Institutional Role in Supporting Open Science

Asia Open Access, Dhaka -2019

March 7, 2019

Miho Funamori
National Institute of Informatics
Outline

1. What is Open Science?
2. Drivers and Responses
3. 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.
The push and resisting force towards Open Science

The degree of Open Science is determined by the balance of two forces.

Way to do Research?

Open Science?

- Traditional Scholarship
- Researcher Career
- HR Skills
- Intl. Competitiveness
- Data Protection

- Transformation of Scholarly Communication System
- Excess in Research Output
- Research Reproducibility
- Inter-disciplinary Research
- Acceleration of Research
- Data Deluge
- Digital Technology
- Accountability
- Avoid duplicate investment
- Solving Social Issues
- Citizen Science
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. Drivers and Responses
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, ...

This work is licensed under a Creative Commons Attribution 3.0 United States License.
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/
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/researchopenscience/pdf/realising_the_european_open_science_cloud_2016.pdf#view=fit&pagemode=none
Open access papers ‘gain more traffic and citations’

Open access science articles are read and cited more often than articles available only to subscribers, a study has suggested.

July 30, 2014

The Research Information Network analysed the web traffic to more than 700 articles published in a hybrid science journal Nature Communications in the first six months of 2013.

It found that, after 180 days, articles whose authors had paid for them to be made open access had been viewed more than twice as often as those articles accessible only to the journal’s subscribers.

A further analysis of more than 2,000 papers published in Nature Communications between April 2010 and June 2013 revealed that open access articles were cited a median of 11 times, compared with a median of seven citations for subscription-only articles. The paper concludes that open access papers enjoy a “small” citation advantage in all disciplines except chemistry.
Alliance of Science Organisations in Germany...Research Data WG

- Principles for the Handling of Research Data (2010.6)
- Research Data at Your Fingertips—A Position Paper (2015.2)

Source: https://www.allianzinitiative.de/en/archive/research-data/working-group-documents/
3-2. Drivers and Responses

Call for Research Transparency and Research Reproducibility
Retraction Watch

The Retraction Watch Leaderboard

with 21 comments

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

1. Yoshitaka Fujii (total rejections: 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)
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
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.

Source: CODATA: Message form President Geoffrey Boulton
http://www.codata.org/message-from-president-geoffrey-boulton
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-ea8dfeb792b3#ke1ie5d3v
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.

Source: Science International: Open Data in a Big Data World
http://www.icsu.org/science-international/accord
Reproducibility Project: Psychology

**Have you failed to reproduce an experiment?**

Most scientists have experienced failure to reproduce results.

<table>
<thead>
<tr>
<th>Field</th>
<th>Someone else's</th>
<th>My own</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemistry</td>
<td>80%</td>
<td>50%</td>
</tr>
<tr>
<td>Biology</td>
<td>80%</td>
<td>50%</td>
</tr>
<tr>
<td>Physics and engineering</td>
<td>70%</td>
<td>50%</td>
</tr>
<tr>
<td>Medicine</td>
<td>60%</td>
<td>50%</td>
</tr>
<tr>
<td>Earth and environment</td>
<td>50%</td>
<td>40%</td>
</tr>
<tr>
<td>Other</td>
<td>30%</td>
<td>20%</td>
</tr>
</tbody>
</table>

Source: Nature, “Over half of psychology studies fail reproducibility test” (2015.9.27)
https://www.nature.com/news/over-half-of-psychology-studies-fail-reproducibility-test-1.18248
In reality, the major driving force for RDM is scientific misconduct prevention.

- **MEXT: “Guideline for Dealing with Scientific Misconduct” (2014)**
  「研究活動における不正行為への対応等に関するガイドライン」
  - Strengthening the guideline in 2006.
  - Holds institutions to be responsible for research transparency and preventing scientific misconduct.

  日本学術会議「(回答)科学研究における健全性の向上について」
  - “Ten-Years Preservation Rule for Research Data”
    研究データ10年保存ルール
3-3. Drivers and Responses
Social Demand
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.
From Access to Research Publications to Access to Research Data

Publicly-funded Research

Research Data

Research Publications
The push which triggered US government to adopt OA policy

Science didn't understand my kids' rare disease until I decided to study it
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.

Source: NIH Data Sharing Policy and Implementation Guidance
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
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
Data Management Plan (DMP)

- **DMP** – A plan where researcher state the treatment of data used during research process.

- More and more, research funders are requiring to submit DMP along with research proposal.

- In Japan, JST is asking for DMP since FY2018.

Librarians assist researchers in writing DMPs.
3. Institutional Role in Supporting 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
Data Repositories

- General

- Disciplinary Data Repositories
  - Numerous
How we classify our tools and services

Data Management Support

Data Management Planning
Active Data Infrastructure
Data Stewardship

Before research  During research  After research
Providing Training for Research Data Management

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
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
RDM implementation in an academic institute

---

**Acad. Inf. Discovery Service**

- Running Inst. Repositories
- Adding metadata
- DMP support
- RDM Training etc.

**Univ. Library**

- Research Strength Info.
- Drafting of research strategy
- Research Evaluation
- Finding research collaborators

**研究所 LIBRARY**

- Running Inst. Repositories
- Adding metadata
- DMP support
- RDM Training etc.

**INST. REPOS.**

- Open/Closed/Embargo
- Publishing func.
- Storing func.
- Preservation func.

**STORAGE**

- Closed
- Access control func.
- Version control func.

**ICT center**

- Provision of e-infrastructures

**Research Strategy**

- Data generation
- Data store
- Data access

---

**Lab**

- Grad students, technicians, lab manager, etc.

---

**Decision-making on RDM implementation**

---

- President
- Research Administration Office
- Researcher
- Research Strategy
- Research Strength
- R Data Info
- DMP Info
- R Integrity (10 yrs preserv)
- Grant Mgmt (DMP)

---

- Research Evaluation
- Finding research collaborators
- Drafting of research strategy
- Research Strength Info.
Multi-stakeholder Approach needed to implement RDM at universities.

Grad students, technicians, lab manager, etc.: data generation, RDM

Univ-wide policies & strategies
- Research VP
- Library
- CIO
- R Integrity
- R admin
- R integrity
- R evaluation
- R support
- Univ. Library
- D preservation
- D publishing
- ICT Center
- E-infrastructure
- IT policies

Dept. admin offices: coordination

Research Admin Office
- R admin
- R integrity

Research Support Unit (URA Station)
- R evaluation
- R support

Univ. Library
- D preservation
- D publishing

ICT Center
- E-infrastructure
- IT policies

President

Professional Assoc.
- Data Protection
- Data Curation
- Learnt Societies

I want to make the university research competitive!
Why an RDM Charter?

Participants at AXIES-RDM session started to claim,

“*We need a charter in order to convince the university administration and to get the researchers and staff engaged!***”

AXIES

- Academic eXchange for Information Environment and Strategy
- Community of CIOs and ICT centers of universities in Japan.
- Counterpart to EDUCAUSE in the US
“RDM Charter for Academic Institutions”

- RDM Charter
  - Not for researchers, but
  - For academic institutions!

- Purpose of RDM Charter
  - Give university administration ideas and options to implement RDM in respective institutions.

Researcher: Don’t dare to tell me how to manage my data! I know what I’m doing!
“RDM Charter for Academic Institutions”

...Composition

- **The Charter** (3 pages)
  - Addresses the viewpoints why academic institutions needs to take RDM seriously.
  - Viewpoints in bullet points:
    1. Role of academic institutions in RDM
    2. Policies and organizations for RDM needed in acad. Institutions
    3. RDM procedures in acad. Institutions
    4. RDM Purpose options in acad. Institutions
    5. Digital platform functions needed for RDM in acad. Institutions
    6. Human resources development for RDM in acad. Institutions
    7. Reuse and service options of research data in acad. institutions

- **Appendix** (12 pages)
- **Glossary** (12 terms, 3 pages)
- **References** (2 pages)
RDM implementation in an academic institute

*Univ. Library*
- Running Inst. Repositories
- Adding metadata
- DMP support
- RDM Training etc.

*INST. REPOS.*
- Open/Closed/Embargo
- Publishing func.
- Storing func.
- Preservation func.

*STORAGE*
- Closed
- Access control func.
- Version control func.

*ICT center*
- Provision of e-infrastructures

*Research Strategy*
- Drafting of research strategy
- Research Evaluation
- Finding research collaborators

*Research Administration Office*
- Grad students, technicians, lab manager, etc.

*Lab*
- Drafting of research strategy
- Research Evaluation
- Finding research collaborators

*President*
- Decision-making on RDM implementation

*Researcher*
- R Integrity (10 yrs preserv)
- Grant Mgmt (DMP)
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
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!
Register at FORCE11!

The Future of Research Communications and e-Scholarship

- Registration is free!
- Newsletter will be sent.
- Travel fellowship will be available for participants from developing countries for below two events!

FORCE11 events:

- **FSCI**
  - August 5-9 @ UCLA
  - FORCE11 Scholarly Communication Institute—a 5-day intensive program!

- **FORCE2019**
  - October 15-17 @ Edinburgh
  - Annual FORCE11 conference