The Open Science Trend and the Role of University Libraries

WS on Next Generation Repositories

April 12, 2018

Miho Funamori
National Institute of Informatics
Outline

1. What is Open Science?
2. Policy Developments
3. Drivers
4. Challenges of Open Science
5. Role of University Libraries in Open Science Era
1. What is Open Science?
What is Open Science?

... Science has always been open!
Definition: Open Science

...Umbrella Term

Pre-print

Data-intensive

Open data

Open code

Open tab books/workflow

Citizen science

Open access

Collaborative bibliographies

Alternative reputation systems

Science blog

Open annotation

Definition: Open Science

- Said to have no fixed definition

- General understanding:
  - New ways of doing research and organizing science
  - Enabled through digital technology
  - Reshaping academic value systems
‘Science 2.0’ describes the on-going evolution in the modus operandi of doing research and organising science. These changes in the dynamics of science and research are enabled by digital technologies and driven by the globalisation of the scientific community, as well as the increasing societal demand to address the Grand Challenges of our times. They have an impact on the entire research cycle, from the inception of research to its publication, as well as on the way in which this cycle is organised.
Open science commonly refers to efforts to make the output of publicly funded research more widely accessible in digital format to the scientific community, the business sector, or society more generally. Open science is the encounter between the age-old tradition of openness in science and the tools of information and communications technologies (ICTs) that have reshaped the scientific enterprise and require a critical look from policy makers seeking to promote long-term research as well as innovation.
Open Science represents 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. The idea captures a systemic change to the way science and research have been carried out for the last fifty years: shifting from the standard practices of publishing research results in scientific publications towards sharing and using all available knowledge at an earlier stage in the research process.
Open science is about the way researchers work, collaborate, interact, share resources and disseminate results. A systemic change towards open science is driven by new technologies and data, the increasing demand in society to address the societal challenges of our times and the readiness of citizens to participate in research.
European Open Science Cloud (EOSC)

- EOSC aims to accelerate and support the current transition to more effective Open Science and Open Innovation in the Digital Single Market.

- **KEY FACTORS:**
  - New modes of scholarly communication
  - Modern reward and recognition practices need to support data sharing and re-use.
  - Core data experts need to be trained and their career perspective significantly improved.
  - A real stimulus of multi-disciplinary collaboration requires specific measures in terms of review, funding and infrastructure.
  - The transition from scientific insights towards innovation needs a dedicated support policy.

Source: European Commission: Realising the European Open Science Cloud
https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf#view=fit&pagemode=none
Open Science Monitor

- Includes open access to scientific results (publication and data)

- However, it is more than that...!
  - Researcher Attitude
  - Open Peer Review
  - Altmetrics
  - Correction and Retractions

Source: Open Science Monitor
https://ec.europa.eu/research/openscience/index.cfm?pg=home&section=monitor
2-1. Policy Developments: Open Access to Research Publications
How it started: “Serials Crisis”

- Journal subscription cost rising faster than the inflation speed
  - Four times higher in 2011 than 1986

*Includes electronic resources from 1999-2011.
We are writing the articles! Isn’t it unfair that the publishers are making profit, and many academics cannot even afford to read the articles?!

The journal subscription is too expensive!
Protest from Academia (2)

- “Subversive Proposal”
  - Steve Harnad (1994)
  - Called for scholarly articles to be freely available on the Internet, instead of published in print for the sake of royalties.

- “An Open Letter to Scientific Publishers”
  - 34,000 scholars worldwide (2001)
  - Called for the establishment of an online public library and pledging to refrain from publishing in traditional non-open-access journals.

- Provided definition of OA
- Two ways to achieve OA:

1. Self-Archiving (green OA)
   - Author’s final manuscript or the publisher’s version after a certain embargo period is archived on a website accessible worldwide.

2. Open-access Journals (gold OA)
   - Subscription fees are omitted instead of a fee charged to the author, usually called the article processing charge (APC).

Source: Budapest Open Access Initiative
http://www.budapestopenaccessinitiative.org/read
Move at Governmental-level

- Protest from a medical patient
  - "It is unfair that taxpayers do not have access to academic articles and thus cannot study their own medical condition, as the price of academic journals is exorbitant”.

- Funding agencies start making OA a mandate for scholarly articles funded publicly
  - NIH(US)-2008-”NIH Public Access Policy”
  - RCUK(UK)-2013-provides grant to universities for APC
2-2. Policy Developments: Open Access to Research Data
From Access to Research Publications to Access to Research Data

Publicly-funded Research

Research Data

Research Publications
Rationale for making Research Data publicly available

- **Accountability**
  - Publicly funded research should be transparent

- **Economic Efficiency**
  - Reuse of data leads to new findings without additional investments

- **Global Challenges Solving and Innovations**
  - Combining data from multiple discipline leads to solving global challenges
  - Industries using data leads to innovations
Governments:
- Australia, Austria, Belgium, Canada, China, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Russian Federation, the Slovak Republic, the Republic of South Africa, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States

Commitments:
- Work towards the establishment of access regimes for digital research data from public funding

Principles:
- Openness, Transparency, Legal conformity, Formal responsibility, Professionalism, Protection of intellectual property, Interoperability, Quality and security, Efficiency, Accountability
In NIH's view, all data should be considered for data sharing. Data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data. To facilitate data sharing, investigators submitting a research application requesting $500,000 or more of direct costs in any single year to NIH on or after October 1, 2003 are expected to include a plan for sharing final research data for research purposes, or state why data sharing is not possible.
Policy Developments in Research Data Sharing

- 2003, NIH, Data Sharing Policy
- 2004, OECD Declaration on Access to Research Data from Public Funding
- 2007, OECD Principles and Guidelines for Access to Research Data from Public Funding
- 2007, Biotechnology and Biological Sciences Research Council (BBSRC-UK), Data Sharing Policy
- 2011, Research Councils UK, Data Sharing Policy
- 2011, NSF, Data Sharing Policy
- 2013, OSTP-US, Increasing Access to the Results of Federally Funded Scientific Research
- 2014-20, Horizon 2020, Open Research Data Pilot
Data Journals and Supplemental Data

- **Data journals established (2014-)**
  - **Nature: Scientific Data**
    - *Scientific Data is an open-access, online-only journal for descriptions of scientifically valuable datasets.*
  - **Elsevier: Data in Brief**
    - *Data in Brief provides a way for researchers to easily share and reuse each other's datasets by publishing data articles.*

- **Supplemental Data**
  - *Supporting material that cannot be included, and which is not essential for inclusion, in the full text of the manuscript, but would nevertheless benefit the reader.*

Data Repositories

- General
  - DRYAD
  - Open Science Framework
  - figshare
  - re3data.org
  - Dataverse
  - Harvard Dataverse

- Disciplinary Data Repositories
  - Numerous
3-1. Drivers
Data-Intensive Scientific Discovery
The Fourth Paradigm: Data-Intensive Scientific Discovery

Tony Hey
Corporate Vice President
Microsoft External Research
A Digital Data Deluge in Research

- Data collection
  - Sensor networks, satellite surveys, high throughput laboratory instruments, observation devices, supercomputers, LHC ...

- Data processing, analysis, visualization
  - Legacy codes, workflows, data mining, indexing, searching, graphics ...

- Archiving
  - Digital repositories, libraries, preservation, ...
Emergence of a Fourth Research Paradigm

1. Thousand years ago – **Experimental Science**
   - Description of natural phenomena

2. Last few hundred years – **Theoretical Science**
   - Newton’s Laws, Maxwell’s Equations...

3. Last few decades – **Computational Science**
   - Simulation of complex phenomena

4. Today – **Data-Intensive Science**
   - Scientists overwhelmed with data sets from many different sources
     - Data captured by instruments
     - Data generated by simulations
     - Data generated by sensor networks
   - eScience is the set of tools and technologies to support data federation and collaboration
     - For analysis and data mining
     - For data visualization and exploration
     - For scholarly communication and dissemination

Science must move from data to information to knowledge

With thanks to Jim Gray

Astronomy has been one of the first disciplines to embrace data-intensive science with the Virtual Observatory (VO), enabling highly efficient access to data and analysis tools at a centralized site. The image shows the Pleiades star cluster from the Digitized Sky Survey combined with an image of the moon, synthesized within the WorldWide Telescope service.
3-2. Drivers

Social Demand
Globalization and Collaboration

- More and more researchers working on international collaboration projects
  - Need for sharing and storing information
  - Need for online collaboration platform

Source: Open Science Framework
https://cos.io/our-products/open-science-framework/
Global Challenge and Innovations

- Global challenges and innovations requiring:
  - Combining data from various discipline
  - Multi-disciplinary collaboration

- Data sharing could:
  - Enable data combination from various discipline
  - Enable research data to be used for industrial purposes and problem solving
Michael Nielsen
Reinventing Discovery

- SPARC honors Michael Nielsen as innovator for bringing Open Science into the mainstream (2012)
- Reinventing Discovery tells the exciting story of an unprecedented new era of networked science.
- It demonstrated various cases with strong emphasis on citizen science.

https://www.amazon.co.jp/Reinventing-Discovery-New-Networked-Science-ebook/dp/B005OQQGZ5
Citizen Science...examples

- Galaxy Zoo
  - Crowdsourced astronomy project where people classify galaxies

- Foldit
  - Online puzzle video game about protein folding

- eBird
  - Online database of bird observations
The shrinking gap between society and the academia

- Tertiary education attainment rate is rising, especially for younger generation.
- Thus, citizens literacy and analytical skills are getting comparable to the academia.
- This results in stronger demand for accountability and societal problem-solving.
3-3. Drivers
Call for Research Transparency
Retraction Watch

The Retraction Watch Leaderboard

with 21 comments

Who has the most retractions? Here's our unofficial list (see notes on methodology), which we'll update as more information comes to light:

1. Yoshitaka Fujii (total retractions: 183) Sources: Final report of investigating committee, our reporting
2. Joachim Boldt (96) Sources: Editors in chief statement, additional coverage
3. Diederik Stapel (58) Source: Our cataloging
4. Adrian Maxim (48) Source: IEEE database
5. Peter Chen (Chen–Yuan Chen) (43) Source: SAGE, our cataloging
6. Hua Zhong (41) Source: Journal
7. Shigeaki Kato (39) Source: Our cataloging
8. James Hunton (37) Source: Our cataloging
10. Hyung-In Moon (35) Source: Our cataloging
11. Naoki Mori (32) Source: PubMed, our cataloging
12. Tao Liu (29) Source: Journal
13. Cheng-Wu Chen (28) Source: our cataloging
14. Gideon Goldstein (26)
15. Scott Reuben (25)
16. Gilson Khang (22) Sources: WebCitation.org, WebCitation.org, journal
17. Friedhelm Herrmann (21)
18. Noel Chia (21)

Source: The Retraction Watch Leaderboard
ClimateGate Scandal

Do E-Mails Reveal Scientist Claims On Climate Change are BUNK?

Hackers break into servers of a major British climate change research facility and purportedly uncover e-mails urging scientists to "hide the decline" of temperatures, manipulate data and silence skeptics.

https://thinkprogress.org/climategate-hacked-emails-reveal-global-warming-deniers-are-crazed-conspiracy-theorists-ea8df6eb792b3#ke1ie5d3v
Recent decades have seen an unprecedented explosion in the human capacity to acquire, store and manipulate data and information and to instantaneously communicate them globally, irrespective of location...

...Effective exploitation of Big Data depends fundamentally upon an international culture of 'Open Data' that involves sharing of data and their availability for re-use and re-purposing.
AREAS FOR ACTION

- Scientists need to be more open among themselves and with the public and media
- Greater recognition needs to be given to the value of data gathering, analysis and communication
- Common standards for sharing information are required to make it widely usable
- Publishing data in a reusable form to support findings must be mandatory
- More experts in managing and supporting the use of digital data are required
- New software tools need to be developed to analyse the growing amount of data being gathered

Source: Royal Society: Final report – Science as an open enterprise
The accord identifies the opportunities and challenges of the data revolution as today’s predominant issue for global science policy. It proposes fundamental principles that should be adopted in responding to them. It adds the distinctive voice of the scientific community to those of governments and inter-governmental bodies that have made the case for open data as a fundamental pre-requisite in maintaining the rigour of scientific inquiry and maximising public benefit from the data revolution in both developed and developing countries.
Open Research Data as Good Science Practice

- Science is by definition knowledge with evidence, which can be reproduced by others.
  - Scientists have to publish their idea openly.
  - Scientists must also show the evidence.

- In print age, only research articles could be published as evidence.

- In digital age, digital research data can also be provided as evidence.

Research data is even better than just articles for the establishment of science!
Academics in Action for Research Transparency

- Reproducibility Project: Psychology
  - Center for Open Science

- Berkeley Initiative for Transparency in the Social Sciences
  - Center for Effective Global Action (CEGA), UC Berkeley

- Peer Reviewers’ Openness Initiative
4. Challenges and Innovation in Academic Reward System
Challenges of Open Science

- Who is going to do the work?
  - Incentive for sharing data
  - Cost of sharing data
  - Training and carrier

- What for sensitive data?
  - Data with privacy concern or industrial confidentiality can be excluded

I want to write my article before sharing data!
Changing Academic Reward System

- Recognizing data
  - Data journals
  - Data citations

- Recognizing social impact
  - Altmetrics

Making data available increases citations

- Alter, Pienta, Lyle
  - 240%, social sciences *

- Piwowar, Vision
  - 9% (microarray data)†

- Henneken, Accomazzi
  - 20% (astronomy) #


Source: Kevin Ashley, Digital Curation Center, UK
Porto research data conference
Changing Scholarly Communication
...Peer Review System

☐ Open Peer Review
  - Reviewer’s comments are open to public with/without the name of reviewer
  - Enabling transparent peer review

☐ Post Publication Peer Review
  - Peer review done after publishing
  - Speeding up publishing, and allowing to count impact in peer review

☐ Cascading Peer Review
  - Peer review comments transferred to next submission
  - Reducing costs and improving efficiencies in peer review
5. The Role of University Libraries In Open Science
Open Science at Academic Institution Level

1. Hold accountability
   ✓ Meeting mandates by funding agencies
   ✓ Research transparency and reproducibility

2. Promoting research
   ✓ Providing proper research environment
   ✓ Accelerating research

3. Disseminating research of the institution

4. Linking academia and society
Tangible actions in Open Science at Academic Institutions

1. Hold accountability for research
   - Research office caring for mandates and transparency
   - Providing infrastructure for data storage

2. Disseminating research
   - Make research discoverable and reusable
   - Provide institutional repositories

3. Promoting research
   - Provide access to scholarly contents, promoting OA
   - Promoting new research paradigm, data-intensive science
   - Advocating researchers for new research paradigm
   - Evaluate research within institution

4. Linking academia and society
   - Matching the needs of society with the seeds of researchers
Tangible actions in Open Science at Academic Institutions

1. Hold accountability for research
   - Research office caring for mandates and transparency
   - Providing infrastructure for data storage

2. Disseminating research
   - Make research discoverable and reusable
   - Provide institutional repositories

3. Promoting research
   - Provide access to scholarly contents, promoting OA
   - Promoting new research paradigm, data-intensive science
   - Advocating researchers for new research paradigm
   - Evaluate research within institution

4. Linking academia and society
   - Matching the needs of society with the seeds of researchers
Role of University Libraries in promoting Open Science

1. Providing scholarly contents within institution
   ✓ Acquiring and locating scholarly contents (books, e-journals, other)
   ✓ Promoting open access

2. Stewarding scholarship within institution
   ✓ Provide storage for active data and long-term preservation
   ✓ Provide DMP tool

3. Disseminating scholarship of institution
   ✓ Provide institutional repository for publishing
   ✓ Adding proper metadata and curating data

4. Advocating for good scholarship practices, i.e. Open Science
   ✓ Working on Knowledge Graph (linking publication, data, researcher, grant, etc)
   ✓ Provide RDM Training
How we classify our tools and services

Data Management Support

Data Management Planning

Active Data Infrastructure

Data Stewardship

Before research  During research  After research
Purr
Purdue University Research Repository

FACT:

Many funding agencies require data management plans with grant proposals.

purr.purdue.edu

2,542 grant proposals
Providing Training for Research Data Management

MANTRA is a free online course for those who manage digital data as part of their research project.

About this course: This course will provide learners with an introduction to research data management and sharing. After completing this course, learners will understand the diversity of data and their management needs across the research data lifecycle, be able to identify the components of good data management plans, and be familiar with best practices for working with data including the organization.

http://datalib.edina.ac.uk/mantra/
https://www.coursera.org/learn/data-management
https://lms.gacco.org/courses/course-v1:gacco+ga088+2017_11/about
University Libraries in Open Science Era

- Expanding the scope of scholarly contents
  - Books, journals, gray-literature, *research data*

- Develop new services for new contents
  - DMP tools, RDM training
  - Data curation, facilitating reuse of scholarship
  - Building knowledge graph

- Advocating of good scholarship practices in the Open Science era!
  - Promoting open access and RDM
  - Disseminating and promoting reuse of scholarship
  - Caring for research transparency and reproducibility
  - Advocating for new research paradigm

*As the steward of scholarship of the institution, be the change agent for new research paradigm!*