International framework of scientific data sharing and open science

Yasuhiro Murayama

Members of Cabinet Office Expert Panel of Open Science Promotion

Research Executive Director,
Big Data Integration Research Center,
National Institute of Information & Communications Technology

WDS International Programme Office Hosted by NICT
Based in Tokyo, Japan
My background

Geophysics, Atmospheric/Space Science

Board member of Japan Geoscience Union (2014-2015)
Visiting Professor, Kyoto University, (2013-2014)
Associate member, Science Council of Japan
Research Executive Director, Big Data Integration Res. Ctr.
Natl. Inst. Information & Communications Technology
ICSU-World Data System Scientific Committee, ex officio membe
Member of Cabinet Office Expert Panel of Open Science
Observer Member of EC’s High Level Expert Group of European
Open Science Cloud

Data Management, Science Policy
Today’s Contents

• Introduction
• Data Sharing/Open Research Data
  – Method of Modern Science and Communications
• International Policy Situation
  – EU, OECD, Japan
• Toward Our Best Practice
  – Data publication, data citation
  – Digital socio-technological data/information infrastructure
• Concluding Remarks
Introduction
G8 Science Ministers Statement London UK, 12

Introduction

We, the G8 Science Ministers met in London on Wednesday of our respective national science academies, as part of this unique meeting we discussed how our nations could increase transparency, coherence and coordination of the global science in order to address global challenges and maximise the scope of research.

3. Open Scientific Research Data

Open enquiry is at the heart of scientific endeavour, and rapid technological change has profound implications for the way that science is both conducted and its results communicated. It can provide society with the necessary information to solve global challenges. We are committed to openness in scientific research data to speed up the progress of scientific discovery, create innovation, ensure that the results of

4. Expanding Access to Scientific Research Results

“Open Government Data”
Open Access to Open Data and Open Science
Overview example

Open Access

Open Source

Self Archiving

Institutional Repository

Science Commons

Full OA (mega) journal

Science 2.0+

Data journal

Open Research Data

Data Sharing

Database, Repository

Research Outputs

ReUse

Research Activity

Open Innovation

Citizen Science

Creative Commons

Open Data (Open gov.)

Code for X

2000’s

2010’s

Improve, Incremental

Redesign, Disruptive?

[6.

[K. Hayashi, 2015]
Scientific Practice and Data Sharing/Open Research Data: Changing Scholarly Communications
An article is not sufficient for validating results.
Reproducibility issue
Research integrity issue

Data as 1st-class research output
Social information asset, provided to the general society

Essential for “irreproducible” natural phenomena
Global change, space, living organisms, health . . .
Approx. 1,300 scientists worked for the IPCC WG1. 
(3,000-4,000 scientists for all WG1-3?)
A crisis of replicability?

Replicability of 53 papers on “good” IF journals are examined;
A paper is cited by a number of secondary works, regardless of its replicability.
Open science: the (unrealized) potential

- ‘Big data’ and ICTs open up new scientific opportunities
- Enable collaboration across disciplines
- Increase efficiency, transparency and reproducibility
- Address global challenges more effectively
- Increase knowledge spill-overs for science, innovation and society
- Promote citizen engagement in science
International Policy Situation
European Open Science Agenda

5 broad policy action lines (from public consultation, validated by stakeholders incl. EU Member States):

- Fostering and creating incentives for open science
- Removing barriers to openness
- Mainstreaming
- Developing an ecosystem
- Embedding open science to societal and economic benefits

Open Science – Big players

- Elsevier
- Springer + Nature
- Google
- Wikimedeia
Defining Open science

Open science includes:
• Open access to scientific publications
• Open and “intelligent” access to research data (and materials)
• Open access to digital applications and source code
• Open access for scientists, the public and commercial companies

➢ re-asserting science as a global public good

“A new scientific paradigm”

• Science is becoming increasingly data-driven
Promoting Open Science in Japan Opening up a new era for the advancement of science

[H. Manago, 2015]
Cabinet Office “Expert Panel of Open Science” (Dec, ‘14 --- March ‘15)
http://www8.cao.go.jp/cstp/sonota/openscience/
⇒ Final Report was published at the Web site 30 March 2015.
Reference:
Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 Version 1.0 11 December 2013 p 4
How to promote “Open Science” in Japan

• The “national principle” is not obligation or mandatory rules, but “guiding principle”.
• The decision by Cabinet Office is now being followed by stakeholders’ discussions (related ministries, scientific societies, universities/national institutes)
• Every scholars do not accept. Depends on their disciplines and past practice/culture.
• New funding mechanism is also required
  – to encourage researchers, journal editors, publishers, data producers, data infrastructure managers/developers.
G7 2016 Science & Technology Ministers’ Meeting
(15-17 May 2016, Tsukuba, Ibaragi, Japan)

MINISTERS’ MEETING AGENDA:
1. Global Health - Health Care and Science and Technology
2. Gender and Human Resource Development for STI
3. The Future of the Seas and Oceans
4. Clean Energy - Developing Innovative Energy Technology
5. Inclusive Innovation - Mainstreaming Inclusiveness Among Innovation Policies
6. Open Science - Entering into a New Era for Science

Agreed to establish a new G7 Open Science Working Group
Toward Our Best Practice
Print & Electronic Technologies as Social Info. Infrastructures

--- 百年の印刷文化の基礎支えと 成長途中的ディジタル・サイエンス

Public library (paper media)

Printing press/Gutenberg: 1445

First scientific journal: 1665

Intl. Assoc. Academies: 1899

ICSU established: 1931

ENIAC, von Neumann: 1946

World Data Center system: 1957

Hard Disk Drive: 1956

TCP/IP, dial-up (64kbps): 1982

WWW (CERN): 1991

New global data initiatives: ICSU-WDS, RDA etc. 2008〜2013
“Data Publication” and “Data Citation”

- Data Publications
  cf. journal publication: review, fix (print), publish with DOI..., metrics (citation index etc.)

- Data Citation
  - ID of dataset (“DOI” is OK?), citation standards? metrics?...

- More outputs from scientists to Society

[Society of Geomagnetism, Earth, Planetary and Space Sciences, 2013]
Building a Culture of Data Citation

CREATE
1. Dataset is stored in a publicly accessible repository
2. Researcher uses ANDS services to mint a Digital Object Identifier (DOI) for the dataset
3. Australian researcher creates a research dataset and a publication related to the dataset

DOI
4. DOI is used in data citation

USE
5. Research community use the DOI to access the dataset and carry out related research
6. Citation metrics services (e.g., Scopus, Web of Knowledge) accumulate citation references to the dataset and publication
7. Funding and research groups review publication and dataset citation metrics

REWARD
8. Researcher future funding and promotion influenced by dataset citation metrics

MEASURE
9. Research community generate new publications using the DOI to reference the dataset

Illustration by Australian National Data Services (ANDS)
Steps by Major (Geophysical) Journals encouraging data deposition & citation

• Willey/AGU publication policy:
  “…in AGU’s journals, all data necessary to understand, evaluate, replicate, and build upon the reported research must be made available and accessible whenever possible…”

• SpringerOpen/”Earth, Planets and Space”, “Geoscience Letters”…
  “…Electronic archiving of data enables readers to replicate, verify and build upon the conclusions published in papers in the journal. It is recommended that all data which are not directly attached to a publication as electronic supplementary files be deposited…”

• Elsevier/JASTP:
  “…Elsevier encourages authors to deposit raw experimental data sets underpinning their research publication in data repositories, and to enable interlinking of articles and data…”
How datasets are cited by articles

(a) Data collection community
- Physical properties of Hole ##
- Reports of the Deep Sea Drilling Project

(b) Data sharing community
- National Social Science Survey
- Working
- Attitudes

(c) Data sharing community
- Pangaea
  - (http://www.pangaea.de/)
  - 384,815 citations from OAI-PMH

(d) Data sharing community
- Australian Data Archive (ADA)
  - (http://www.ada.edu.au/)
  - 16,062 citations from HTML

(d) Referential context
- Inter-university Consortium for Political and Social Research (ICPSR)
  - (http://www.icpsr.umich.edu/) (c) Downloaded 05/31/14
  - 115,154 citations from OAI-PMH

How are datasets cited by articles?

1. Data collection community
   - Physical properties of Hole ##
   - Reports of the Deep Sea Drilling Project

2. Data sharing community
   - National Social Science Survey
   - Working
   - Attitudes

3. Data sharing community
   - Pangaea
     - (http://www.pangaea.de/)
     - 384,815 citations from OAI-PMH

4. Data sharing community
   - Australian Data Archive (ADA)
     - (http://www.ada.edu.au/)
     - 16,062 citations from HTML

5. Referential context
   - Inter-university Consortium for Political and Social Research (ICPSR)
     - (http://www.icpsr.umich.edu/) (c) Downloaded 05/31/14
     - 115,154 citations from OAI-PMH
Concluding Remarks

• Open Science is an emerging focus of international Science & Technology policy
• A new mode of scholarly communication is required.
• Scholarly work eco-cycle
  – Create, Use, Measure, Reward/Recognition
  – Librarians, editors, publishers, data managers...
• Journal editors can play an active role for not only publishing articles, but also datasets behind them.

“Science is built of facts the way a house is built of bricks”  (Henri Poincare, 1902)
...And then, RDA Plenary 9
Date: 5-7 April 2017
Place: Barcelona, Spain