples of transnational initiatives:

1. Belmont Forum e-Infrastructures & Data Management
2. Science Europe
3. European Open Science Cloud

Two examples of emerging minimum ‘standards’ (good practice)

1. FAIR Data Principles
2. CoreTrustSeal Data Repositories certification

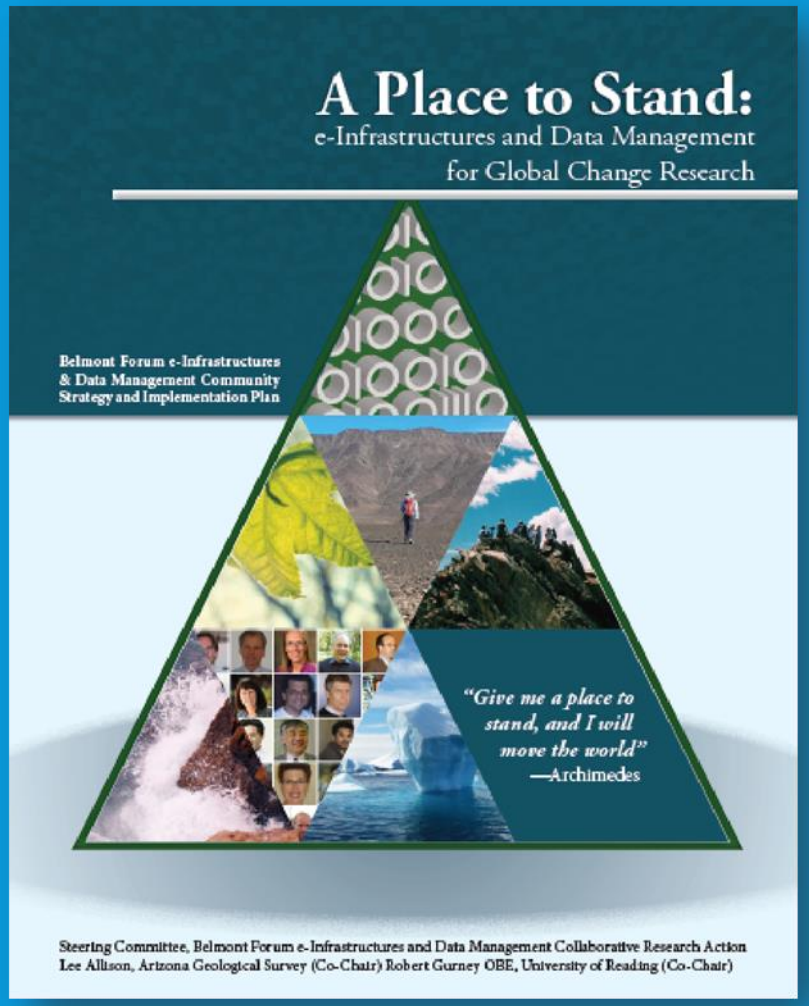
BELMONT FORUM

INTERNATIONAL PARTNERSHIP OF FUNDING AGENCIES

e-INFRASTRUCTURE DATA MANAGEMENT PROJECT

- ***More than 25 national science-funding agencies,*** *science councils* + some NGOs
- ***Increase impact of environmental research and data:*** through transnational, transdisciplinary Collaborative Research Actions
- ***Promote and leverage policies, methods and systems for: transnational data openness and FAIRness, sharing research products and data stewardship***

BELMONT FORUM DATA POLICY AND PRINCIPLES



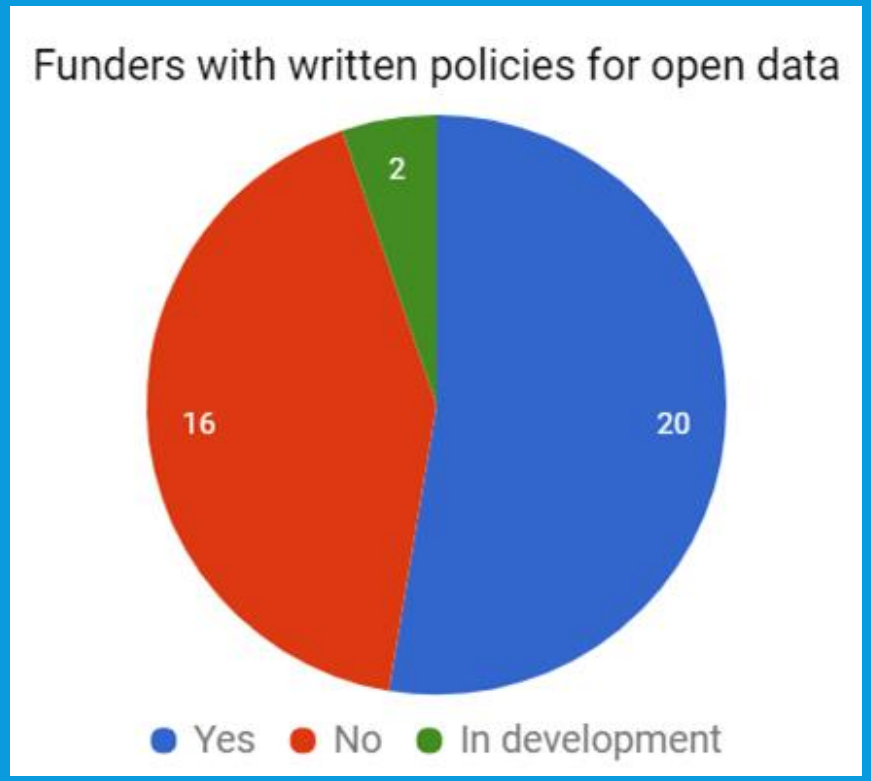
Belmont Forum adopted Open Data Policy and Principles October 2015

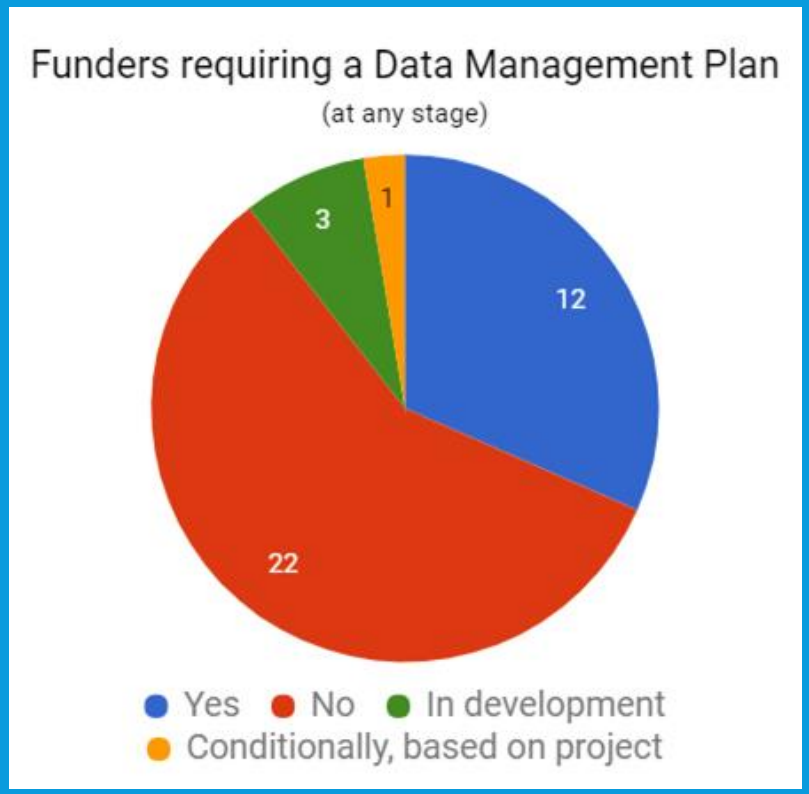
Data should be:

- **Discoverable** through catalogues and search engines
- **Accessible** as open data by default, and made available with minimum time delay
- **Understandable** in a way that allows researchers—including those outside the discipline of origin—to use them
- **Manageable and protected** from loss for future use in sustainable, **trustworthy repositories**

The Belmont Forum and its members will support and promote this data policy and principles with the intent of making these data principles enforceable over time. **This will allow member agencies to foster reproducibility in the science proposed, measure impact, and reduce the costs of research by allowing data reuse.**

BELMONT FORUM MEMBERS DATA POLICIES





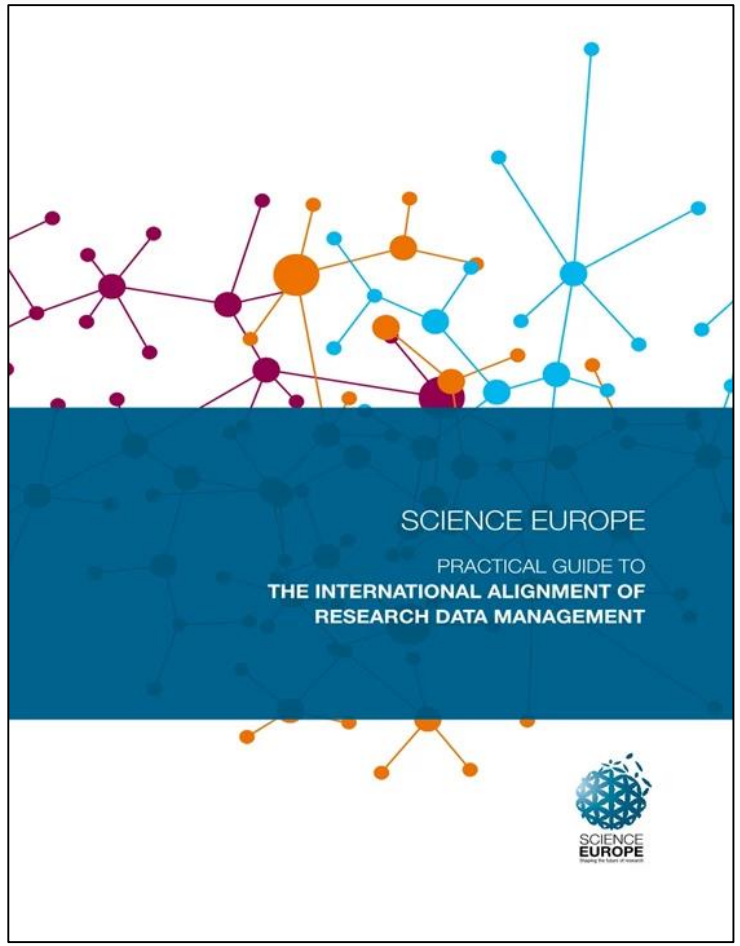
CHANGING DATA POLICY, DATA STEWARDSHIP AND RESEARCH PRACTICES

Action Theme 2: Data Planning

Promote effective data planning and stewardship in all Belmont Forum agency-funded research

The Belmont Forum should promote active and effective data stewardship in all Belmont Forum funded research, and enable harmonization of e-infrastructure through enhanced project data planning, monitoring, review and sharing. **Current Data Management Plan (DMP) requirements vary greatly across Belmont Forum agencies** in their level of detail and compliance with general stewardship principles, **including use of certified, trusted data repositories**…

Science Europe: International alignment of RDM



Align research data management (RDM) requirements across various research/funding organisations:

The core requirements for DMPs and criteria for the selection of trustworthy repositories presented in this guide are compliant with the **FAIR Data Principles**…

- Core requirements for data management plans
- Criteria for the selection of trustworthy repositories:
  - Some repositories have been **certified as trustworthy repositories**…**it is strongly recommended that repositories that have not yet been certified seek certification by such a body.**

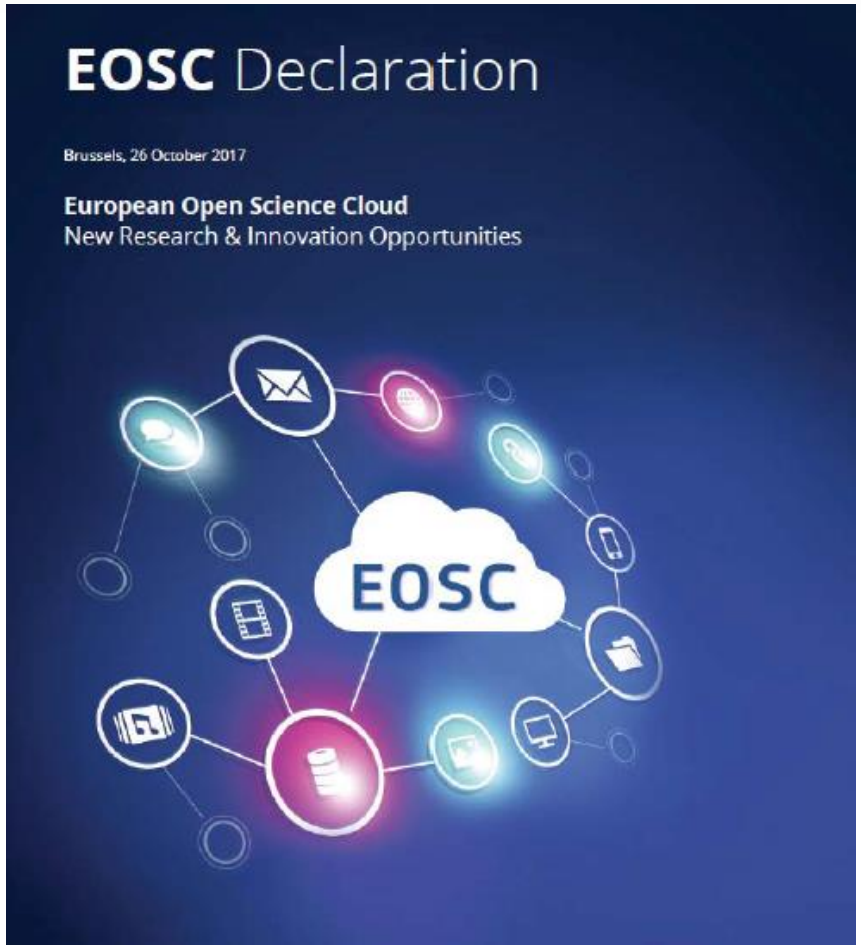
https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf

Science Europe: Implementing RDM Policies across EU

1. When **developing** RDM policies, collaborating within and beyond the own organisation is key, as involving all stakeholders concerned increases acceptance of the new policy.
2. When **communicating** on new requirements, research organisations should engage directly with all actors concerned and explain the underlying concepts and benefits. The actors essentially include scientific, legal, financial and communication staff.
3. When **implementing** RDM policies, both RFOs and RPOs need to work together if they want to provide sufficient support for researchers, such as training or online tools to set up and managing DMPs.

https://www.scienceeurope.org/media/jikjlb2g/se_rdm_best_practices.pdf

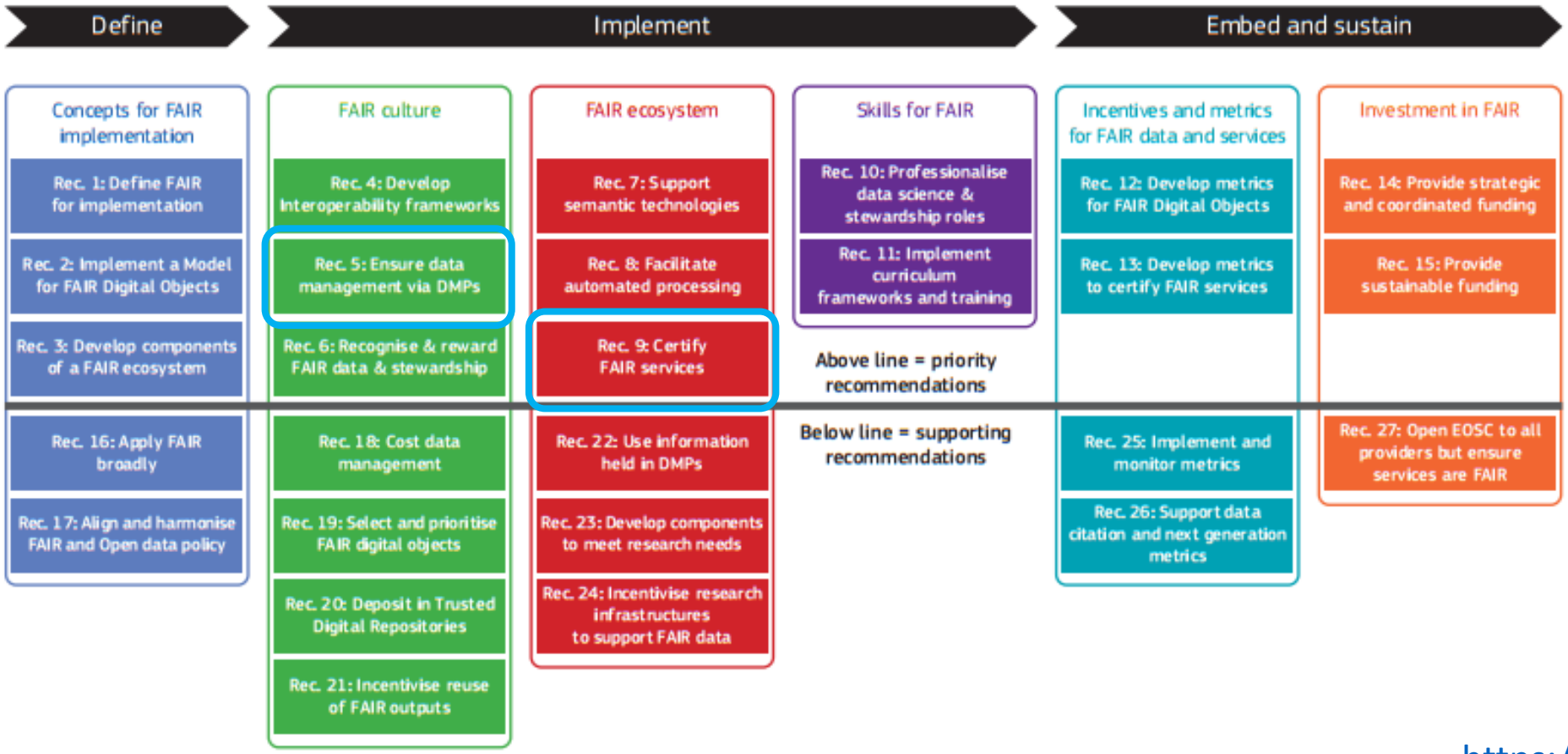
European Open Science Cloud (EOSC)



- [Data culture]
- [Open access by-default]
- [Skills]
- [Data stewardship]
- [Rewards and incentives]
- [FAIR (data) principles]
- [Implementation & transition to FAIR]
- [FAIR Data governance]
- [Standards]
- [Research data repositories]
- [Accreditation/certification]
- [Data Management Plans]
- …

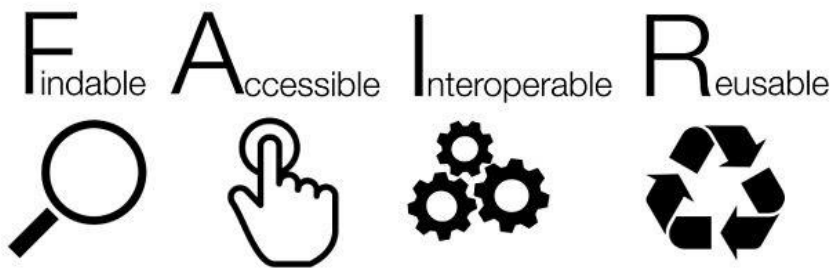
https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

European Open Science Cloud (EOSC)



https://doi.org/10.2777/54599

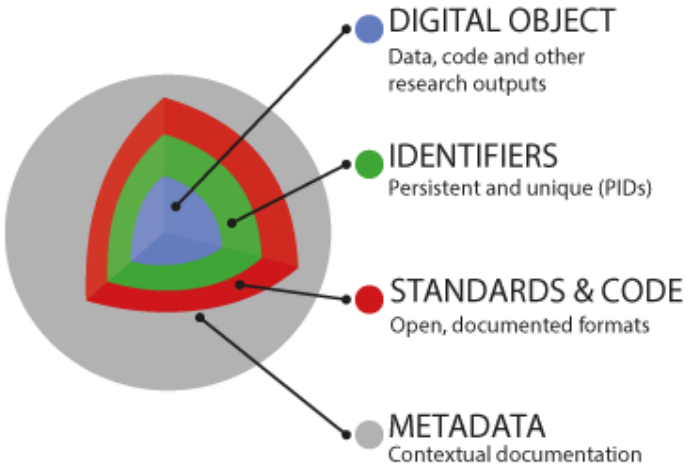
FAIR data guiding principles



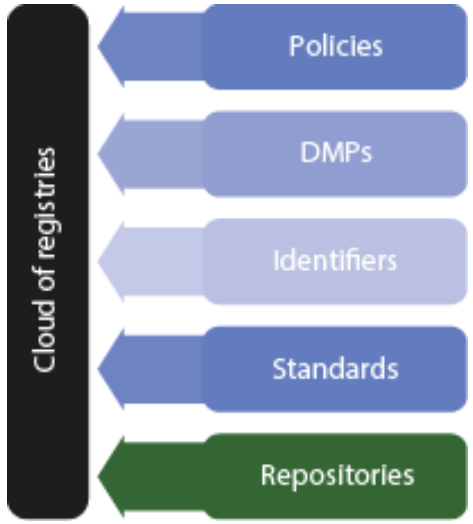
https://doi.org/10.1038/sdata.2016.18

Focus: Enable discovery and reuse of data

Process: Data management & stewardship



A model for FAIR Digital Objects



The components of a FAIR Ecosystem

CoreTrustSeal: Trustworthy Data Repositories certification

Universal framework for the **certification** of **data repositories** at the **core level** to support **long-term access** to **reusable data**.

- Gives **data producers assurance** that data are preserved and remain reusable in the future;
- Provides **funding bodies confidence** that investments are maximized;
- Enables **data consumers to choose** the repositories where data are held;
- Supports **repositories improve processes** for efficient data archiving and distribution.

CoreTrustSeal: Trustworthy Data Repositories certification

The CoreTrustSeal Trustworthy Data Repositories Requirements reflect the characteristics of trustworthy repositories -> **de facto minimal standard**

16 Requirements:

- Background information (R0)
- Organizational infrastructure (R1-6)
- Digital object management (R7-14)
- Technology and security (R15-16)

2019-07 CoreTrustSeal Draft Requirements 2020–2022 v00.01

Core Trustworthy Data Repositories Requirements

Background & General Guidance

The *Core Trustworthy Data Repositories Requirements* describe the characteristics of trustworthy repositories. All Requirements are mandatory and evaluated as standalone items. Although some overlap is unavoidable, duplication of evidence sought for each Requirement has been kept to a minimum. The options in checklists (e.g., repository type and curation level) are not considered to be comprehensive and may be refined in the future. Applicants are encouraged to add 'other' options.

Each Requirement is accompanied by Guidance text describing the information and evidence that applicants must provide to enable an objective review.

The applicant must indicate a compliance level for each of the Requirements:

0 – Not applicable
1 – The repository has not considered this yet
2 – The repository has a theoretical concept
3 – The repository is in the implementation phase
4 – The guideline has been fully implemented in the repository

Compliance levels are an indicator of the applicant's self-assessed progress, but reviewers judge compliance against response statements and supporting evidence. If an applicant believes a Requirement is not applicable (0), then this must be justified in detail. Compliance Levels of 1 or 2 are not sufficient for a successful application. Certification may be granted if some Requirements are in the implementation phase (3).

Responses statements provided by applicants should include links to supporting evidence online. As the core certification process does not include a site visit by an auditor, such publically available evidence provides transparent assurance of good practice. URL links should be verified immediately before submitting applications.

All responses must be in English. Although attempts will be made to match reviewers to applicants in terms of language and discipline, this is not always possible. Full translations of evidence are not required, but if non-English evidence is provided, then an English summary must be included in the response statement.

No sensitive information disclosure is required to acquire the CoreTrustSeal, but provisions are made within the certification process for repositories that want to share evidence materials also containing confidential information.

The CoreTrustSeal is valid for three years from the date it is awarded. Though repository systems and capabilities evolve continuously according to technology and user needs, they might not undergo major changes in this timeframe. An organization with well-managed business processes and records should be able to reapply with minimal revisions after three years unless:

- The organization, its data collection, or Designated Community has changed significantly.
- The CoreTrustSeal Requirements have been updated in ways that impact the applicant.

The CoreTrustSeal Requirements are subject to review and revision every three years. This does not affect a successful applicant until they seek renewal.

Glossary of Terms

Please refer to the Core Trustworthy Data Repositories Requirements Glossary: https://goo.gl/rQK5RN.

1

https://www.doi.org/10.5281/zenodo.667809

Data Archiving and Networked Services
DANS

Thank you for your attention!

mustapha.Mokrane@dans.knaw.nl

Twitter: @MokraneMA

www.dans.knaw.nl/en and www.fairsfair.eu