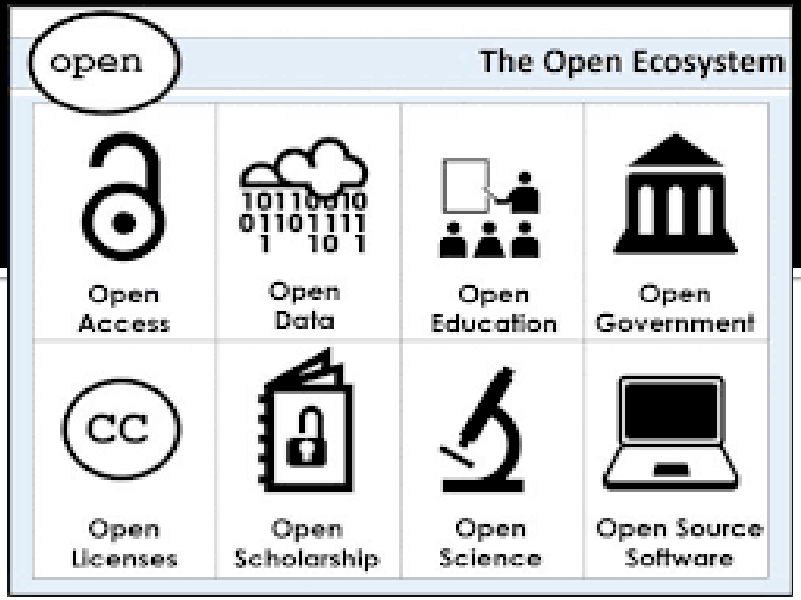
The Alan Turing Institute

Opening the Research Process: From Publications to Data, and Back Again

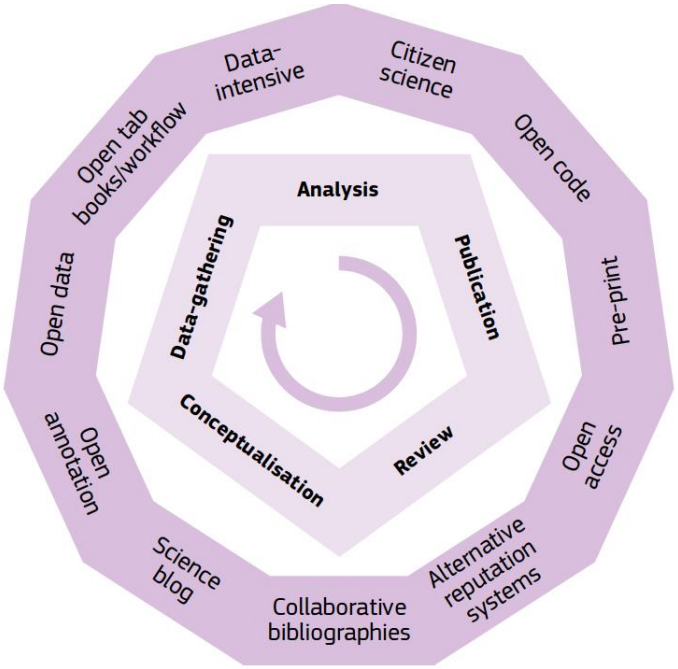
Sabina Leonelli
Exeter Centre for the Study of Life Sciences (Egenis)
& Department of Sociology, Philosophy and Anthropology
University of Exeter
@sabinaleonelli

Open Science



Variously defined by

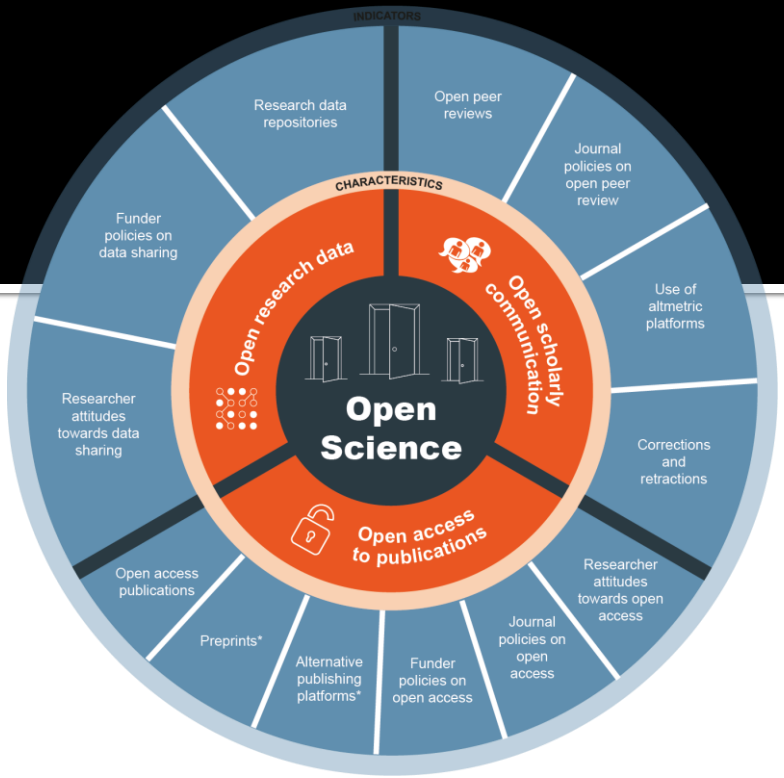
- the use of new digital tools
- practices of collaboration and sharing across disciplines and beyond academia
- a specific set of values
  e.g. openness, transparency & reproducibility
- a specific view of the research workflow (e.g. *when* to open) and related governance & communication



Open Science

"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)*



European Open Science Agenda:

1. OA publishing models
2. FAIR open data
3. Open Science Cloud
4. Altmetrics
5. Rewards & careers
6. Education & skills
7. Citizen Science
8. Research integrity

Open Science

Widespread agreement on three aspects:

- GLOBAL SCOPE: affects all stages of the research process, and its implementation involves a wide set of governance structures
- SYSTEMIC REACH: involves a systemic shift in current practices of research, publishing and evaluation
- LOCAL IMPLEMENTATION: implications for any one research systems need to be considered with reference to its specific characteristics, and thus the mechanisms through which OS is implemented are likely to vary → **key worry for researchers**

What's new?!
The long history of openness

Very long **history** of openness as norm for science:

- Not just a Mertonian norm (though this matters!), but a widespread practice
- Comes from research communities: natural history, meteorology, geology, astronomy
- More recently particle physics, genomics

High levels of **recognition** for this norm among researchers

What's new?!
The trouble with incentives

- Self-referential academic publishing:
  - Devaluation of quality and reproducibility of research outputs in favor of high volume and prestige
  - Dominance of publication in high impact factors journals over more desirable research goals
  - Publishing industry parasitic on publicly funded research
- Lack of incentives and rewards for:
  - responsible sharing of research components
  - collaboration and community building
  - public engagement and co-production
  - focus on social challenges

Result: trouble with science

- Loss of research excellence and long-term reliability
- Loss of access to research outputs
- Disconnection between knowledge production and social role of research
- Disincentive to international and interdisciplinary collaboration, as well as public engagement and communication across expertises
- Undermining of humanities and social sciences
- Increasing divide between high-resources and low-resourced environments (within and beyond research)
- Lack of transparency and credibility, public trust

Open Science as a "solution", and data management as the lynchpin

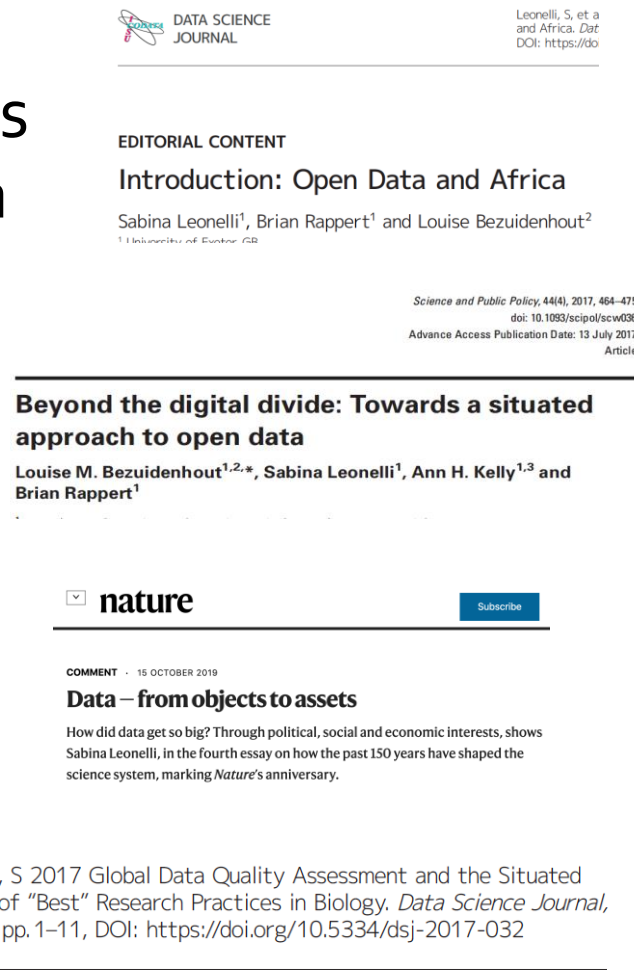
- Open Science as an opportunity for explicit discussion of research conditions and presumptions
- Data play a key role and exemplify key challenges:
  - new prominence as research outputs
    - recognised as valuable in their own right
    - mobility and re-use are central to data value
    - relation to articles (and related credit) needs to be redefined
  - significant resources required for re-use

Open Science as a "solution", and data management as the lynchpin

- Responsible data management can 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?

Empirical research on meanings and practices of openness (2013-2017)

- Interviews and fieldwork on
  - Researchers perspectives on openness in the UK, Kenya, South Africa, Ghana
    - Including a range of seniority (PIs to technicians)
    - Focusing on biology, biochemistry and engineering
  - Perceived obstacles to openness and particularly Open Data
  - Perceived obstacles to taking advantage of existing data infrastructures & Open Software


DATA SCIENCE JOURNAL

Leonelli, S, et a
and Africa. *Dat*
DOI: https://do

EDITORIAL CONTENT

Introduction: Open Data and Africa

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

Science and Public Policy, 44(4), 2017, 464–47
doi: 10.1093/scipol/scw03
Advance Access Publication Date: 13 July 201
Articl

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]

nature

Subscribe

COMMENT · 15 OCTOBER 2019

Data – from objects to assets

How did data get so big? Through political, social and economic interests, shows Sabina Leonelli, in the fourth essay on how the past 150 years have shaped the science system, marking *Nature*'s anniversary.

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

DATA_SCIENCE (2014-19): The Epistemology of Data-Intensive Science

Tracking **data journeys**

To understand how data move from sites of *production* to sites of *dissemination* and *interpretation/use*, and with which consequences

- **Approach:** philosophy, history and social studies of science
- **Focus:**
  1. **Databases** as windows on material/conceptual/institutional labor required to make data widely accessible and useable
     - labels & software to classify, model, visualize, retrieve data
     - management of infrastructure and communications
  2. **Data re-use cases** to investigate
     - conditions under which data can be interpreted
     - implications for discovery & what counts as good research
     - role of Open Science movement in knowledge generation

European Commission

RESEARCH & INNOVATION

Research and Innovation Observatory – Horizon 2020 Policy Support Facility

Jan 2017 – Jan 2018

European Commission > Research & Innovation > RIO - H2020 PSF > Policy Support Facility

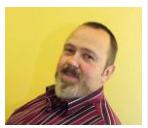
Home | Country analysis | Policy Support Facility | Library | Statistics | About | Search

MLE on Open Science


Frank Miedema
Chair


Katja Mayer
Rapporteur and expert


Sabina Leonelli
Expert


Kim Holmberg
Expert

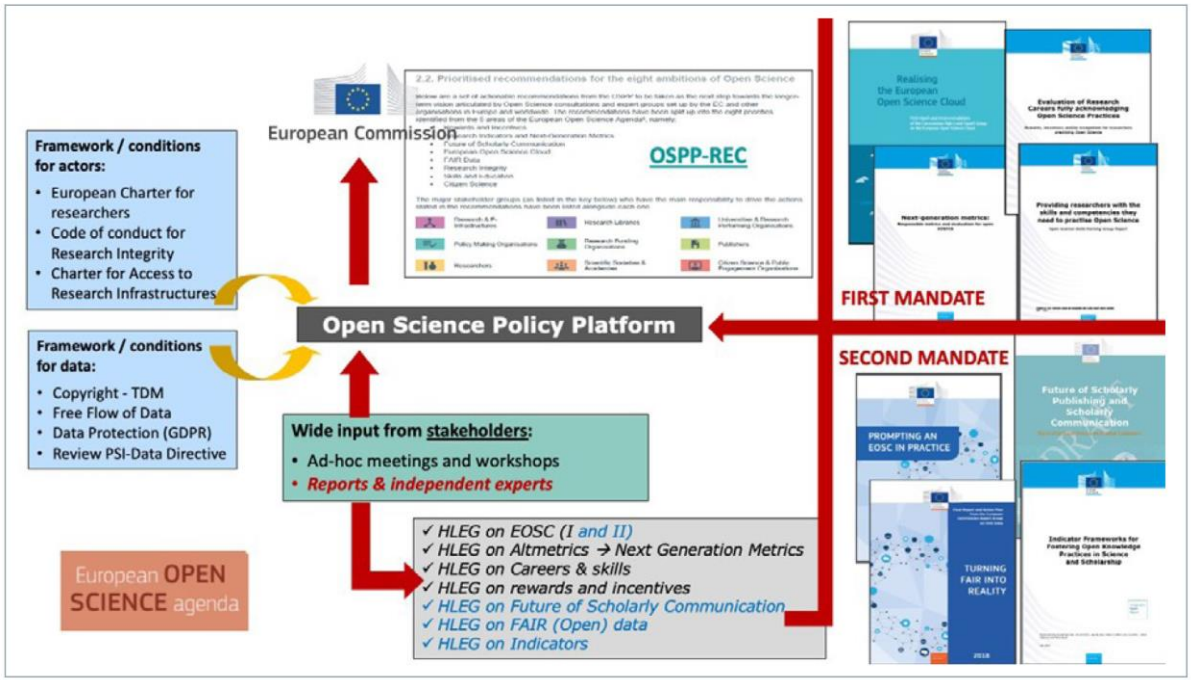
PSF Knowledge Centre: https://ec.europa.eu/h2020-policy-support-facility

Three topics:

1. The potential of altmetrics – alternative (i.e. non-traditional) metrics that go beyond citations of articles – 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

Armenia
Austria
Belgium
Bulgaria
Croatia
France
Latvia
Lithuania
Moldova
Portugal
Slovenia
Sweden
Switzerland

Open Science Policy Platform (2016-20)



European Commission

EUROPEAN OPEN SCIENCE CLOUD

Progress on Open Science: Towards a Shared Research Knowledge System

Final Report of the Open Science Policy Platform

Key challenges for Open Data

1. Enhancing skills and training
2. Distributing costs and accountabilities
3. Adapting intellectual property regimes
4. **Confronting semantic ambiguity**
5. **Recognising and promoting diversity in research cultures**
6. **Countering high resource bias**
7. **Integrating ethical and social concerns**

MLE findings: OS incentives and rewards

Career & Research Evaluation	Fairer assessment of research efforts (that take the complexity of scientific activities into account)
OS Training & Education Resources	Better training and support for research dissemination and data curation
Citation, Authorship & Publication system	Shifts in publishing and citation cultures, and recognition (and rewarding) of other activities, such as peer review
Long-Term Sustainability	Reliable Open Science infrastructures, that guarantee long term support and sustainability
OS Role Models	Visible recognition of Open Science activities
RRI & Public Engagement	Promotion of responsible research and innovation and improved public engagement (citizen science)
Transparency & Accountability	Transparency of research assessments, through for instance Open Peer Review.
International Coordination & Science Diplomacy	Enhanced international and diplomatic relations

Confronting semantic ambiguity

- **What openness means in practice**
  - Some common interpretations: "free of license", "free of ownership", "under CC-BY license", "common good", "good enough to share", "unrestricted access or use", "accessible without payment" (Grubb & Easterbrook 2011; Levin, Leonelli et al 2016)
- **How can it be implemented**
- **What is legal** (how does openness apply to sensitive research?)
- **What is ethical** (how to protect individuals & groups from harm?)
- **What is recommended by whom** (funders, learned societies, publishers, research institutions, governments..)

Recognising and promoting diversity in research cultures

- Enormous variation in methods, outputs and criteria for assessing excellence and quality
- Disciplines are not the only unit: large diversity in methods and habits within each field
- Significance of target objects: much research is exquisitely tailored to the characteristics of the phenomena of interest

Recognising and promoting diversity in research cultures

Should not throw the baby out with the bathwater: value of long-standing research traditions and reviewing methods

Crucial to remain user-friendly and fulfil expectations of users

- reliance on overly rigid standards creates exclusions and obliterates system-specific knowledge

Need case-by-case judgments on research quality and fruitful modes of data sharing

- data linkage methods are best when it is possible to disaggregate

Complexity of tools, skills and judgements required to implement OS

Data management and best practice for plant science

Sabina Leonelli, Robert P. Davey, Elizabeth Arnaud, Geraint Parry and Ruth Bastow

Table 1 | General tools for data management.

Type of tool	Function	Examples of relev
Open lab books	Digital and shareable version of traditional lab books	RSpace (http://wv
Generic open data repositories	General storage for many different data types	Figshare (http://w DataVerse (http://www.dataverse.org)
Specific databases	Fine-grained datasets that require subject-specific metadata	The Arabidopsis Information Resource (http://www.arabidopsis.org) The Bio-Analytic Resource for Plant Biology (http://www.bar.utoronto.ca) iHub (http://www.ionomicshub.org/home/PiiMS)
Data portals	Aggregating and providing visibility for various databases and resources	Araport (http://www.araport.org) Biosharing (http://www.biosharing.org) Agroportal[22]
Bio-ontologies	Keywords for the annotation, ordering and retrieval of data	Plant Ontology[15] Crop Ontology[21]
Metadata standards	Standardization of experimental data collection	Minimal Information on Biological and Biomedical Investigations (http://www.biosharing.org/standards) Minimal Information about a Microarray Experiment[27] Minimal Information about Plant Phenotyping Experiments (http://www.cropnet.pl/phenotypes/?page_id=15)
Identifiers for research materials	Annotation and retrieval of research materials on which experiments were originally performed	Germplasm Resource Information Network – Global (http://www.grin-global.org/) Multi-Crop Passport Descriptors (http://www.bioversityinternational.org/e-library/publications/detail/faobioversity-multi-crop-passport-descriptors-v2-mcpd-v2) Genesys (http://www.genesys-pgr.org)
Informatics standards	Software tools helping to format, store and visualize data	Breeding API (http://www.docs.brapi.apiary.io/) InterMINE (http://www.intermine.org)
Data annotation pipelines	Annotation of data from generation to reuse	Integrated Breeding Platform (http://www.integratedbreeding.net/) CropStore (http://www.cropstoredb.org/description.php) eDal (http://www.edal.ipk-gatersleben.de)
Guidelines of good practice	Articulation of data management principles and actions fostering data reuse	FAIR Data (http://www.force11.org/group/fairgroup/fairprinciples) Wheat Data Interoperability Guidelines[31]

(source: Leonelli et al 2017, *Nature Plants*)

Recognising and valuing (the epistemic reasons for) diversity is key to trust

- OS needs to foster trust among researchers, which in turns requires mechanisms to guarantee reliability of outputs
- Quality criteria for data (as well as knowledge claims!) are community-specific
- Variation of criteria needs to be studied and integrated into data infrastructures and mechanisms for responsible sharing

Countering high resource bias

- OS as playground for powerful research groups
- Many OD initiatives are led by rich, English-speaking labs within visible and popular research traditions..
- ..deal only with 'tractable' data formats and utilize resource-intensive methods and instrumentation

Integrating ethical and social concerns into OS tools and practices

- Role of ethics in Open Data:
  - Privacy of individuals and groups (GDPR)
  - Equity and ownership concerns
  - Socially damaging implications of data sharing
- Aim of data governance: human flourishing (Royal Society/British Academy 2017)
  - Who decides what counts as 'common' or 'public' good, and how?
  - What role do social goals play in research assessment?
  - Data fairness as legal, social and methodological

From data to publications: lessons learnt on OS knowledge production

How 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
- Disincentives to international and interdisciplinary collaboration
- Undermining of humanities and social sciences
- Increasing divide between high-resources and low-resourced environments
- Lack of transparency and credibility, public trust

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
- 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 measures act as disincentive to international and interdisciplinary collaboration**
- **Even worse undermining of humanities and social sciences**
- **Continuing to increase divide between high-resourced and low-resourced environments**
- **Lack of understanding, public trust; opinion vs evidence**

A Good Scenario

- Loss of research excellence and long-term reliability
- 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
- Increasing divide between high-resources and low-resourced environments (within and beyond research)
- Lack of transparency and credibility, public trust

- **Increased excellence 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**
- **Fostering research in low-resourced environments (within and beyond research)**
- **Increased engagement and public trust**

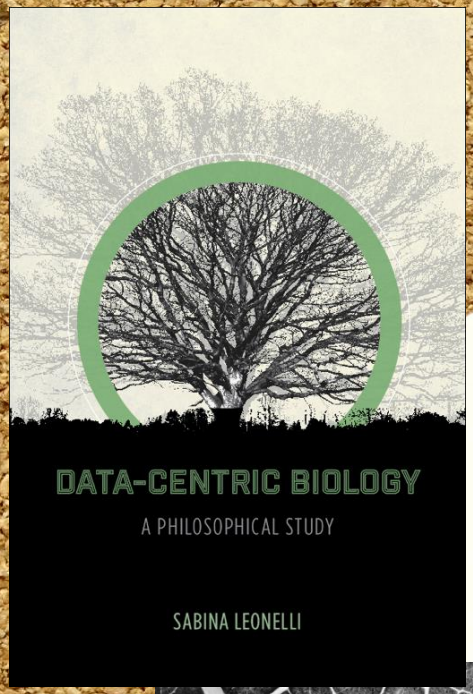
OS as a platform for critical, informed and open debate

- Acknowledge systemic nature of OS implementation
- Value researchers' diverse perspectives
  - Promoting dialogue on what counts as science, scientific infrastructures and scientific governance, and how results should be credited and disseminated
- Distribute burdens associated with transformation
  - Recognizing inequity of global research landscape and urgency of decreasing the digital and resource divide
  - Fostering research that documents such inequity and its implications (social science & humanities)

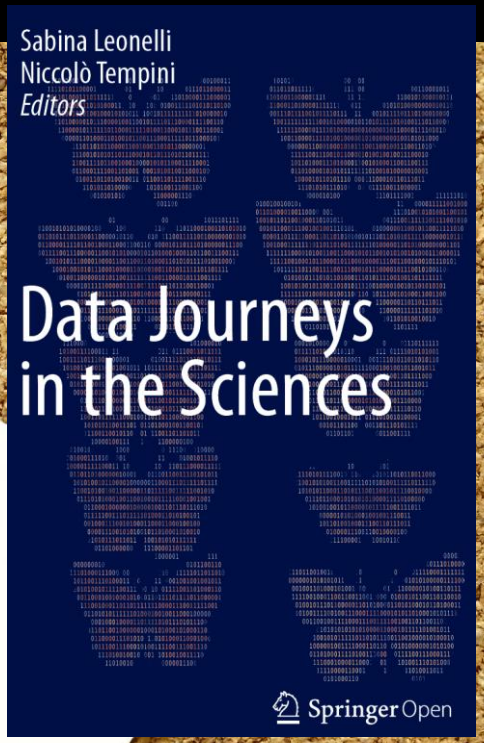
"Open Science for a Global Transformation": CODATA response to UNESCO Consultation

Open Science is best characterised as the necessary transformation of scientific practice to adapt to the changes, challenges and opportunities of the 21st century digital era to advance knowledge and to improve our world. This requires changes in scientific culture, methodologies, institutions and infrastructures. These changes are already present in many research domains and institutions, where their transformative effects can be witnessed, but they are unevenly distributed. One of the purposes of Open Science viewed as a call for transformation, is to ensure that 'no-one is left behind' https://doi.org/10.5281/zenodo.3935461

Coming soon:
Editorial and call for papers in the Data Science Journal



The Alan Turing Institute



Thank you!!

abstract

I reflect on the impact of Open Science on all stages of the research process, paying particular attention to the practical obstacles standing in the way of Open Science implementation within and across research domains. My discussion is grounded on (1) qualitative empirical studies on whether and how researchers from different disciplines and regions are practicing Open Science; (2) collaborations with infrastructures, repositories and institutions devoted to Open Science implementation; and (3) consultations with the European Commission and various national governments on how research evaluation should be conducted in order to incentivise responsible and sustainable forms of Open Science.

Some relevant publications

- Reports & Position Statements
  - (2018) Opportunities and Challenges for Implementing Plan S: The View of the Young Academies.
  - (2018) *OSPP-REC: Recommendations of the Open Science Policy Platform*.
  - (2018) *Open Science: Altmetrics and Rewards*. Final Report for the Mutual Learning Exercise Open Science: Altmetrics and Rewards of the European Commission
  - (2018) *Global Access to Research Software: The Forgotten Pillar of Open Science Implementation. A Global Young Academy Report*.
  - (2017) Position Statement on Open Science Publishing by the Open Science Policy Platform of the European Commission
  - (2016) Open Data Position Statement of the Global Young Academy and the European Young Science Academies
  - (2012) : Global Young Academy Position Statement on Open Science

Some relevant publications

- Studies of history and current perceptions of OS
  - Leonelli, S. (2018) Re-Thinking Reproducibility as a Criterion for Research Quality. *Research in the History of Economic Thought and Methodology*
  - Leonelli, S., Rappert, B and Bezuidenhout, L. (2018) Introduction: Open Data and Africa. *Data Science Journal*
  - Leonelli, S. (2017) Global Data Quality Assessment and the Situated Nature of "Best" Research Practices in Biology. *Data Science Journal*
  - Leonelli, S., Davey, R., Arnauld, E., Parry, G. and Bastow, R. (2017) Data Management and Best Practice in Plant Science. *Nature Plants*
  - Bezuidenhout, L., Leonelli, S., Kelly, A. and Rappert, B (2017) Beyond the Digital Divide: Towards a Situated Approach to Open Data. *Science and Public Policy*
  - Levin, N. and Leonelli, S. (2016) How Does One "Open" Science? Questions of Value in Biological Research. *Science, Technology and Human Values*
  - Levin, N., Leonelli, S., Weckowska, D., Castle, D., and Dupré, J. (2016) How Do Scientists Understand Openness? Exploring the Relationship between Open Science Policies and Research Practice. *Bulletin for Science and Technology Studies*
  - Leonelli, S., Spichtinger, D. and Prainsack, B. (2015) Sticks AND Carrots: Incentives for a Meaningful Implementation of Open Science Guidelines. *Geo*
  - Leonelli, S, Smirnoff, N., Moore, J., Cook, C. and Bastow, R. (2013) Making Open Data Work in Plant Science. *Journal for Experimental Botany*
  - Leonelli, S. (2013) Why the Current Insistence on Open Access to Scientific Data? Big Data, Knowledge Production and the Political Economy of Contemporary Biology. *Bulletin of Science, Technology and Society*