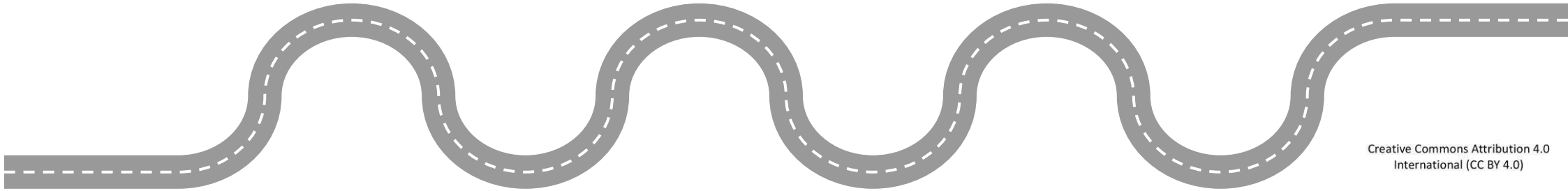
SciLifeLab

The road beyond Open

Sara El-Gebali

0000-0003-1378-5495

@yalahowy



09-May-2022

Open Science is…

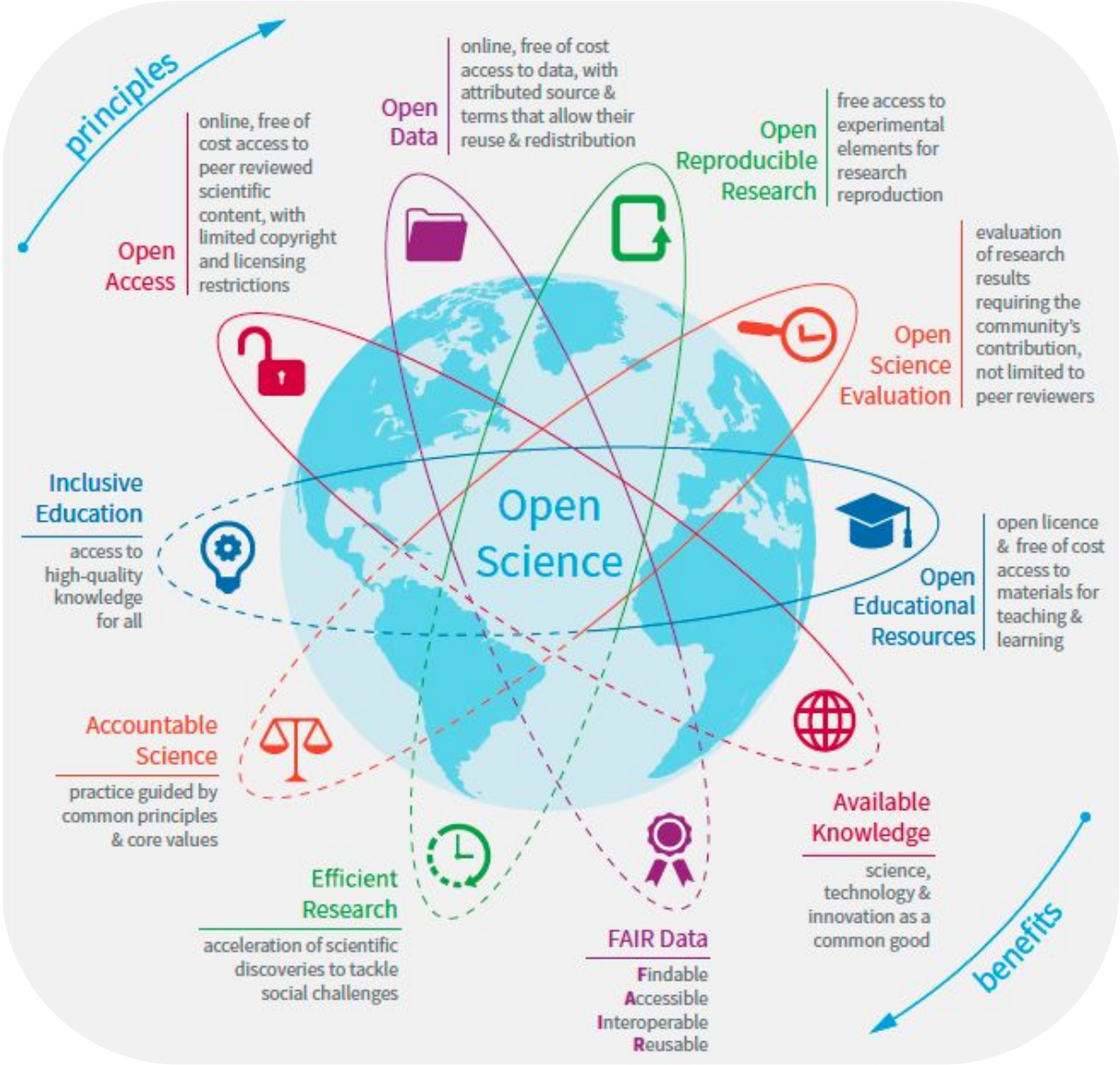
... an essential path to global scientific development that enhances research quality and efficiency and improves public trust in research results.

... a requisite for an inclusive society that makes science available to all, fosters the integration of scientific knowledge across disciplines, and assumes responsibility for the social impact that results from scientific advancement.

https://www.scientificinfographics.com/21-towards-an-open-science

Open science

more than scholarly publications



[1] Towards a global consensus on open science: report on UNESCO's global online consultation on open science. UNESCO, 2020
[2] FOSTER portal https://www.fosteropenscience.eu/, accessed in March 2021
[3] Science ouverte à l'Université de Genève : feuille de route pour un partage de connaissances scientifiques 2020-2023

Open Data

Open data is a natural expansion on open science beyond scholarly publications.

Open Data

Open data is a natural expansion on open science beyond scholarly publications.

“**Open data** is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike.”

The Open Data Handbook- Open Knowledge Foundation

Open Data integral to Open science

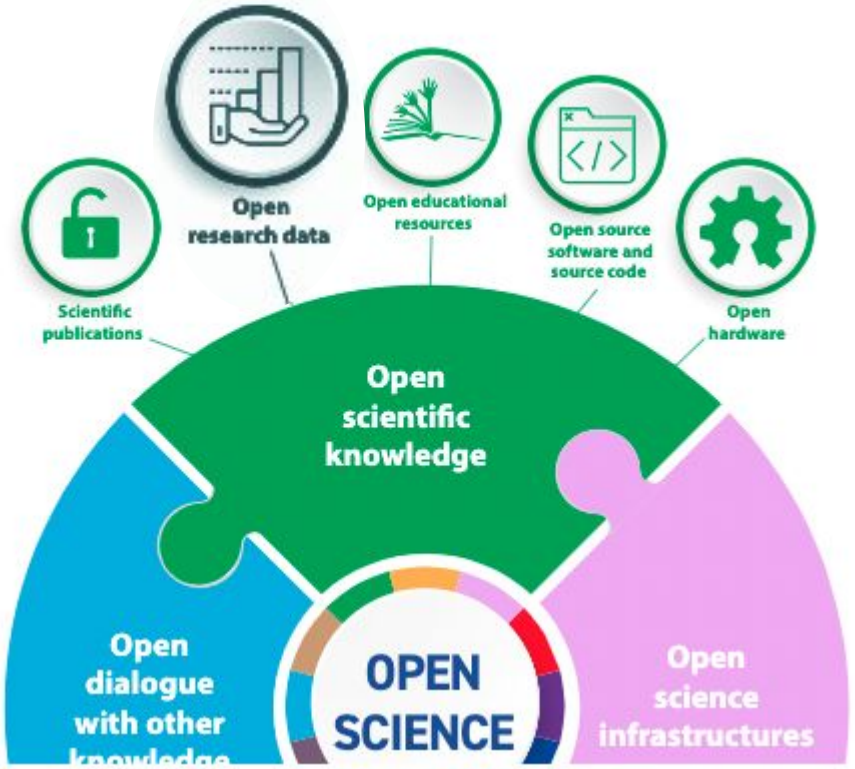
8 ambitions of the EU's open science policy

Open Data

FAIR (Findable, Accessible, Interoperable and Re-usable data) and open data sharing should become the default for the results of EU-funded scientific research.

European commission

https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science_en



UNESCO Recommendation on Open Science https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en

Open Data in face of global emergencies

Open research data has accelerated investigations during the pandemic.

TECHNOLOGY FEATURE | 24 April 2020

Open science takes on the coronavirus pandemic

NEWS | 03 February 2021

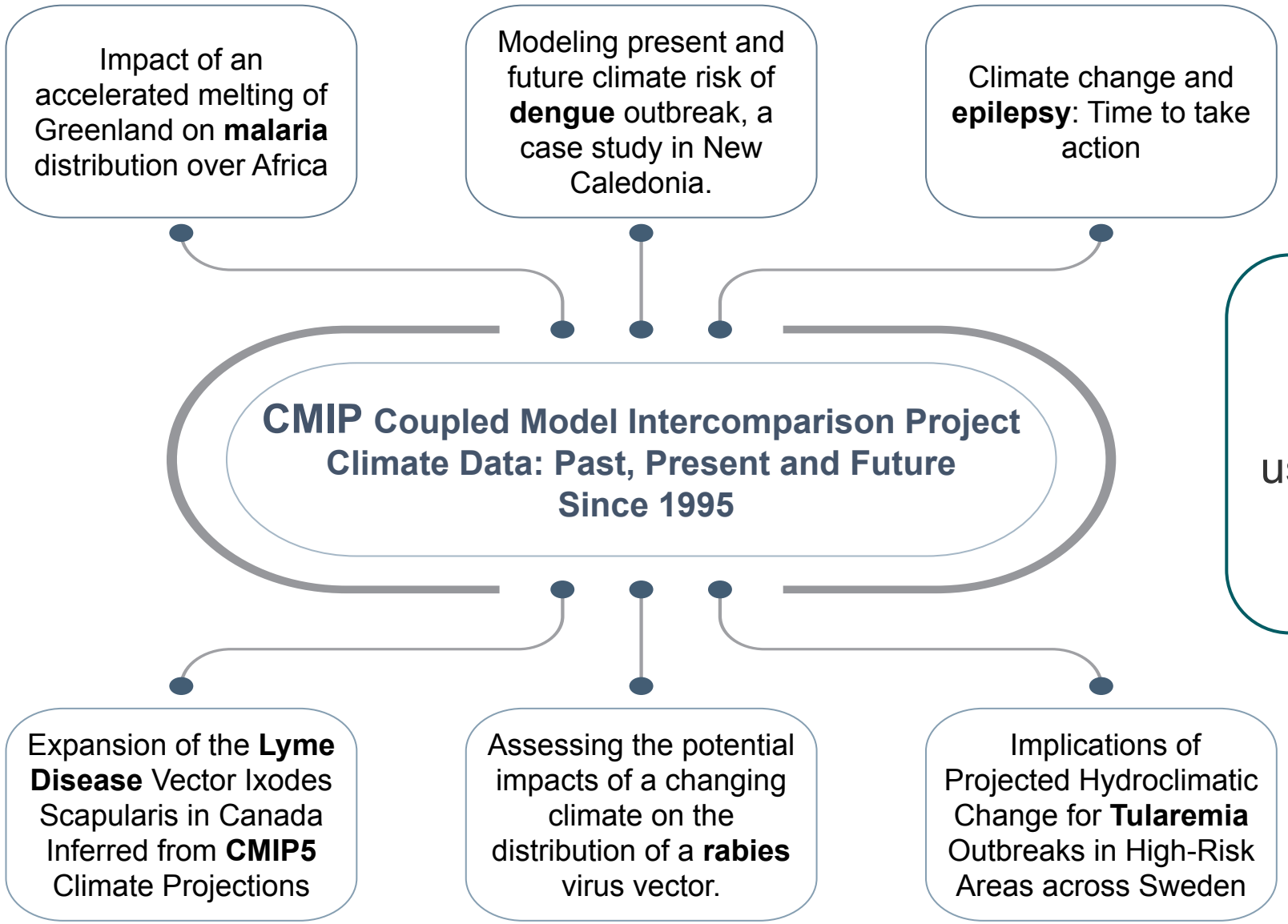
Scientists call for fully open sharing of coronavirus genome data

Open science saves lives: lessons from the COVID-19 pandemic

Lonni Besançon ✉, Nathan Peiffer-Smadja, Corentin Segalas, Haiting Jiang, Paola Masuzzo, Cooper Smout, Eric Billy, Maxime Deforet & Clémence Leyrat

BMC Medical Research Methodology **21**, Article number: 117 (2021) | Cite this article

Open Data in a multidisciplinary universe



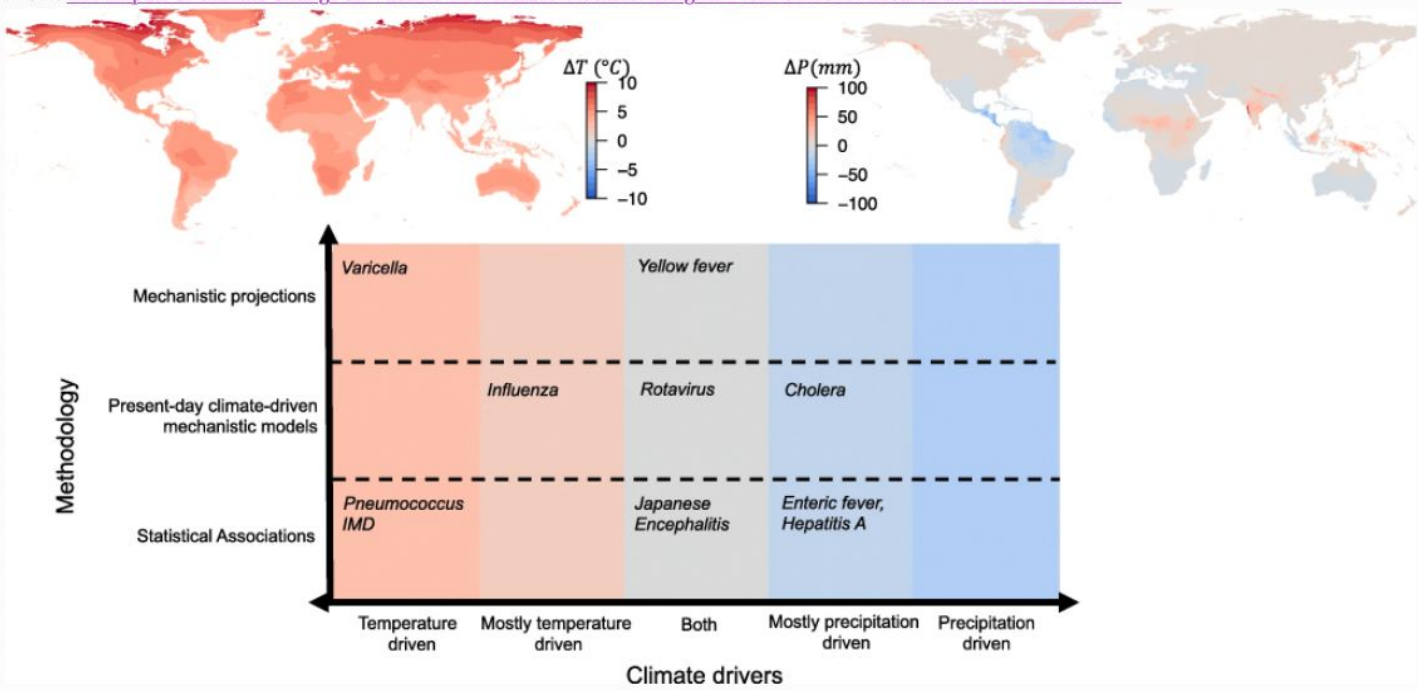
Open research data has facilitated cross-disciplinary use where shared data is being used beyond one discipline.

Open Data in a multidisciplinary universe

Open research data allows new scientific questions to be asked, through massive analysis or the federation of heterogeneous datasets

The Impact of Climate Change on **Vaccine-Preventable Diseases:** Insights From Current Research and New Directions

From: The Impact of Climate Change on Vaccine-Preventable Diseases: Insights From Current Research and New Directions



Top panel: maps show the CMIP6 multi-model mean projected change in temperature (ΔT) and precipitation (ΔP) in 2100 relative to 2000 under the Shared Socioeconomic Pathway (SSP) 3 "middle of the road" scenario, generated using Worldclim data [6]. Bottom panel: plot shows a summary of climate drivers (temperature or precipitation) for different vaccine-preventable diseases and the "best case" modeling effort reviewed, where we assume the best case is a fully mechanistic model using projection data. Absolute humidity drivers are counted under temperature-driven given the functional dependence of the two variables

Open Data real life implications

Open research data
informs policy

Environmental science

How Landscape Ecology Informs Global Land-Change Science and Policy FREE

Audrey L. Mayer, Brian Buma, Amélie Davis, Sara A. Gagné, E. Louise Loudermilk, Robert M. Scheller, Fiona K.A. Schmiegelow, Yolanda F. Wiersma, Janet Franklin

BioScience, Volume 66, Issue 6, 1 June 2016, Pages 458–469,
https://doi.org/10.1093/biosci/biw035
Published: 27 April 2016

Infrastructures and Urban design

Article | Open Access | Published: 11 December 2018

Degrading permafrost puts Arctic infrastructure at risk by mid-century

Jan Hjort ✉, Olli Karjalainen, Juha Aalto, Sebastian Westermann, Vladimir E. Romanovsky, Frederick E. Nelson, Bernd Etzelmüller & Miska Luoto

Nature Communications **9**, Article number: 5147 (2018) | Cite this article

Open Data real life implications

Environmental science

How Landscape Ecology Informs Global Land-Change Science and Policy FREE

Audrey L. Mayer, Brian Buma, Amélie Davis, Sara A. Gagné, E. Louise Loudermilk, Robert M. Scheller, Fiona K.A. Schmiegelow, Yolanda F. Wiersma, Janet Franklin

BioScience, Volume 66, Issue 6, 1 June 2016, Pages 458–469,
https://doi.org/10.1093/biosci/biw035
Published: 27 April 2016

Infrastructures and Urban design

Article | Open Access | Published: 11 December 2018

Degrading permafrost puts Arctic infrastructure at risk by mid-century

Jan Hjort ✉, Olli Karjalainen, Juha Aalto, Sebastian Westermann, Vladimir E. Romanovsky, Frederick E. Nelson, Bernd Etzelmüller & Miska Luoto

Nature Communications 9, Article number: 5147 (2018) | Cite this article

Open research data
informs policy

COVID-19

The Biden-Harris plan to beat COVID-19

The American people deserve an urgent, robust, and professional response to the growing public health and economic crisis caused by the coronavirus (COVID-19) outbreak. President Biden believes that the federal government must act swiftly and aggressively to help protect and support our families, small businesses, first responders, and caregivers essential to help us face this challenge, those who are most vulnerable to health and economic impacts, and our broader communities – not to blame others or bail out corporations.

The Biden-Harris administration will always:

- **Listen to science**
- **Ensure public health decisions are informed by public health professionals**

SciLifeLab

Is Open enough?

SciLifeLab

Is Open enough?

SciLifeLab

Is Open enough?

+ Reuse

Open data vs crate digging

Open data vs crate digging

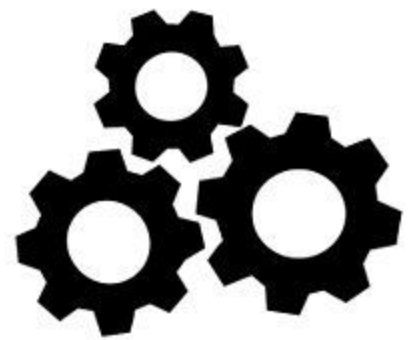
Discoverability is key!

Open data vs crate digging

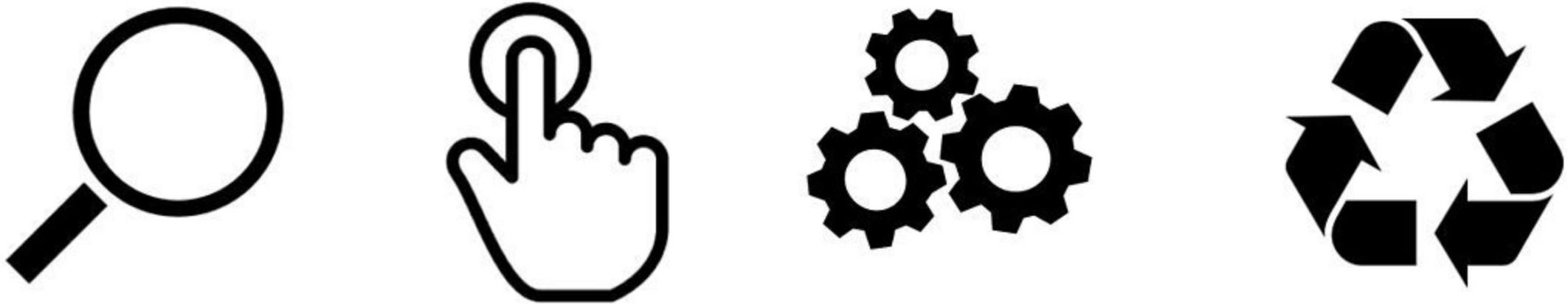
Discoverability is key!

“comprehensive, easy to obtain, easy to manipulate, and believable”

Findable Accessible Interoperable Reusable



Findable Accessible Interoperable Reusable



FAIR ≠ Open

FAIR Principles

Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons

Scientific Data **3**, Article number: 160018 (2016) | Cite this article

An Addendum to this article was published on 19 March 2019

Abstract

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles

Findable

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the **FAIRification process**.

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 they describe

F4. (Meta)data are registered or indexed in a searchable resource

Accessible

Accessible

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

Interoperable

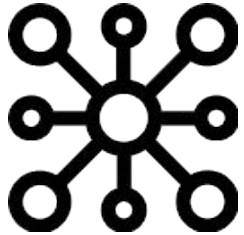
Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

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



Reusable

Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

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

SciLifeLab

How do we get there?

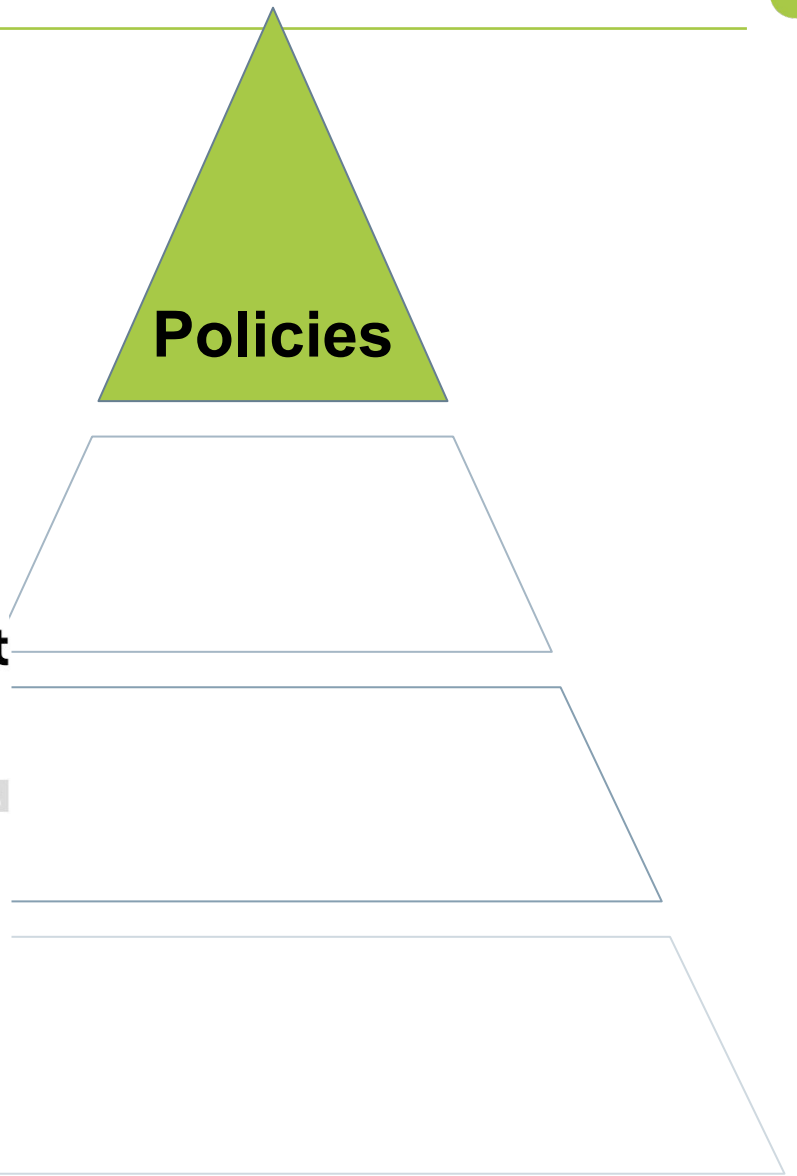
SciLifeLab

How do we get there?

Beyond the Open road lies change

Beyond the Open road lies change

Open science policies can remove obstacles, set the tone, promote funding and change the reward system



Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century

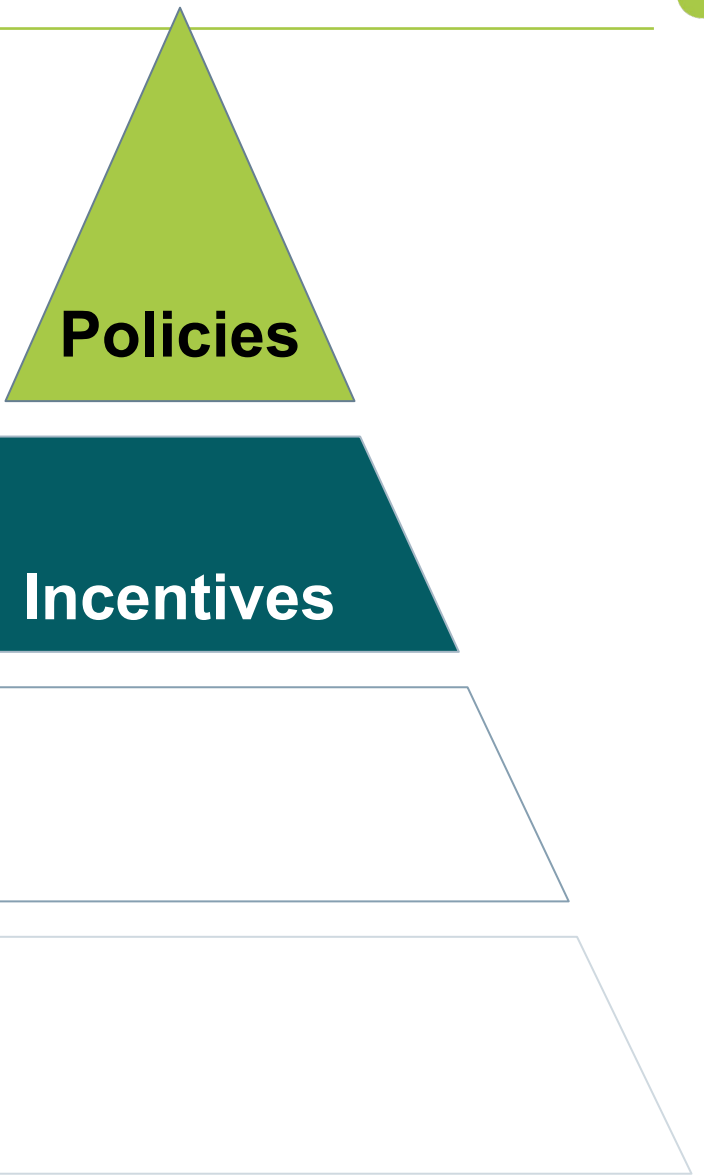
Jean-Claude Burgelman*, Corina Pascu*, Katarzyna Szkuta, Rene Von Schomberg, Athanasios Karalopoulos, Konstantinos Repanas and Michel Schouppe

Open Science, DG Research and Innovation, European Commission, Brussels, Belgium

Open science will make science more efficient, reliable, and responsive to societal challenges. The European Commission has sought to advance open science policy from its inception in a holistic and integrated way, covering all aspects of the research cycle from scientific discovery and review to sharing knowledge, publishing, and outreach. We present the steps taken with a forward-looking perspective on the challenges laying ahead, in particular the necessary change of the rewards and incentives system for researchers (for which various actors are co-responsible and which goes beyond the mandate of the European Commission). Finally, we discuss the role of artificial intelligence (AI) within an open science perspective.

Beyond the Open road lies change

Reward and recognition structures need new metrics accounting for diverse research outputs (incl. citation and acknowledgement for data & code)



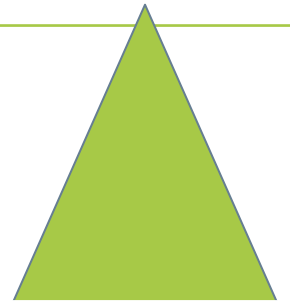
Problems/concerns with sharing data

over the last 4 years

	2018	2019	2020	2021
Concerns about misuse of data	518 36%	2082 37%	1881 38%	1920 43%
Not receiving appropriate credit or acknowledgement	478 33%	1834 32%	1531 31%	1739 39%

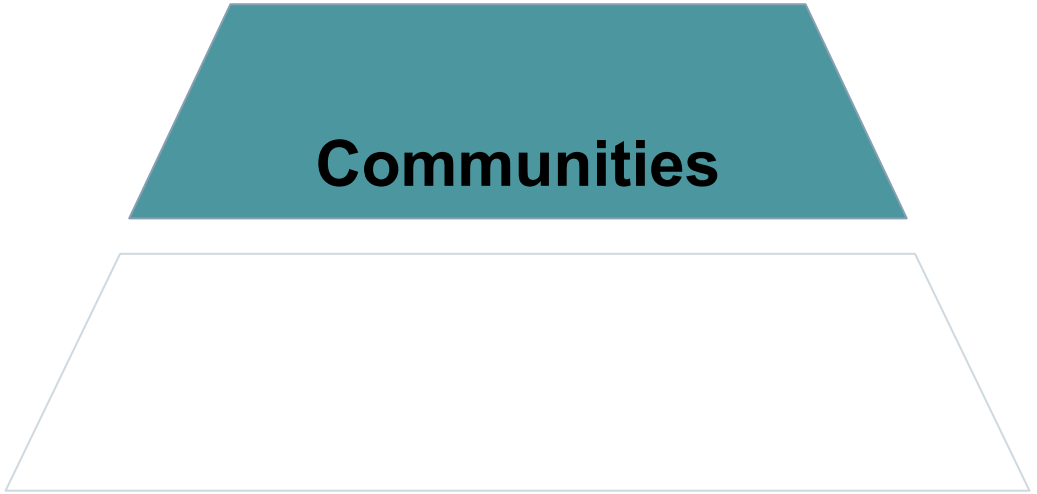
Beyond the Open road lies change

Communities drive cultural change, raise awareness and increase researchers engagement and adoption.



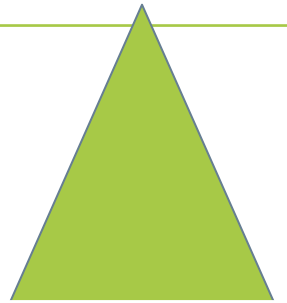
Community versus individual responsibility

In 2020, nearly 40% of surveyed researchers stated that they had never heard of the FAIR principles[97]. Even for those who had heard of the principles, only 25% of researchers felt they were "familiar" with them. The FAIR principles are comprehensive but technical, describing practices which touch on metadata, persistent identifiers, access protocols and repositories. Can we assume that even those researchers who are "familiar" with FAIR really understand its practical application?



Beyond the Open road lies change

Communities drive cultural change, raise awareness and increase researchers engagement and adoption.



Community versus individual responsibility

In 2020, nearly 40% of surveyed researchers stated that they had never heard of the FAIR principles[97]. Even for those who had heard of the principles, only 25% of researchers felt they were "familiar" with them. The FAIR principles are comprehensive but technical, describing practices which touch on metadata, persistent identifiers, access protocols and repositories. Can we assume that even those researchers who are "familiar" with FAIR really understand its practical application?

ISSUE BRIEF

May 13, 2019

Data Communities

A New Model for Supporting STEM Data Sharing

Danielle Cooper, Rebecca Springer

DOI: https://doi.org/10.18665/sr.311396

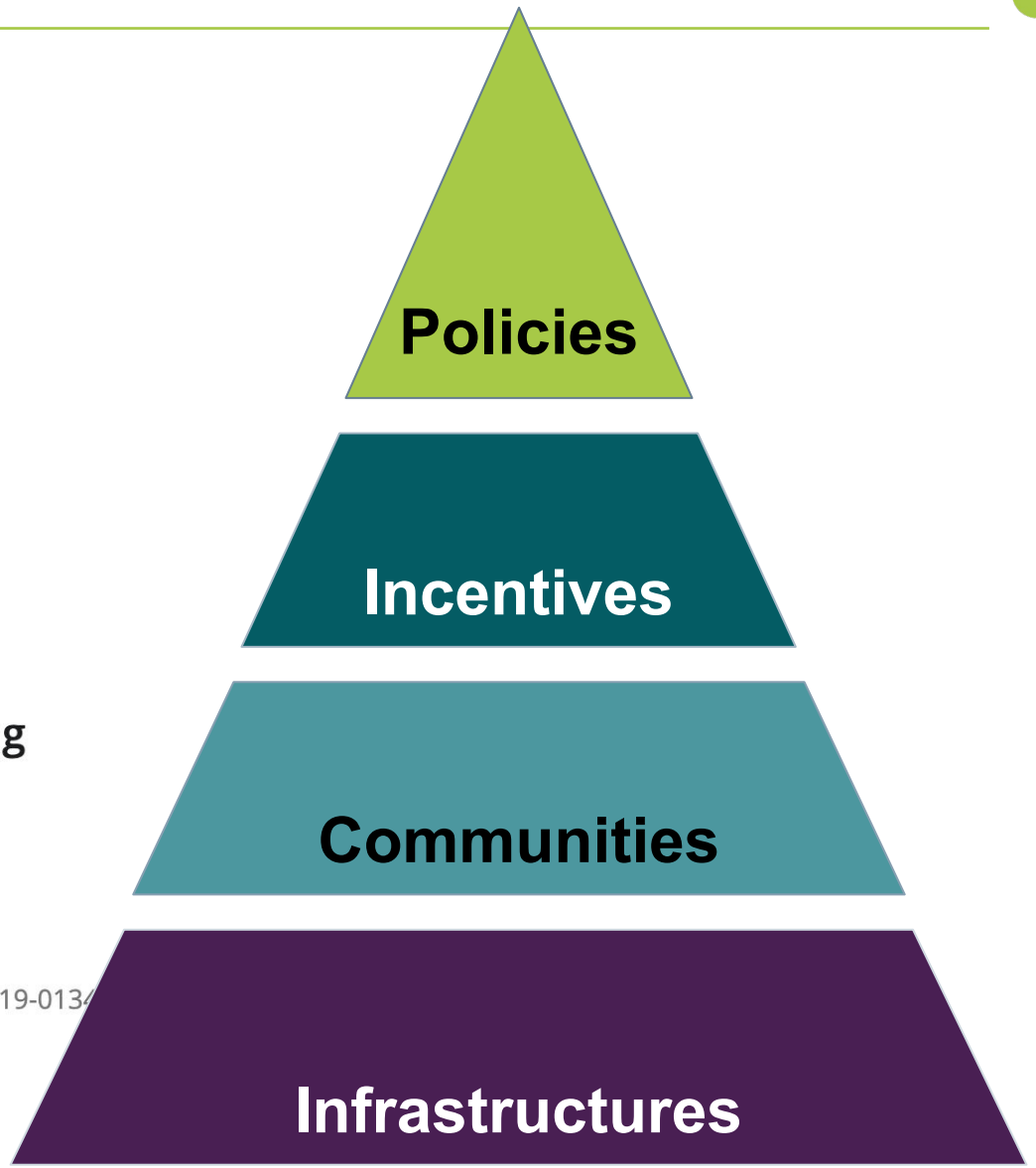
Topics: Digital scholarship and data management, Libraries, Research practices, Scholarly communication

Tags: Data communities

Beyond the Open road lies change

Human infrastructure;

- Research support staff for better quality data and higher reuse,
- Increased digital competency



RESEARCH ARTICLE | Open Access

How do properties of data, their curation, and their funding relate to reuse?

Libby Hemphill, Amy Pienta, Sara Lafia, Dharma Akmon, David A. Bleckley

First published: 23 March 2022 | https://doi.org/10.1002/asi.24646

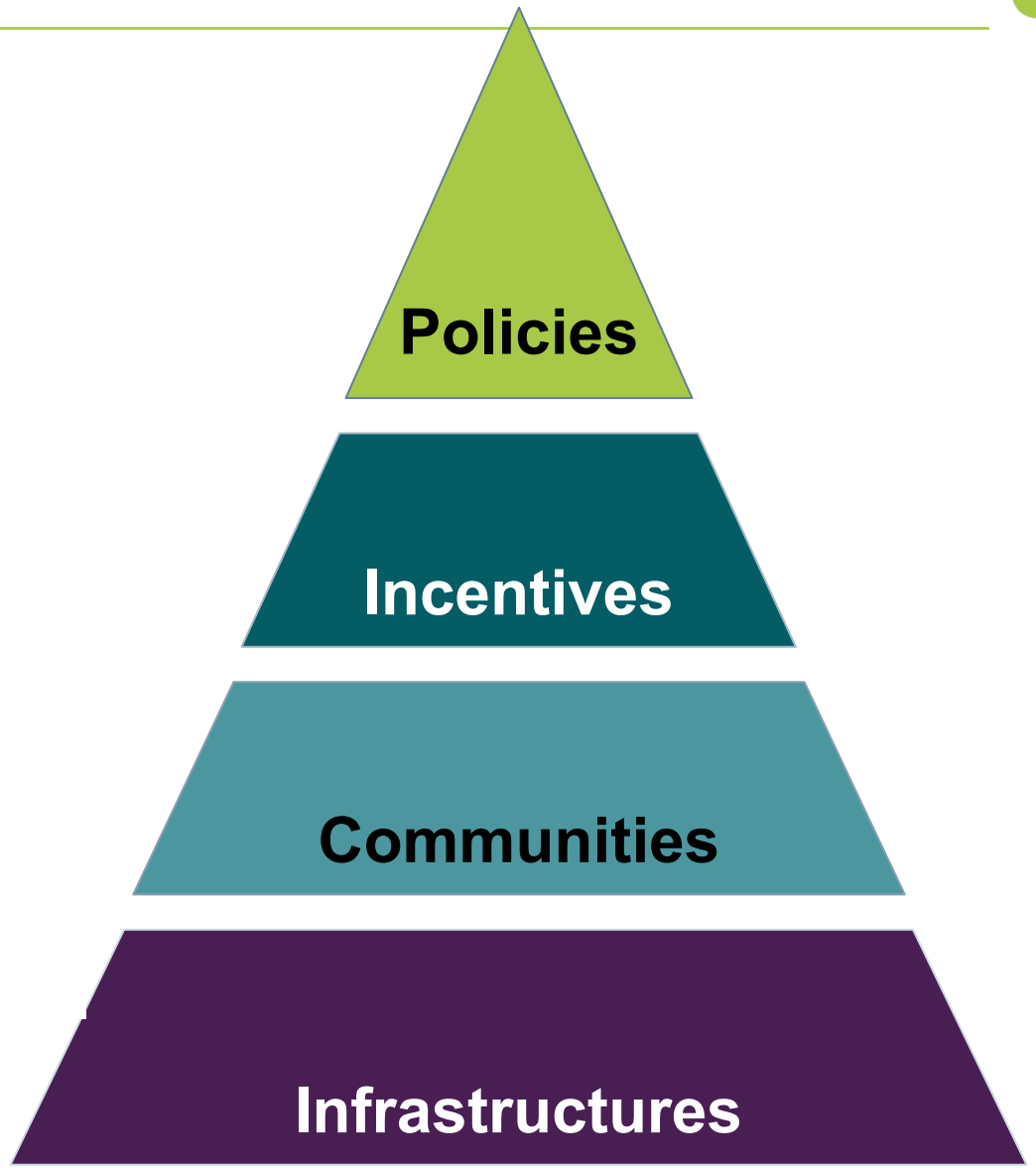
Funding information: Institute of Museum and Library Services, Grant/Award Number: LG-37-19-013
19; National Institute on Drug Abuse, Grant/Award Number: N01DA-14-5576; National Science Foundation, Grant/Award Number: 1930645

Beyond the Open road lies change

Technical infrastructure that is scalable and sustainable is necessary for handling large amounts of data.

" Hurdles to data sharing in the area of policy and cultural change will fall short if we do not have underpinning research infrastructure and the experts needed to run the infrastructure. "

[The State of Open Data 2021]
The longest-running longitudinal survey and analysis on open data

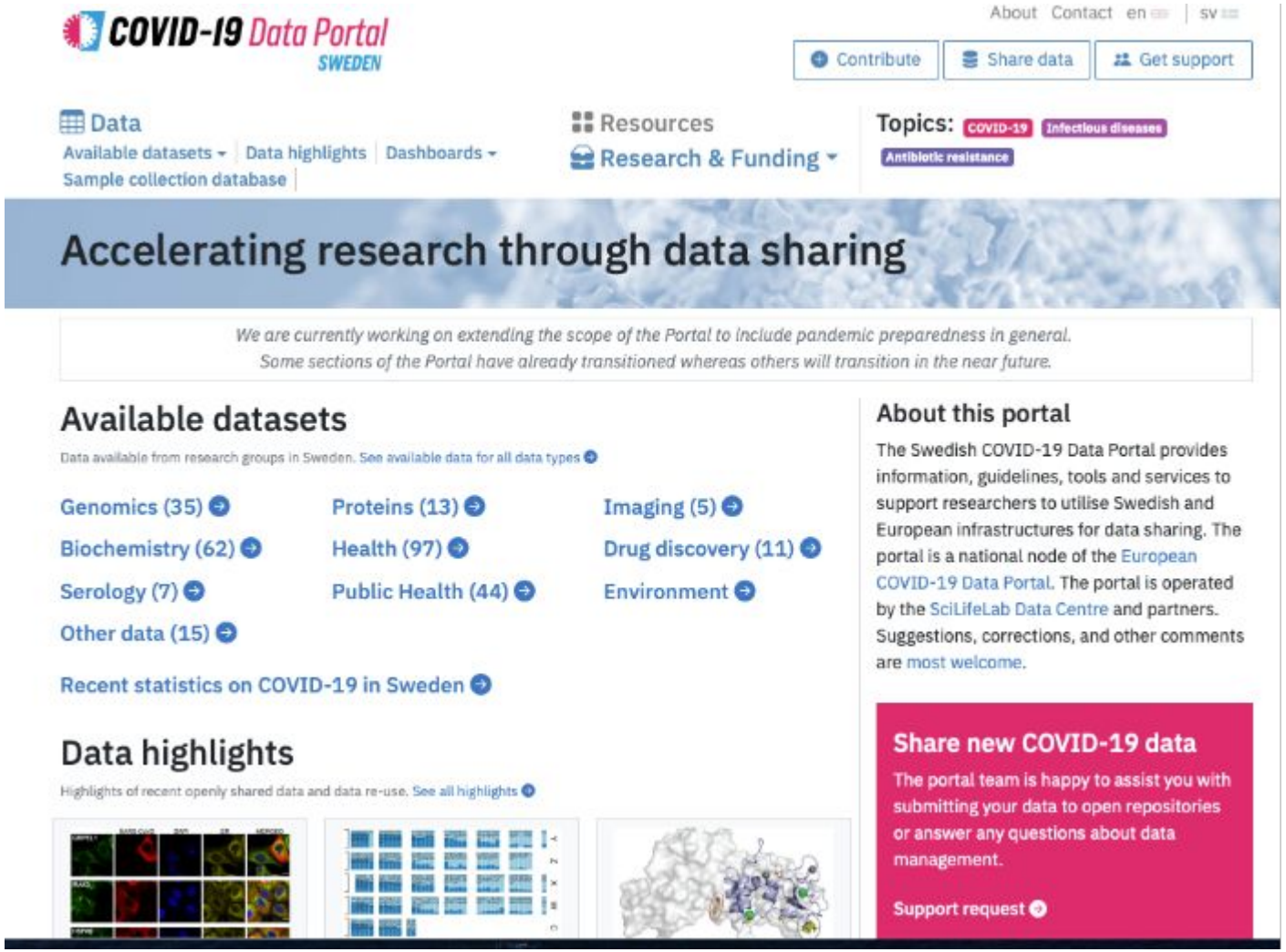


SciLifeLab

Tools & Services offered at the SciLifeLab

Swedish COVID-19 Data Portal

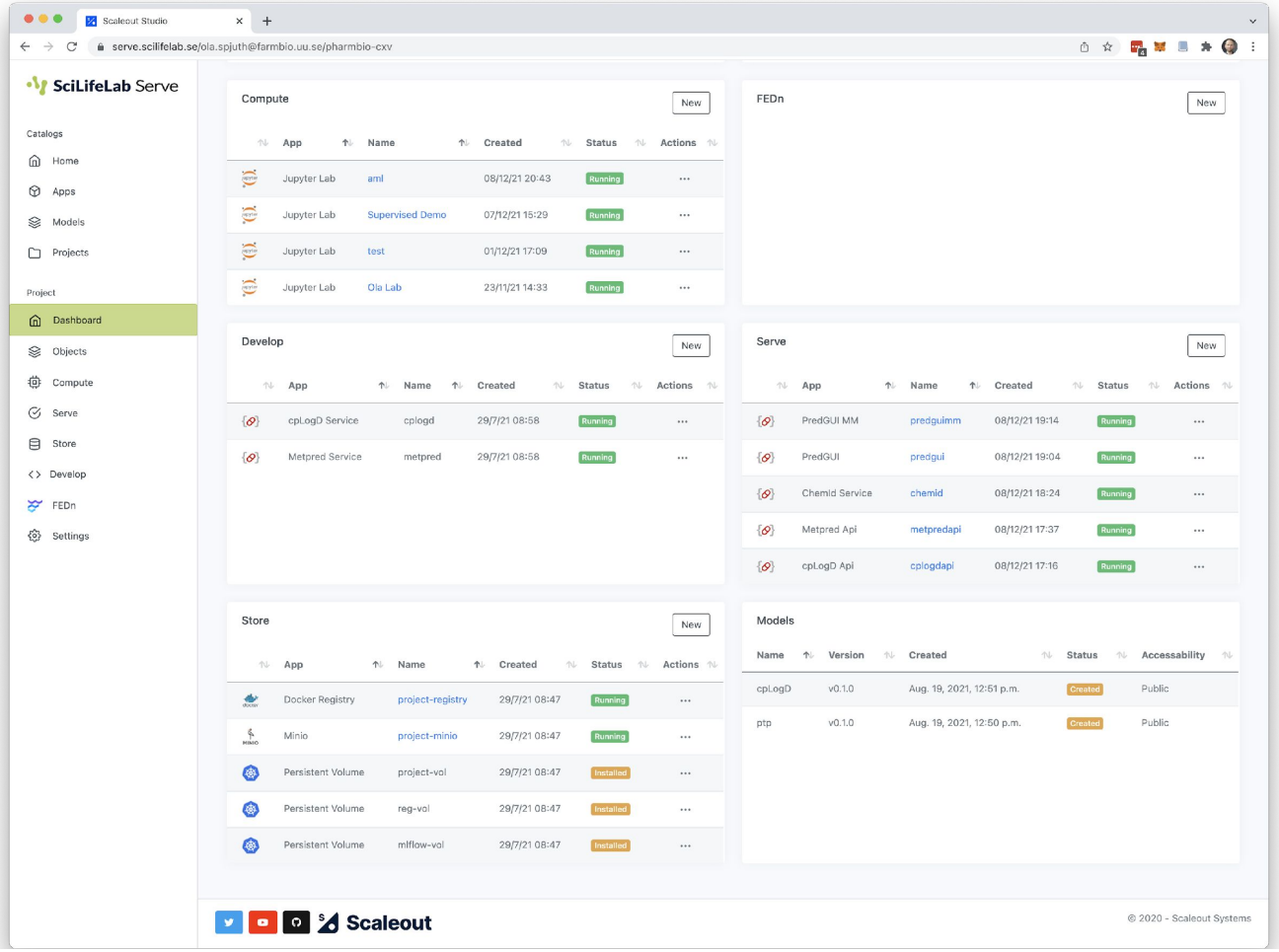
- Launched June 2020
- Focus on FAIR data sharing
- Provides information, guidelines, tools and services to support researchers
- Data management support for Swedish COVID-19 research programs



https://covid19dataportal.se

SciLifeLab Serve

- Serve apps, models, notebooks etc
- Manage life cycle of AI models (MLOps), improve FAIR[1]
- Running on Kubernetes
- Effort at SciLifeLab Data Center



A high-level overview of the machine learning life cycle

https://serve.scilifelab.se

[1] Spjuth O, Frid J, and Hellander A.
The Machine Learning Life Cycle and the Cloud: Implications for Drug Discovery
Expert Opinion On Drug Discovery. 16, 9, 1071-1079. (2021).
DOI: 10.1080/17460441.2021.1932812

SciLifeLab Data Repository

- Institutional Figshare instance
- General institutional repository for publishing
- Publish any kind of research-related data, e.g. documents, figures, or presentations
- Log in using SWAMID
- Data is made citable through its DOI

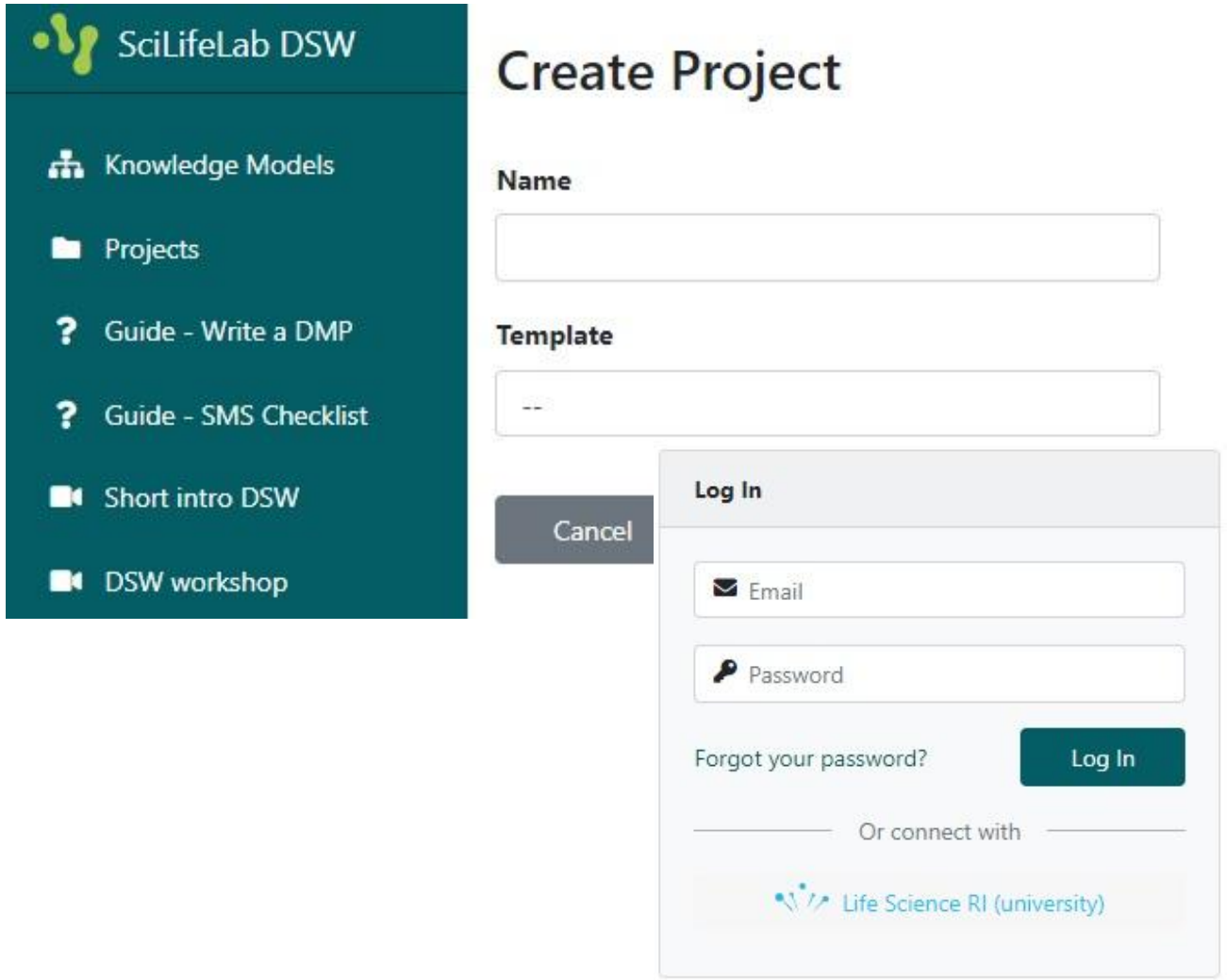
Repository information: https://www.scilifelab.se/data/repository/

Submission guidelines: https://www.scilifelab.se/data/repository/submission/

Repository URL: https://scilifelab.figshare.com

Data Stewardship Wizard

- A tool to **create data management plans** based on interactive questionnaires
- LifeScience **University log-in**, collaborative editing, versioning, and exports to PDF/Word using templates
- SciLifeLab provide templates and questionnaires **based on national and life-science specific guidelines**



https://www.dsw.scilifelab.se

https://dsw.scilifelab.se/appendix/dsw_instructions

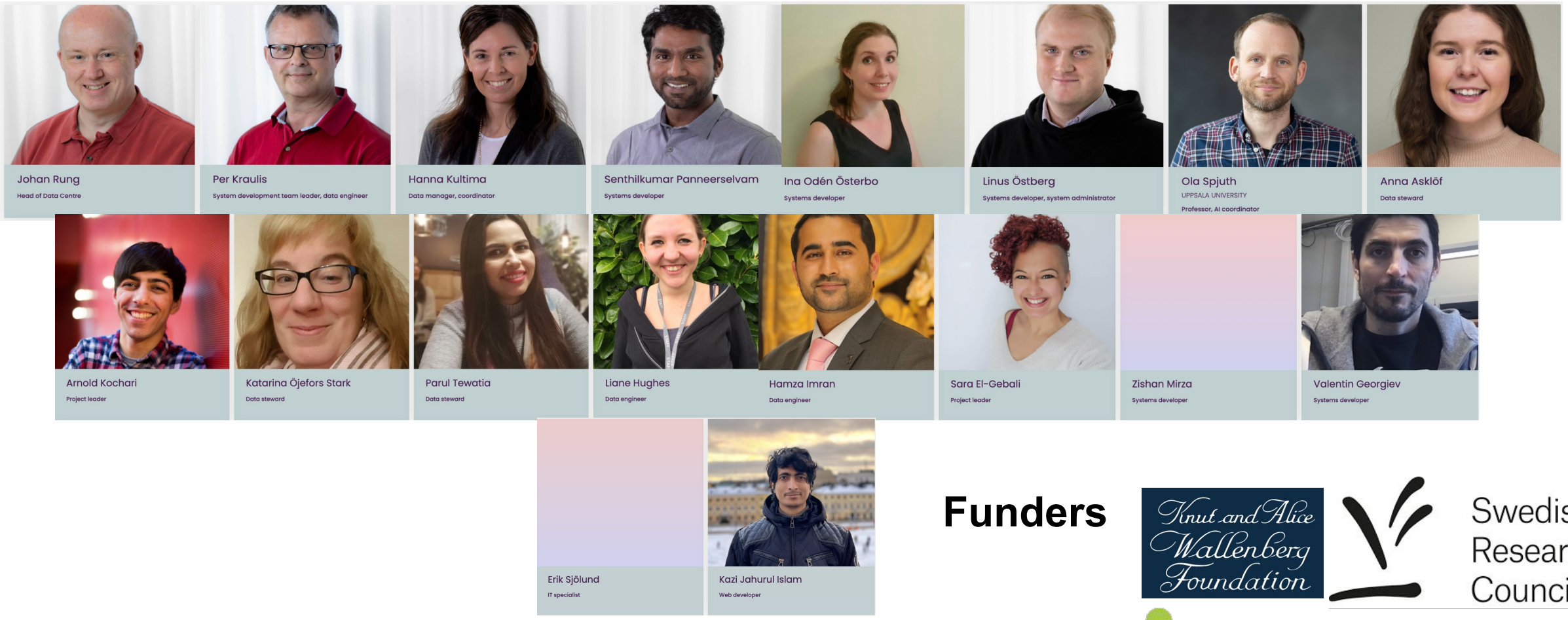
Stay in touch

Email: datacentre@scilifelab.se

Twitter: @scilifelab_DC

Linkedin: scilifelab-data-centre

Thank you!



Funders

Knut and Alice Wallenberg Foundation

Swedish Research Council

SciLifeLab

References

- Open Science takes on the coronavirus pandemic https://www.nature.com/articles/d41586-020-01246-3
- Scientists call for fully open sharing of coronavirus genome data https://www.nature.com/articles/d41586-021-00305-7
- Impact of an accelerated melting of Greenland on malaria distribution over Africa. https://doi.org/10.1038/s41467-021-24134-4
- Modeling present and future climate risk of dengue outbreak, a case study in New Caledonia.https://doi.org/10.1186/s12940-022-00829-z
- Climate change and epilepsy: Time to take action. https://doi.org/10.1002/epi4.12359
- Expansion of the Lyme Disease Vector Ixodes Scapularis in Canada Inferred from CMIP5 Climate Projections https://pubmed.ncbi.nlm.nih.gov/28599266/
- Assessing the potential impacts of a changing climate on the distribution of a rabies virus vector. https://doi.org/10.1371/journal.pone.0192887
- Implications of Projected Hydroclimatic Change for Tularemia Outbreaks in High-Risk Areas across Sweden https://doi.org/10.3390/ijerph17186786
- The Impact of Climate Change on Vaccine-Preventable Diseases: Insights From Current Research and New Directions. https://doi.org/10.1007/s40572-020-00293-2
- How Landscape Ecology Informs Global Land-Change Science and Policy https://academic.oup.com/bioscience/article/66/6/458/2754255?login=false
- Degrading permafrost puts Arctic infrastructure at risk by mid-century https://www.nature.com/articles/s41467-018-07557-4
- The Biden-Harris plan to beat COVID-19 https://www.whitehouse.gov/priorities/covid-19/
- Has Biden followed the science? What researchers say https://www.nature.com/articles/d41586-022-00108-4
- The FAIR Guiding Principles for scientific data management and stewardship https://www.nature.com/articles/sdata201618
- Go FAIR- FAIR Principles https://www.go-fair.org/fair-principles/
- Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century https://www.frontiersin.org/articles/10.3389/fdata.2019.00043/full
- The State of Open Data 2021 https://digitalscience.figshare.com/articles/report/The_State_of_Open_Data_2021/17061347
- THE FUTURE OF FAIR- Highlights and reflections from the Better Research Through Better Data roundtable https://go.sn.pub/the-future-of-fair
- Data Communities A New Model for Supporting STEM Data Sharing https://sr.ithaka.org/publications/data-communities/
- How do properties of data, their curation, and their funding relate to reuse? https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/asi.24646
- Turning FAIR into reality https://data.europa.eu/doi/10.2777/1524
- Invest 5% of research funds in ensuring data are reusable https://www.nature.com/articles/d41586-020-00505-7
- The EU's open science policy https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science_en#documents
- UNESCO Recommendation on Open Science https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en
- European commission https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science_en