OPEN SCIENCE AND DATA:

FROM POLICIES AND PRACTICES TO OPPORTUNITIES AND CHALLENGES

Carthage Smith, OECD Global Science Forum

OECD
BETTER POLICIES FOR BETTER LIVES

What do we mean by Open Science?

2015: Open science refers to efforts to make the outputs of publicly funded research more widely accessible in digital format to the scientific community, the business sector or society more generally

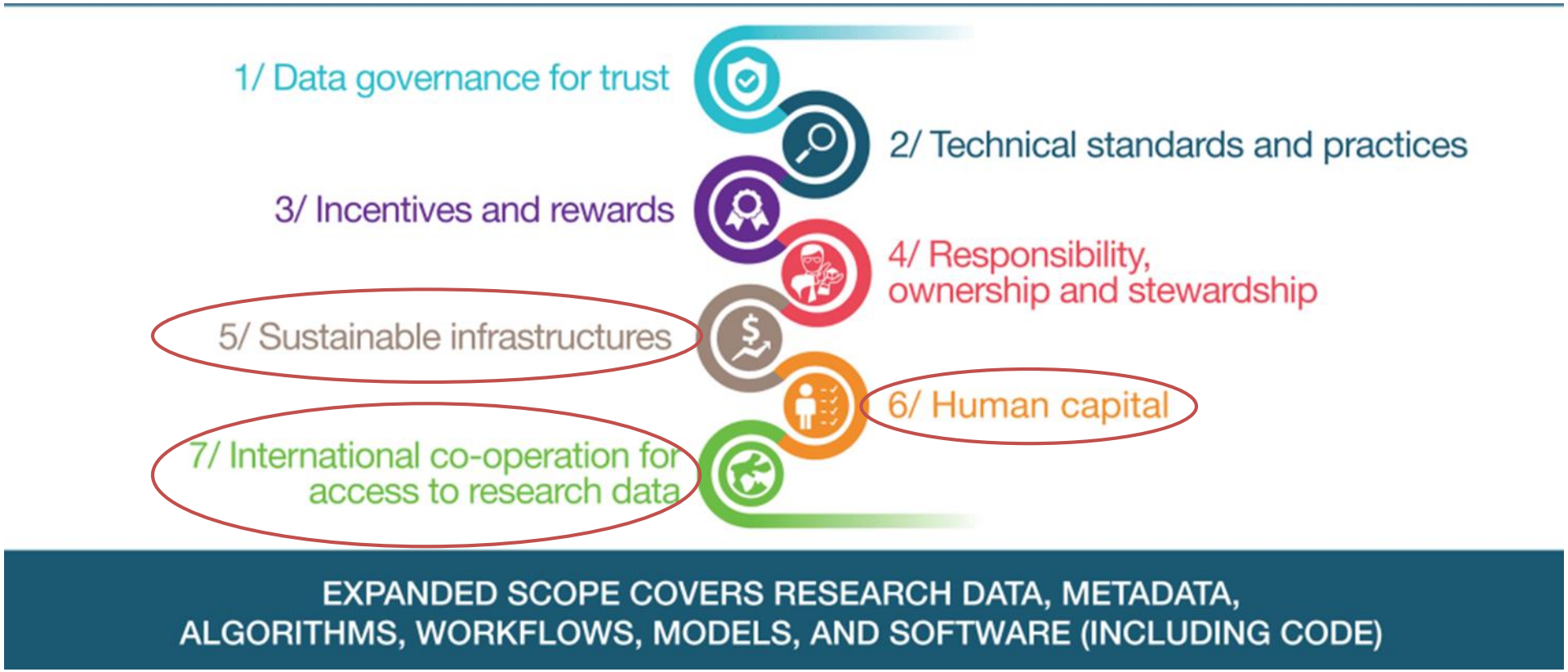
2018: Open Science in its broadest sense refers to efforts to make the scientific process more open and inclusive for all relevant actors

- Open access to publications
- FAIR data access
- Citizen engagement

Open Science and data (2016-)

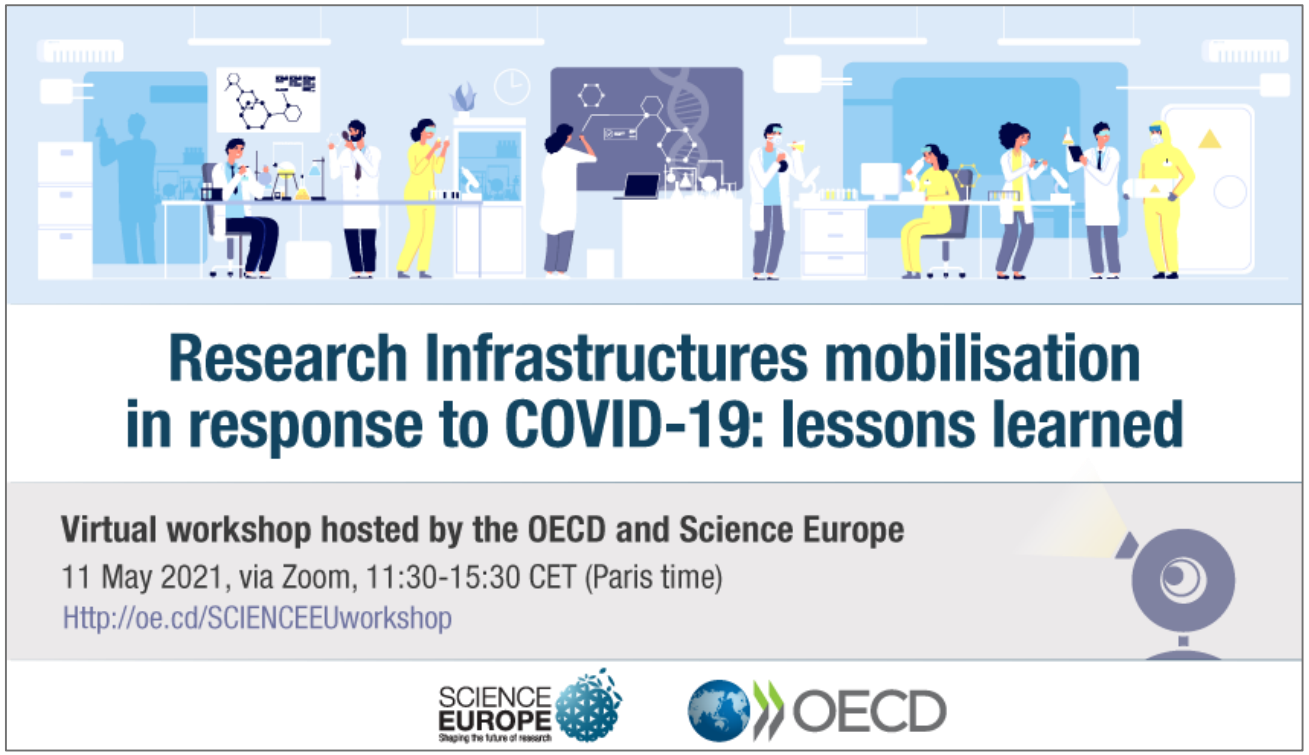
- Integrity and security in the global research ecosystem

OECD Recommendation on Access to Research Data from Public Funding (Jan 2021)



Full text available at https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347

1. Sustainable infrastructure



Objectives

1. Identify and describe existing revenue sources and business models
2. Test potential business models with various stakeholders, including funders
3. **Make policy recommendations to promote sustainable business models** for research data repositories

Elements of a successful business model

Funding Source	Pros	Cons
Structural funding	• Compatible with open data principles. • Longer-term stability. • Larger-scale and efficiencies. • Flexible as to allocation.	• Fixed, multi-year may not scale easily. • Competes with research funding. • Too many eggs in few baskets.
Host or institutional funding	• Compatible with open data principles. • Longer-term stability. • Efficiencies through sharing services. • Close to researchers (customers).	• Limited purview, with focus on local community. • May lead to fragmentation of domain data and lower interoperability. • Limited incentive to add value to data and develop related services.
Data deposit fees	• Compatible with open data principles. • Demand oriented and scales with demand (data ingest). • Researchers price sensitivity ensures cost constraint. • Open data is part of research and its funding.	• Cost disincentive to depositing, so depends on strong mandates. • May lead to low level of curation to contain costs (price). • May be difficult for repository to compete for deposits with comparable repositories that do not charge.
Data access charges (subscriptions or use fees)	• Users pay for what they want, so funding reflects value. • More market-oriented approach may provide incentive for cost constraint.	• Not compatible with open data principles and many funder mandates, limiting the potential market size. • Charges limit use and will reduce the value of data. • Revenue scales with use and not ingest or curation costs. • Vulnerable to funding cuts.
Diversification of revenue sources	• No single source of failure. • Can maintain compatibility with open data principles. • Flexible and enables experimentation with new services.	• May lead to higher transaction costs (managing multiple funding sources). • May lead to Mission drift.

Recommendations

1. All stakeholders should recognise that **research data repositories are an essential part of the infrastructure for open science**.
2. All research data repositories should have **a clearly articulated business model**.
3. Sponsors need to **consider the ways in which data repositories are funded**, and the pros and cons of various funding mechanisms in different circumstances.
4. Research data repository business models are **constrained by, and need to be aligned with, policy regulation** (mandates) **and incentives** (including funding).
5. In the context of financial sustainability, opportunities for **cost optimisation should be explored**.

2. International cooperation

Objectives

Establish principles and policy actions that can support open and sustainable international research data networks:

1. When is a data network needed?
2. How can governments use networks to maximize research data openness and reuse?
3. What is the best governance model for a particular network?
4. What interoperability arrangements are necessary for the effective operation of the network?
5. What business models can sustain a network over time?

Findings

- The most successful networks have engaged and **supportive users** who clearly understand and value the services of the network.
- The top issue faced by data networks in open sharing of data is **the varying attitudes and policies across countries.**
- Different research communities require **different data networks** because **the cultures** of data sharing vary.
- The most difficult aspects of **interoperability** are rooted in **human relationships and trust.**
- Developing a **coherent and sustainable business model** is a central challenge for virtually all data networks.

Recommendations

1. work toward **common definitions of, and agreements on, open data**. What is open data in different domains?
2. work toward commonly agreed and enforced **legal and ethical frameworks for the sharing of different types of public research data**.
3. Funders and host institutions should view internationally coordinated data networks as a **long-term strategic investment**
4. Networks should have clear business models, including value propositions and **measures of success** that are relevant to their different stakeholders and these measures **should be monitored**.

3. Digital Skills

Objectives

Address the following questions:

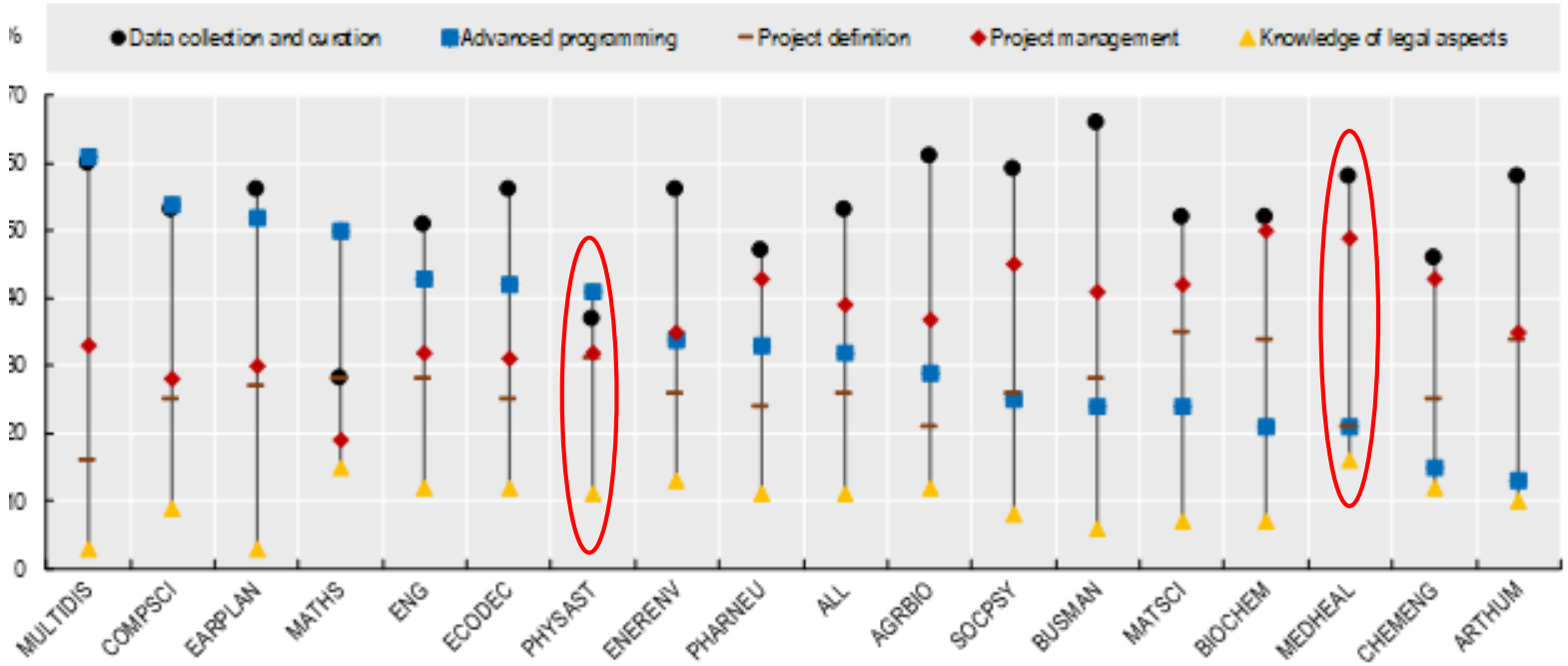
1. What is known about the digital workforce needs for data intensive science?
2. What is needed to build a digitally skilled research workforce?

➢ And make recommendations for policy action

Digital skills: Different needs in domains

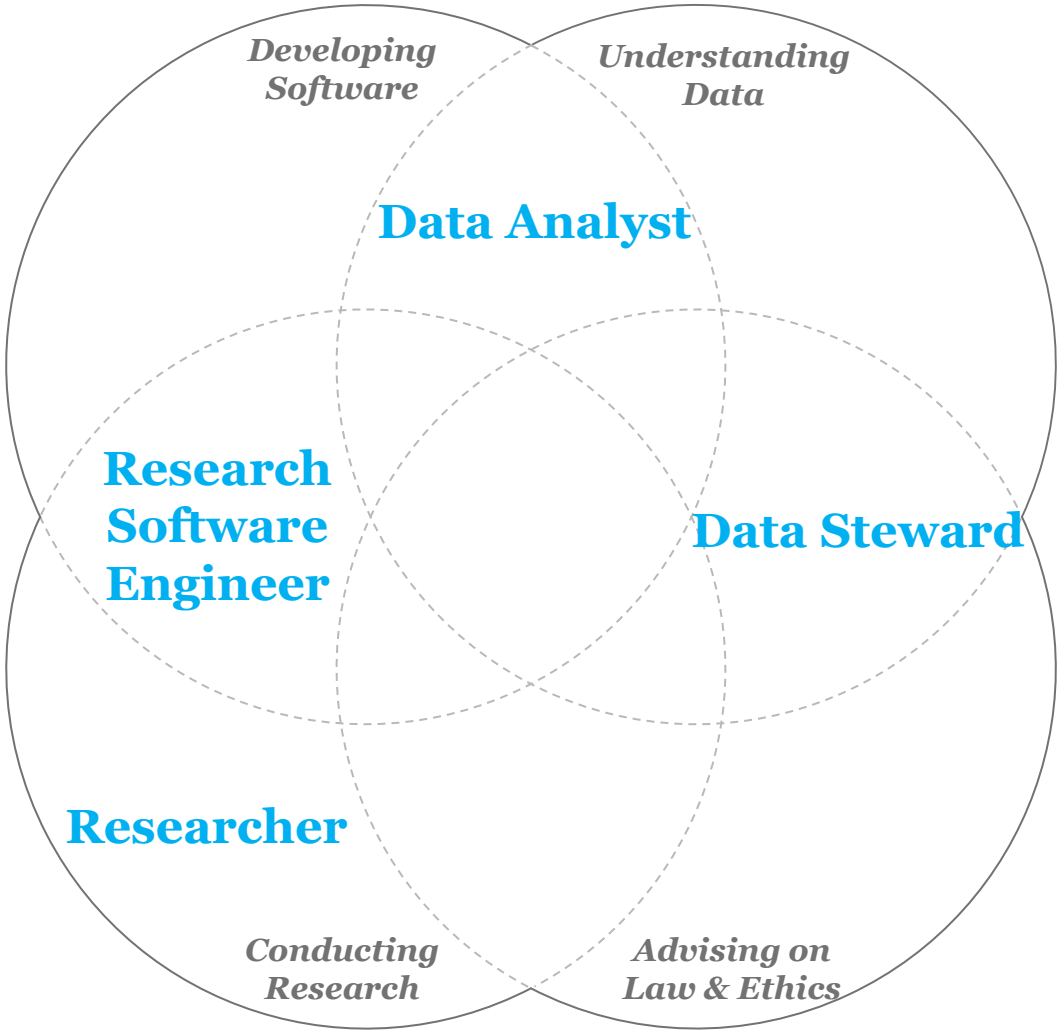
Figure 5.3. Most important skills for scientific authors' research work

Percentage of authors who deem each type of skill as important



Source: OECD, 2020. *Charting the Digital Transformation of Science*

Digital skills, frameworks & roles



Policy actions required in 5 key areas

Integrate digital workforce capacity development into broader science policy frameworks and actions, e.g. for open science and research integrity.

Identify the key competencies, skills and roles required for data-intensive science in different contexts.

Support training in foundational digital skills and more specialized skills for scientists and research support professionals.

Implement changes in academic evaluation and reward systems in order to attract and retain diverse digitally skilled staff.

Support development of communities for new professional roles, learners and trainers.

[Enhancing access to research data during crises: lessons learned from the COVID-19 pandemic - OECD]

Lessons learned

1. **Prior investment in** data infrastructure and networks is there to be leveraged during crises
2. There are technical obstacles in relation to achieving FAIR data but **the major challenge is data governance and this stretches beyond research**. There is need for a pandemic data response and management system that is publically monitored, controlled and accountable.
3. **Citizens and data subjects need to be included in co-designing data governance arrangements** for individual projects and for data systems as a whole. A global system of systems, based on national and regional networks is more likely to be effective than a single rigid global architecture.
4. Real time data access is difficult but is essential for managing a pandemic. From the outset, there is a need to **integrate health services and public health surveillance data, with research needs** and vice-versa.
5. There is a need to support incremental steps towards a global FAIR data system and **keep moving forward with 'a coalition of the willing'.** Top down visions, frameworks and direction need to integrated with bottom-up community-led efforts.

Lessons learned

6. There are significant gaps and biases in the global data on COVID-19, both in terms of geographic coverage and lack of representation of certain populations groups. **Inclusivity and inclusiveness should be an explicit aim for pandemic data response systems**
7. Not all data can, or should, be open but it should be FAIR and **all meta-data should be open and FAIR**. Open data (and software and workflows) enable many - public and private sector - actors to contribute in the response to crises.
8. There is a balance to be achieved between **standardisation and conformity versus diversity and innovation** but making data FAIR should be part of routine research practice and the adoption of community recognised standards is critical.
9. **The incentive system for researchers needs to change** and how we evaluate the contribution of researchers in response to COVID-19 will be a good test of this.
10. International fora, such as the Research Data Alliance, which bring together **different actors from different disciplines and different countries play an essential role in building the technical and social bridges** that are necessary to ensure the optimal use of data for research and policy-making during crises.

Thank You

OECD iLibrary (oecd-ilibrary.org)

OECD Global Science Forum – OECD
https://www.oecd.org/science/inno/global-science-forum.htm