# Data Democratisation: Cultural, Institutional and Infrastructural Dilemmas of Data Sharing

Sabina Leonelli
Exeter Centre for the Study of Life Sciences (Egenis)
& Department of Sociology, Philosophy and Anthropology
University of Exeter
@sabinaleonelli

The Alan Turing Institute

erc
European Research Council
Established by the European Commission

University of Exeter

egenis

# Trouble in Science*

Global context rife with

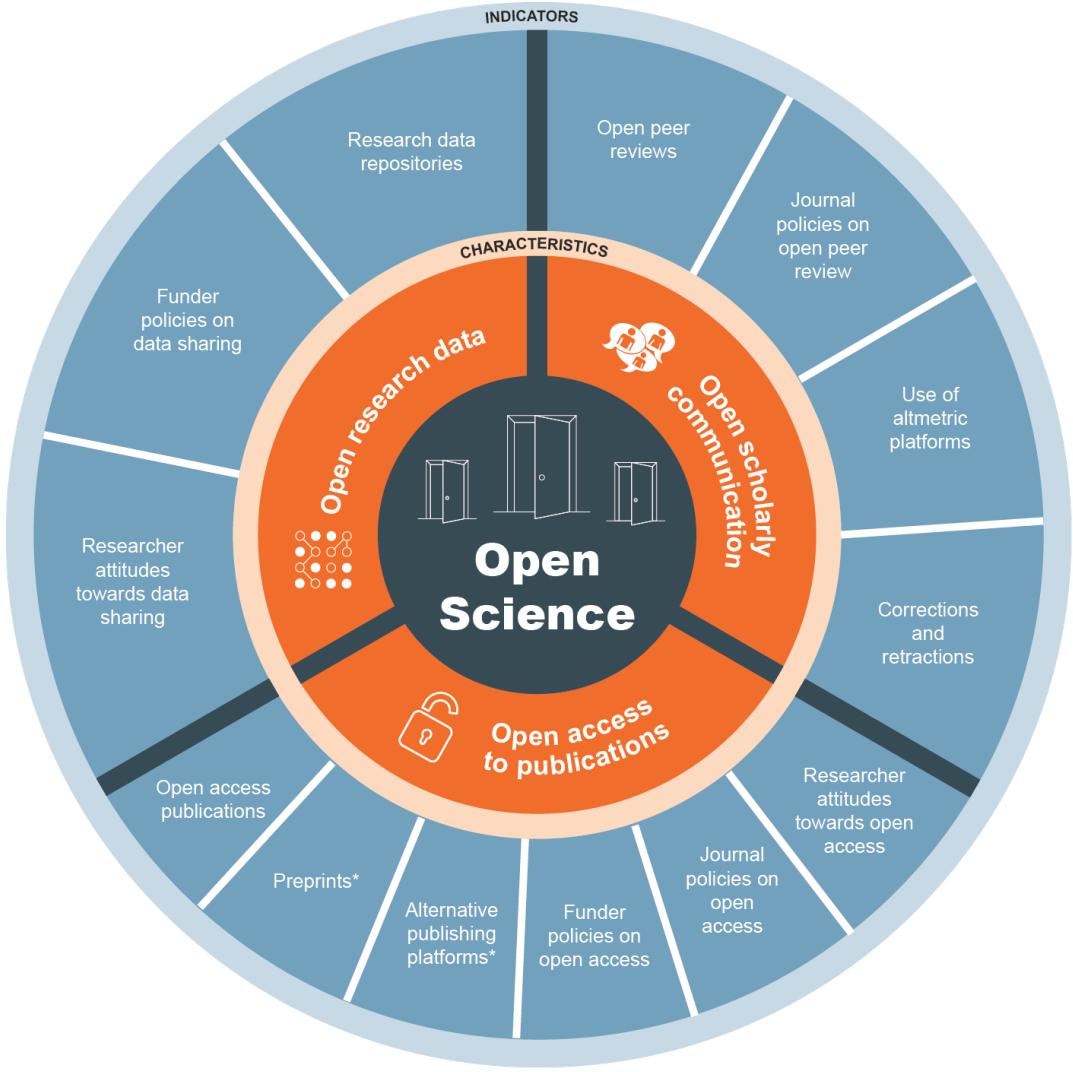
- inequity
- diverse and often problematic incentive systems
- control by / dependence on inscrutable companies and infrastructures
- various forms of discrimination
- lack of accountability and public trust
- a short-term understanding of scientific, political and economic benefits

* ..and well beyond of course!! But let’s zoom in on the world of research for now..

# Is moving to ”open” a solution?

“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.. [..] .. sharing and using all available knowledge at an **earlier stage** in the research process”

Carlos Moedas, *Open Innovation, Open Science, Open to the World (2015)*



# Open Data: The Central Challenge

- Data play a key role and exemplify key challenges:
  - new prominence as research outputs
    - recognised as valuable in their own right
    - mobility and re-use are central to data value
    - relation to articles (and related credit) is being redefined
    - serious issues with quality and with responsible dissemination
  - responsible data management *could* foster:
    - post-COVID global transformation of research and its role in decision-making
    - equitable participation in the creation of knowledge, through data stewardship that is transparent, subject to scrutiny and grounded on a commitment to justice and fairness
    - rethinking of policy, funding, evaluation and practice of science systems
- Under which conditions can this work?

# Consider the pandemic

- Unassailable demonstration of the value of OS?
  - Acceleration of discovery (e.g. Open Access shift)
  - Revindication of value of big OS platforms
    - USA: “COVID for you” initiative
    - International: RDA COVID-19 Working Group
    - UK
      - Open Data from UK Office for National Statistics; SAIL; “Data Loch” repository of all routine health and social care data for the Edinburgh and South East Scotland Region; CHESS (Covid-19 Hospitalisation in England Surveillance System) adapted from the UK Severe Influenza Surveillance System by Public Health England
    - Existing activist networks around specific diseases
      - E.g. EULAR COVID-19 Database established to capture how rheumatology conditions and their treatment affected the risk of and severity of COVID-19

# Consider the pandemic

History and Technology
An International Journal

Historiographic Essay

**Mobilizing the Transnational History of Knowledge Flows. COVID-19 and the Politics of Research at the Borders**

John Krige & Sabina Leonelli

Published online: 10 Mar 2021

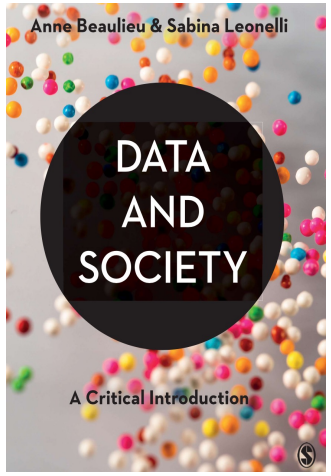
- But unassailable success this is not!
  - Huge technical “sharing” issues (data, models, software etc)
  - Lack of clarity over rights and obligations pertaining to “sharing”
  - Lack of consultation and collaboration with relevant communities and disciplines
  - Large-scale exploitation of data accumulated on and through patients from around the world
  - Polemics on data governance and access, esp. transnationally

# Consider the pandemic

- Turing Report on “data readiness” (June 2021):
  - UK (like many other countries) woefully unprepared in data infrastructures
  - lots of unusable medical data due to lack of metadata and appropriate domain expertise
  - international data sharing (e.g. from Northern Italy) proved essential

- The debate around GISAID platform:
  - Set up to share influenza genomic data *securely* and *responsibly*
  - Grounded on database agreement governing access and re-use of data
  - Fostered trust and exchange over SARS-COV-2 data
  - Attacked for “not being open enough” by prominent Global North researchers (*Nature*, Jan 2021) and Leopoldina report (Febr 2021), among others

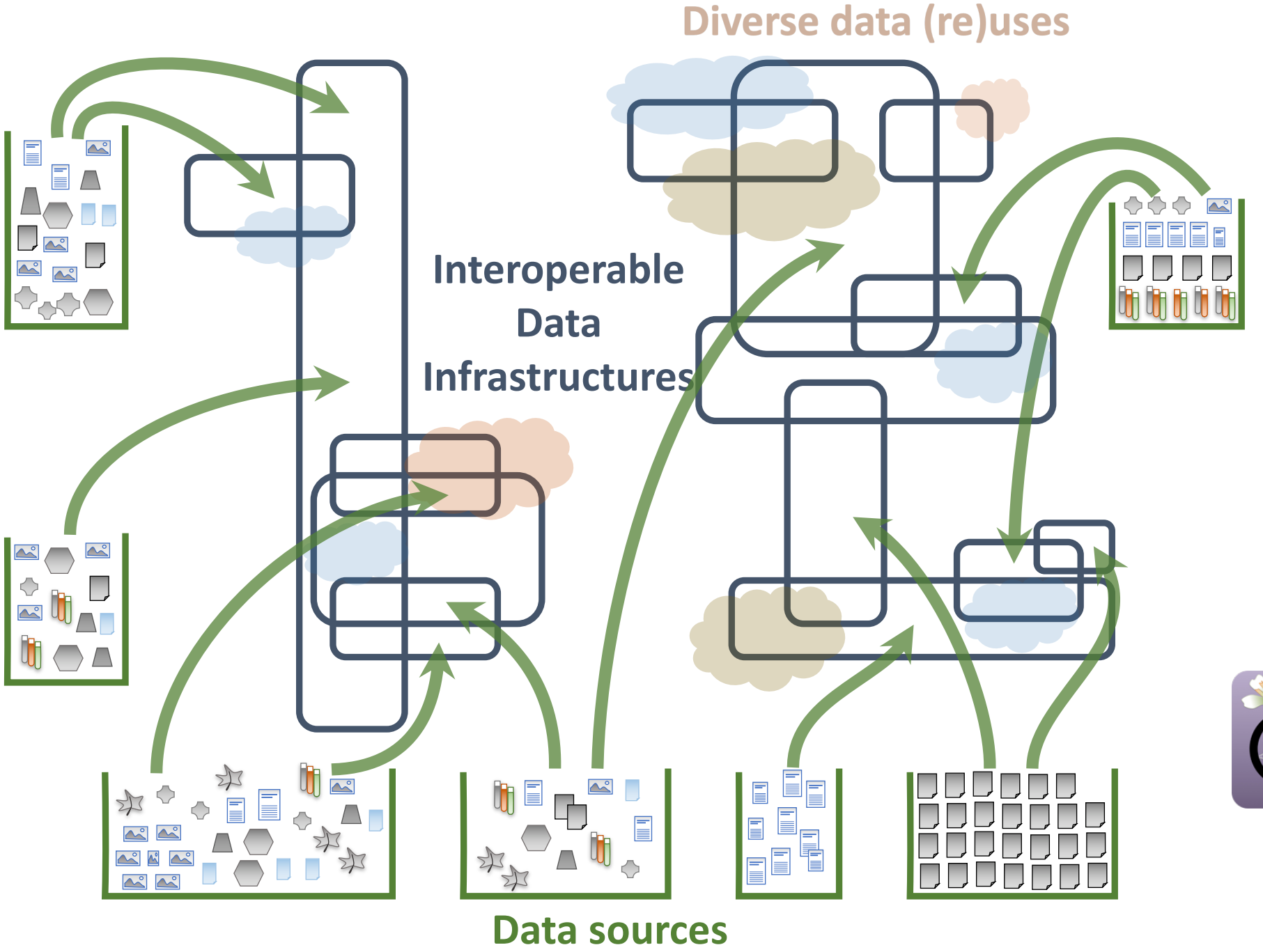
# The Challenges of Open Data


- Forging tools for unregulated mass surveillance of human behavior at individual as well as community levels
- Expanding existing divides and silencing knowledge from low-resourced environments and 'unfashionable' topics
- Privileging re-use: what does it mean for creativity and innovation?
- Eroding expertise and centuries-old methodological wisdom: 'anything online goes'
- Eroding trust and credibility of science: exponential growth of opportunities for marketing "alternative facts"
- Producing unreliable knowledge that does not help to tackle urgent social challenges (Leonelli 2016; Bezuidenhout et al 2017; Leonelli 2018, 2019; Beaulieu & Leonelli 2021)

# From Theory to Practice

- Opening new spaces, challenging traditional communication channels and power structures, encouraging participation “from below”…
- .. or reinforcing conservatism, bias, exclusion, discrimination and inequity?
- The proof is in the pudding… in this case, **implementation**




European Research Council

Established by the European Commission

DATA_SCIENCE

datastudies.eu

# ESRC: “Open Innovation” + “Beyond the Digital Divide”; GYA “Global Access to Open Software”

*Science and Public Policy*, 44(4), 2017, 464–475
doi: 10.1093/scipol/scw036
Advance Access Publication Date: 13 July 2017
Article

OXFORD

## Beyond the digital divide: Towards a situated approach to open data

Louise M. Bezuidenhout[1,2,*], Sabina Leonelli[1], Ann H. Kelly[1,3] and Brian Rappert[1]


Koen Vermeir, Sabina Leonelli, Abdullah Shams Bin Tariq, Samuel Olatunbosun Sojinu, Augustine Ocloo, Md. Ashraful Islam Khan, Louise Bezuidenhout

*Article*

## How Do Scientists Define Openness? Exploring the Relationship Between Open Science Policies and Research Practice

Nadine Levin[1], Sabina Leonelli[2], Dagmara Weckowska[3], David Castle[4], and John Dupré[2]

Bulletin of Science, Technology & Society
2016, Vol. 36(2) 128–141

DOI: 10.1177/0270467616668760
bst.sagepub.com

SAGE

## How Does One “Open” Science? Questions of Value in Biological Research

Nadine Levin[1] and Sabina Leonelli[2]

Science, Technology, & Human Values
1-26

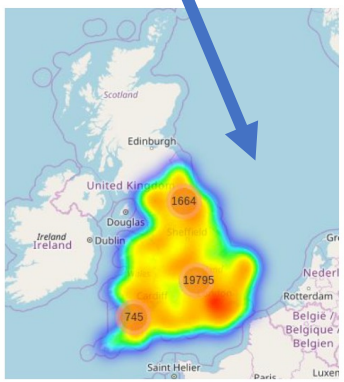
DOI: 10.1177/0162243916672071
sthv.sagepub.com

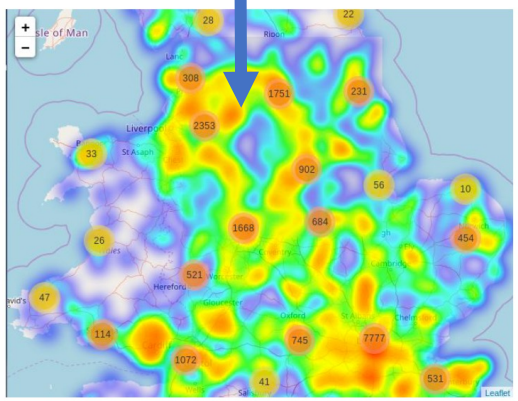
SAGE

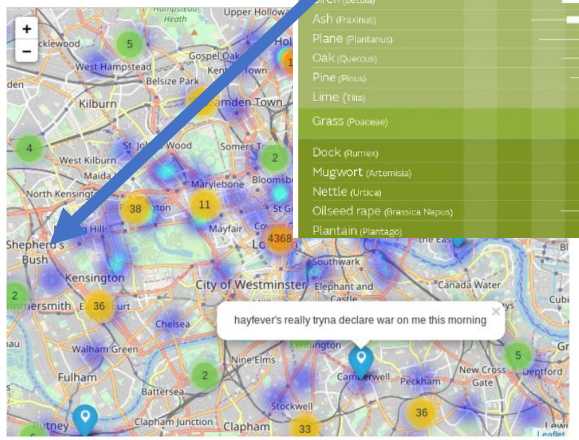
GLOBAL YOUNG ACADEMY

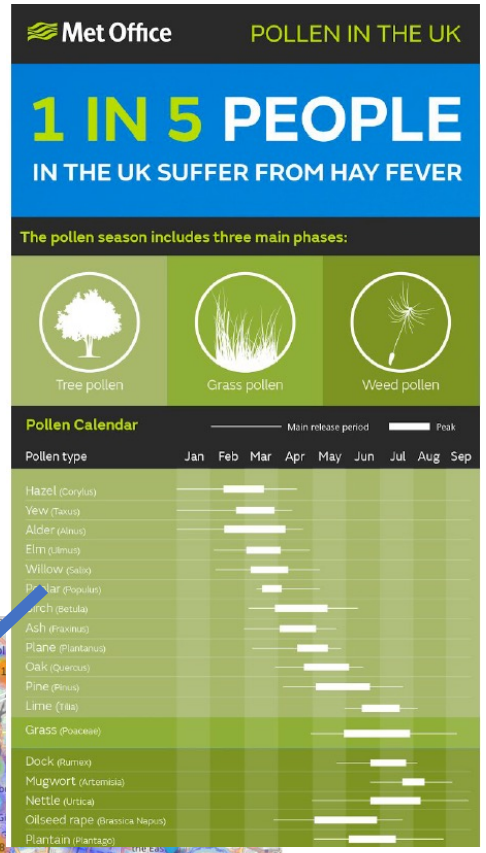
# ESRC - from social media to public health





The Alan Turing Institute

Home + Research + Research projects

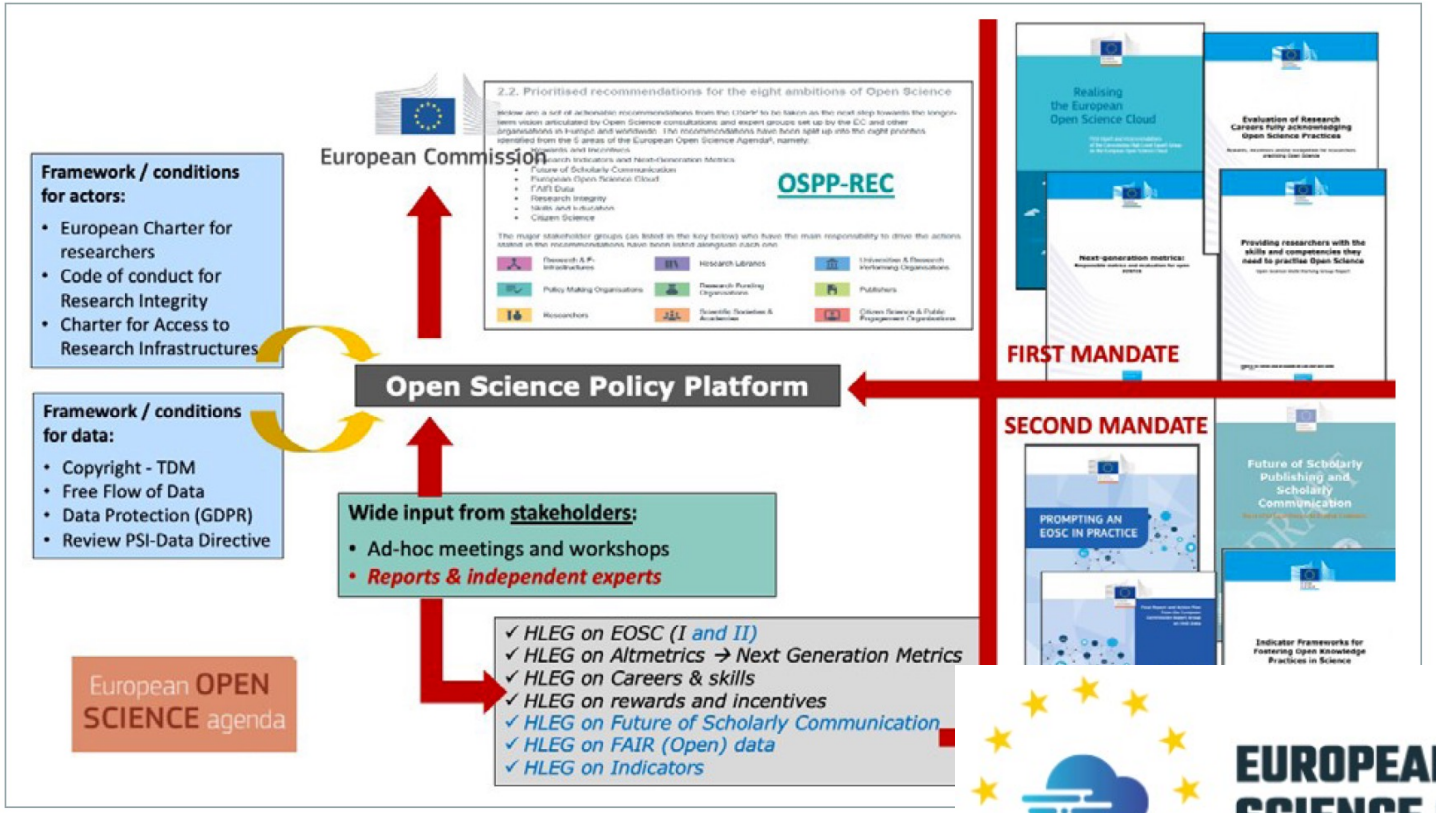
# From field data to global indicators

## Investigating how plant data can be efficiently and reliably linked across data infrastructures around the world

Related programmes

**Data science for science**

# Open Science Policy Platform (2016-20)




European Commission

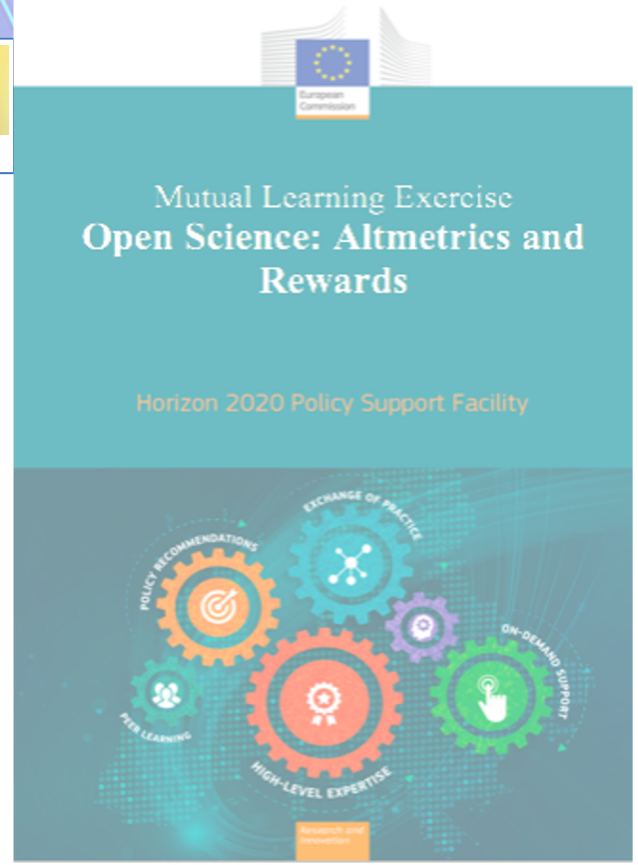
**Progress on Open Science: Towards a Shared Research Knowledge System**

EUROPEAN OPEN SCIENCE CLOUD

Final Report of the Open Science Policy Platform


Jan 2017 – Jan 2018

https://ec.europa.eu/h2020-policy-support-facility




Three topics:

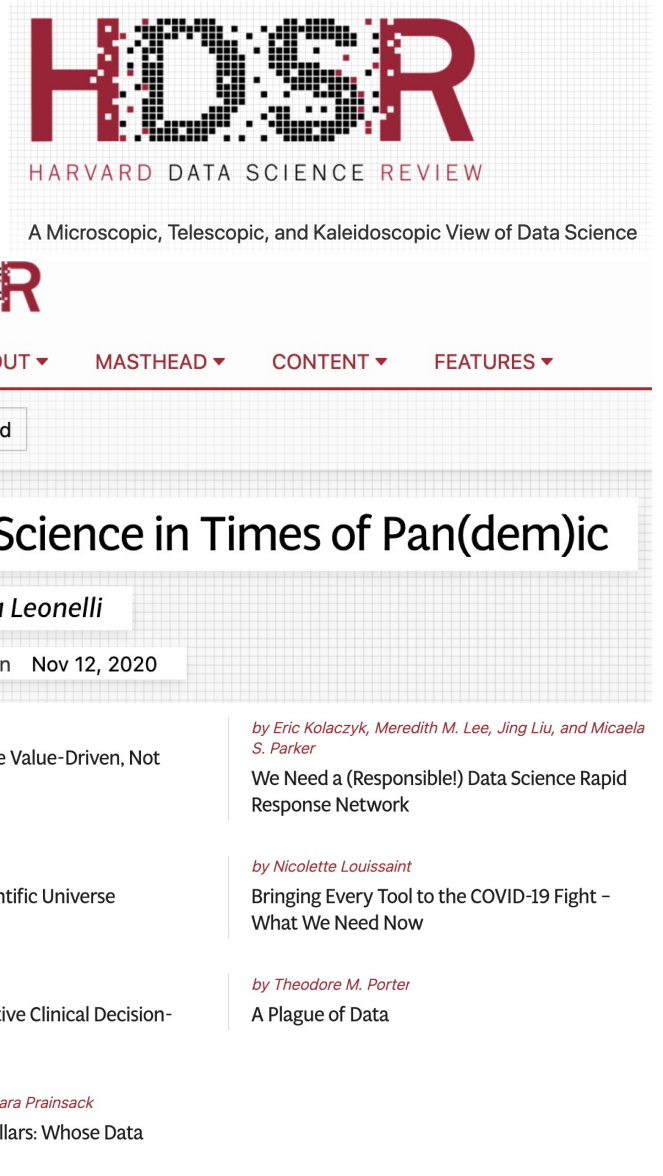
1. The potential of altmetrics to foster Open Science
2. Incentives and rewards for researchers to engage in Open Science activities
3. Guidelines for developing and implementing national policies for Open Science

# Dilemmas for Data Sharers

1. evaluation and credit systems
2. diversity in research cultures
3. costs and accountabilities
4. skills and training
5. intellectual property regimes
6. diversity and applicability of legal frameworks
7. semantic ambiguity
8. ethical concerns
9. high resource bias
10. infrastructural inequity and discrimination

# Data sharing requires hard thinking

- Trying to make data scrutinizable and re-usable, while remaining mindful of their social and political value
  - No, data are not neutral facts..
- Important to investigate reasons for mistrust
  - how to interpret collaboration and co-authorship
  - downstream data use
  - Open Data as strong expectation and yet difficult given existing inequities
- Community engagement & regular debate over what counts as “public benefit”
- Research design is key



# OS needs to be a platform for critical, informed and inclusive debate

- Shape OS policies and practices around consultation across diverse perspectives
  - Quality criteria for all research components are community-specific and value-laden
  - Variability in research conditions needs to be studied and integrated into OS governance, infrastructures and mechanisms for responsible sharing
- Distribute burdens associated with OS implementation
  - Recognize inequity of global research landscape and urgency of decreasing the digital and resource divide
  - Foster research that documents such inequity and its implications (social science & humanities)
- OS needs to promote dialogue on what counts as science, scientific infrastructures and scientific governance, and how results should be credited and disseminated

**A Philosophy of Open Science for Diverse Research Environments (PHIL_OS)**

1 September 2021 - 31 August 2026

# Thank you for your attention!


European Research Council

Established by the European Commission

And many thanks to my colleagues in Exeter,
international collaborators,
the Wissenschaftskolleg zu Berlin,
the European Research Council (PHIL_OS award ) and
the Alan Turing Institute (EPSRC award “From Local Fields to Global Indicators”)
for their crucial support

Science and Public Policy, 44(4), 2017, 464–475
doi: 10.1093/scipol/scw036
Advance Access Publication Date: 13 July 2017
Article

OXFORD

## Beyond the digital divide: Towards a situated approach to open data

Louise M. Bezuidenhout[1,2,*], Sabina Leonelli[1], Ann H. Kelly[1,3] and Brian Rappert[1]

DATA SCIENCE JOURNAL

Leonelli, S, et al. 2018 In
and Africa. *Data Science*
DOI: https://doi.org/10.5

EDITORIAL CONTENT

## Introduction: Open Data and Africa

Sabina Leonelli[1], Brian Rappert[1] and Louise Bezuidenhout[2]

*Article*

## How Does One "Open" Science? Questions of Value in Biological Research

Science, Technology, & Human Values
1-26

DOI: 10.1177/0162243916672071
sthv.sagepub.com
SAGE

Nadine Levin[1] and Sabina Leonelli[2]

- Upcoming **2021**: Beaulieu & Leonelli *Data and Society: A Critical Introduction*. SAGE.
- Upcoming **2022**: Leonelli *The Philosophy of Open Science.* Cambridge University Press.

## From FAIR data to fair data use: Methodological data fairness in health-related social media research

Sabina Leonelli[1], Rebecca Lovell[2], Benedict W Wheeler[2], Lora Fleming[2] and Hywel Williams[3]

DATA SCIENCE JOURNAL

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

Open Science for a Global Transformation

Leonelli, S 2017 Global Data Quality Assessment and the Situated Nature of "Best" Research Practices in Biology. *Data Science Journal*, 16: 32, pp. 1–11, DOI: https://doi.org/10.5334/dsj-2017-032

History and Technology
An International Journal

Historiographic Essay

## Mobilizing the Transnational History of Knowledge Flows. COVID-19 and the Politics of Research at the Borders

John Krige & Sabina Leonelli

Published online: 10 Mar 2021

HDSR

HOME ABOUT MASTHEAD CONTENT FEATURES

Just Accepted

## Data Science in Times of Pan(dem)ic

*by Sabina Leonelli*

Published on Nov 12, 2020

*Article*

SSS

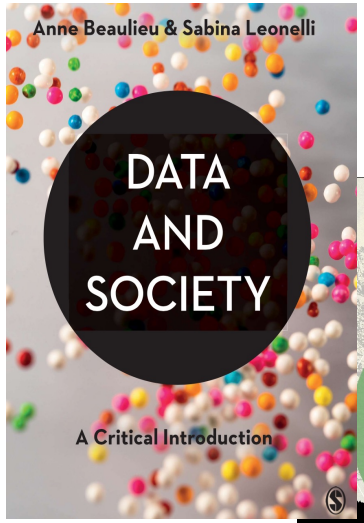
## Concealment and discovery: The role of information security in biomedical data re-use

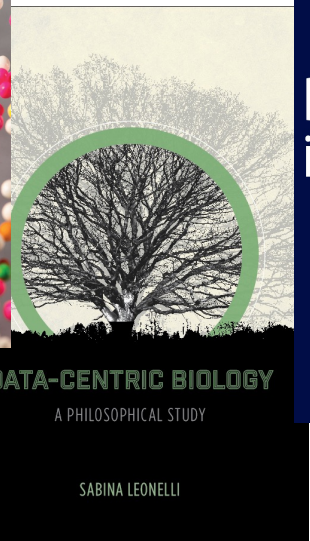
Social Studies of Science
2018, Vol. 48(5) 663–690

DOI: 10.1177/0306312718804875
journals.sagepub.com/home/sss
SAGE

Niccolò Tempini
Egenis, The Centre for the Study of Life Sciences, Department of Sociology, Philosophy and Anthropology, University of Exeter, Exeter, UK

Sabina Leonelli
Egenis, The Centre for the Study of Life Sciences, Department of Sociology, Philosophy and Anthropology, University of Exeter, Exeter, UK; School of Humanities, The University of Adelaide, Adelaide, SA, Australia





RETHINKING REPRODUCIBILITY AS A CRITERION FOR RESEARCH QUALITY

Sabina Leonelli


# Conclusions: Can Open Science Help With..

- Loss of research excellence and long-term reliability
- Increase of burden on (young) researchers
- Loss of access to publicly funded research outputs
- Disconnection between knowledge production and social role of research
- Disincentive to international and interdisciplinary collaboration
- Undermining of humanities and social sciences
- Stronghold of corporate infrastructures over research
- Increasing divide  between high-resourced and low-resourced environments
- Lack of transparency and credibility, public trust

# Conclusions: A Bad Scenario

- Loss of research excellence and long-term reliability
- Increase of burden on researchers
- Loss of access to publicly funded research outputs
- Disconnection between knowledge production and social role of research
- Disincentive to international and interdisciplinary collaboration
- Undermining of humanities and social sciences
- Stronghold of corporate infrastructures over research
- Increasing divide  between high-resourced and low-resourced environments (within and beyond research)
- Lack of transparency and credibility, public trust

- **Loss of creativity and increased bureaucracy**
- **OS demands piled on top of existing reward& evaluation system**
- **Loss of freedom to publish**
- **Continuing disconnection between knowledge production and social role of research**
- **Diversity of OS measure act as disincentive to international and interdisciplinary collaboration**
- **Even worse undermining of humanities and social sciences**
- **Further entrenchment of corporate grip**
- **Continuing increase of divide between high-resourced and low-resourced environments**
- **Lack of understanding, public trust; opinion vs evidence**

# Conclusions: A Good Scenario

- Loss of research excellence and long-term reliability
- Increase of burden on researchers
- Loss of access to publicly funded research outputs
- Disconnection between knowledge production and social role of research
- Disincentive to international and interdisciplinary collaboration
- Undermining of humanities and social sciences
- Stronghold of corporate infrastructures over research
- Increasing divide  between high-resourced and low-resourced environments (within and beyond research)
- Lack of transparency and credibility, public trust

- **Increased excellent and creativity**
- **Sustainable free access with no charge to authors**
- **Stronger links between knowledge production and social role of research**
- **Strong incentives to international and interdisciplinary collaboration**
- **Refocusing on humanities and social sciences as crucial to OS**
- **Regulation of corporate infrastructures, corsortia of public and private funders**
- **Fostering research in low-resourced environments (within and beyond research)**
- **Increased engagement and public trust**