Open Science and the Academy:
A Theoretical Discussion

Open Science, Open Innovation, and Institutional Aspect (DSIR-SS2)
6th International Congress on Advanced Applied Informatics

July 12, 2017

Miho Funamori
National Institute of Informatics

Research Question

- How would academics adopt the idea of Open Science?
  - Given the fact that academics are largely ignorant on the issue.
- This paper presents a theoretical discussion on this question by scrutinizing the background and evolution of the policy agenda, the issues arising from it, and the academia's perception of and behavior toward Open Science.

Outline

1. What is Open Science?
2. Policy Developments
3. Drivers
4. Actors
5. Challenges of Open Science
6. Discussion:
   - Possibility of Universities adopting the idea of Open Science

1. What is Open Science?

What is Open Science?

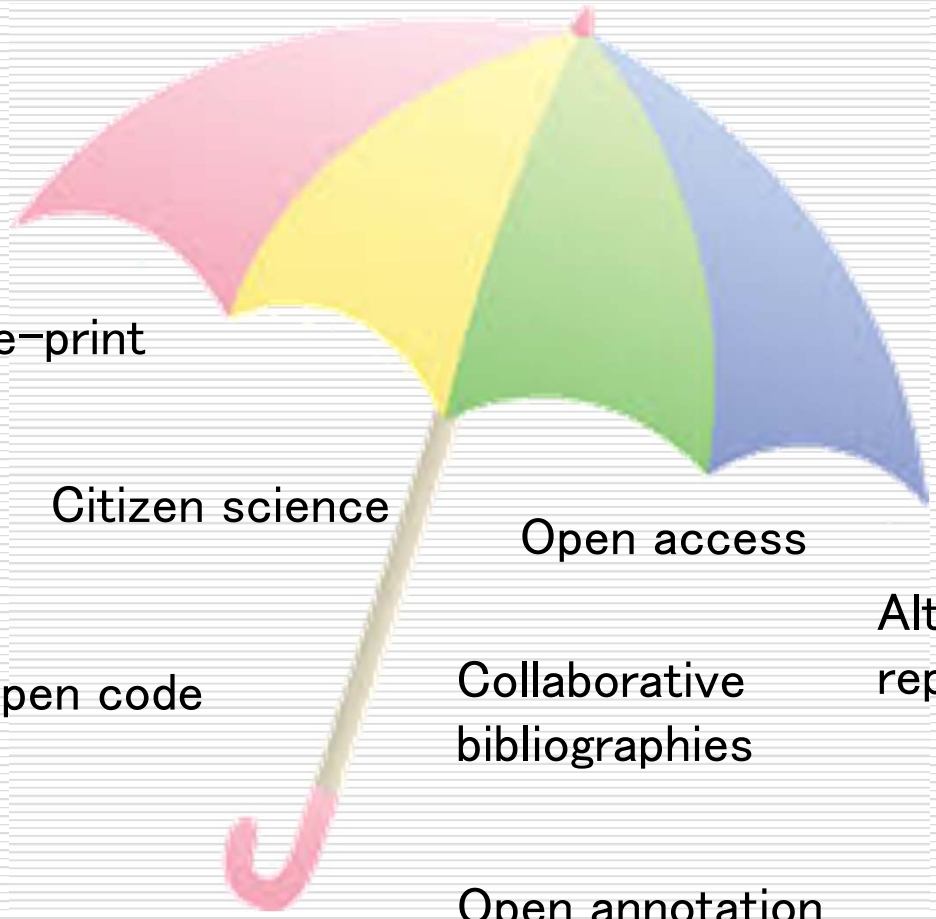
... Science has always been open!

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

Definition: Open Science …Umbrella Term



Pre-print

Data-intensive

Citizen science

Open access

Open data

Alternative reputation systems

Open code

Collaborative bibliographies

Open tab books /workflow

Science blog

Open annotation

Source: European Commission, "Open Innovation, Open Science, Open to the World—a vision for Europe," 2016

Open Science...European Commission (2014)

...Public Consultation 'Science 2.0': Science in Transition

EUROPEAN COMMISSION

DIRECTORATES-GENERAL FOR RESEARCH AND INNOVATION (RTD) AND COMMUNICATIONS NETWORKS, CONTENT AND TECHNOLOGY (CONNECT)

BACKGROUND DOCUMENT

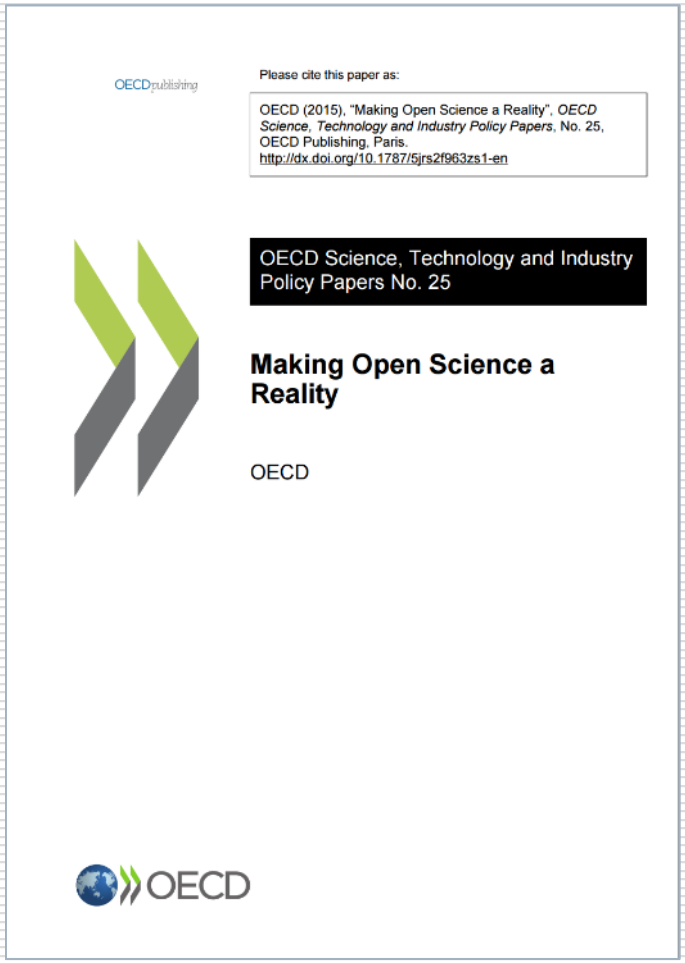
PUBLIC CONSULTATION

'SCIENCE 2.0': SCIENCE IN TRANSITION

'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...OECD (2015)

...Making Open Science a Reality



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...European Commission (2016)

...Open Innovation, Open Science, Open to the World

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...European Commission (2016)

...Amsterdam Call for Action on Open Science

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)

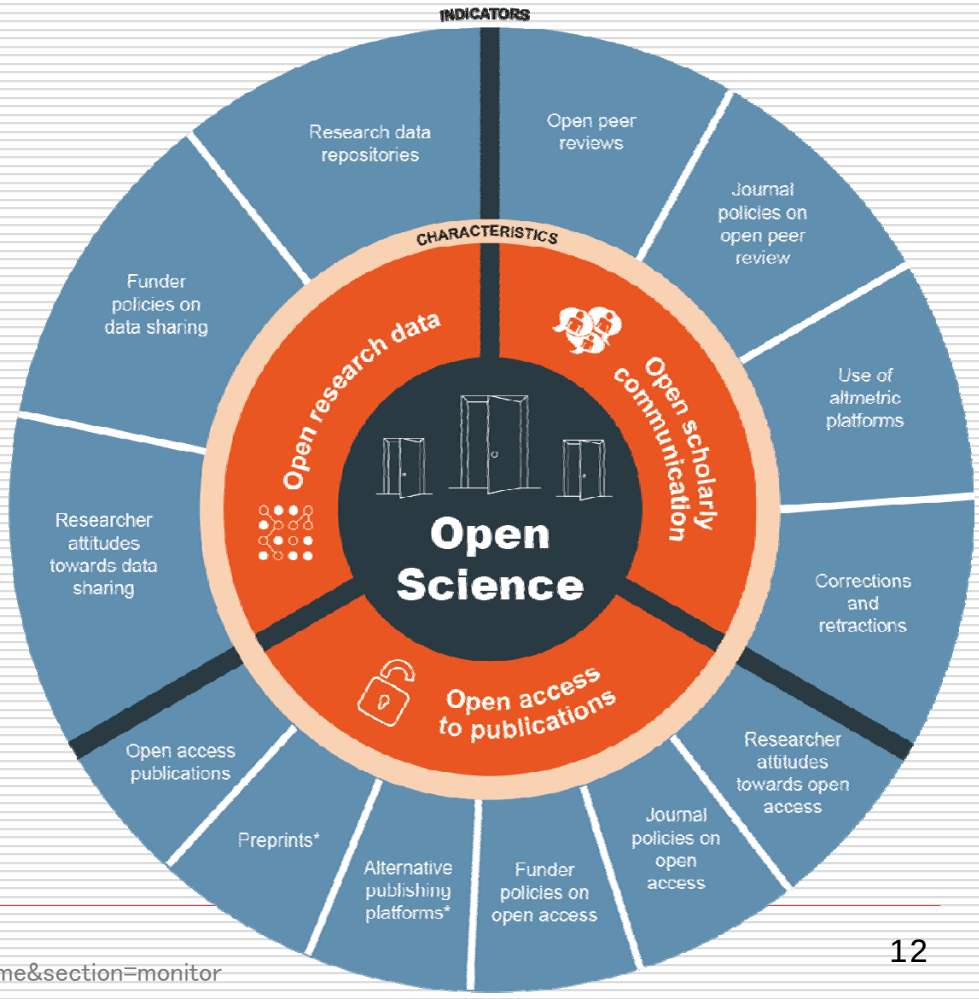
- EOSC aims to accelerate and support the current transition to more effective Open Science and Open Innovation in the Digital Single Market.
- KEY FACTORS:
  - New modes of scholarly communication
  - Modern reward and recognition practices need to support data sharing and re-use.
  - Core data experts 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

Open Science Monitor



Source: Open Science Monitor
https://ec.europa.eu/research/openscience/index.cfm?pg=home&section=monitor

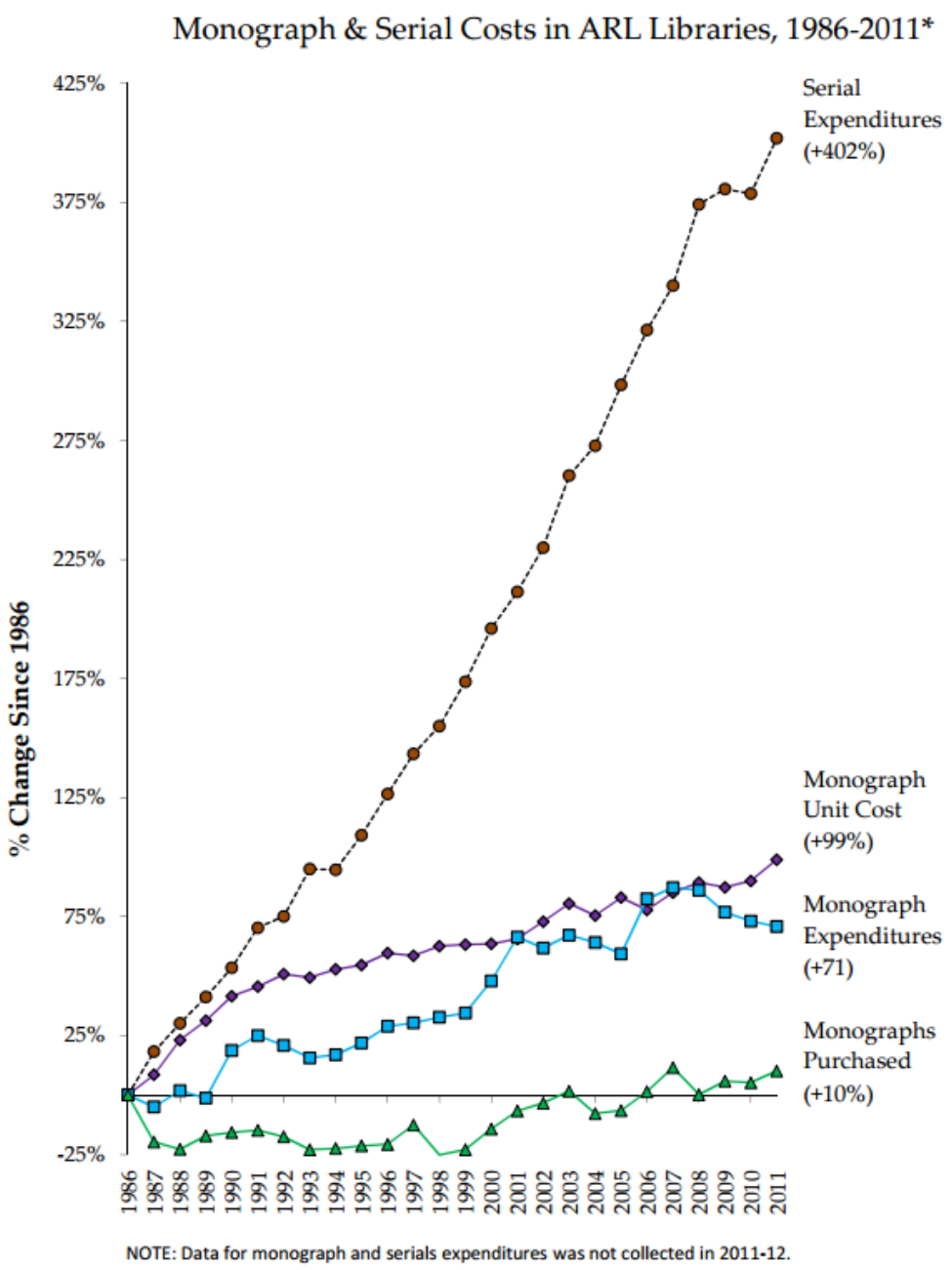
Do researchers have to change the way to perform science?

- Research has always been conducted by researchers, and will be.
- Why not let the researchers decide how to conduct research?

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



Source: ARL Statistics 2010-11 Association of Research Libraries, Washington, D.C.
*Includes electronic resources from 1999-2011.
http://www.arl.org/storage/documents/monograph-serial-costs.pdf

Protest from Academia (1)

Isn't it unfair
that the publishers are making profit,
and many academics cannot even afford
to read the articles?!

We are
writing 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.

Protest from Academia (3)

- "Budapest Open Access Initiative (BOAI)", (2002)
  - 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

Definition:
"Open Access (OA)"

- To provide free and unrestricted online access to peer-reviewed scholarly research.

- It may also include:
  - theses, book chapters, monographs

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

Move at University-level

☐ Provide APC to their own researchers

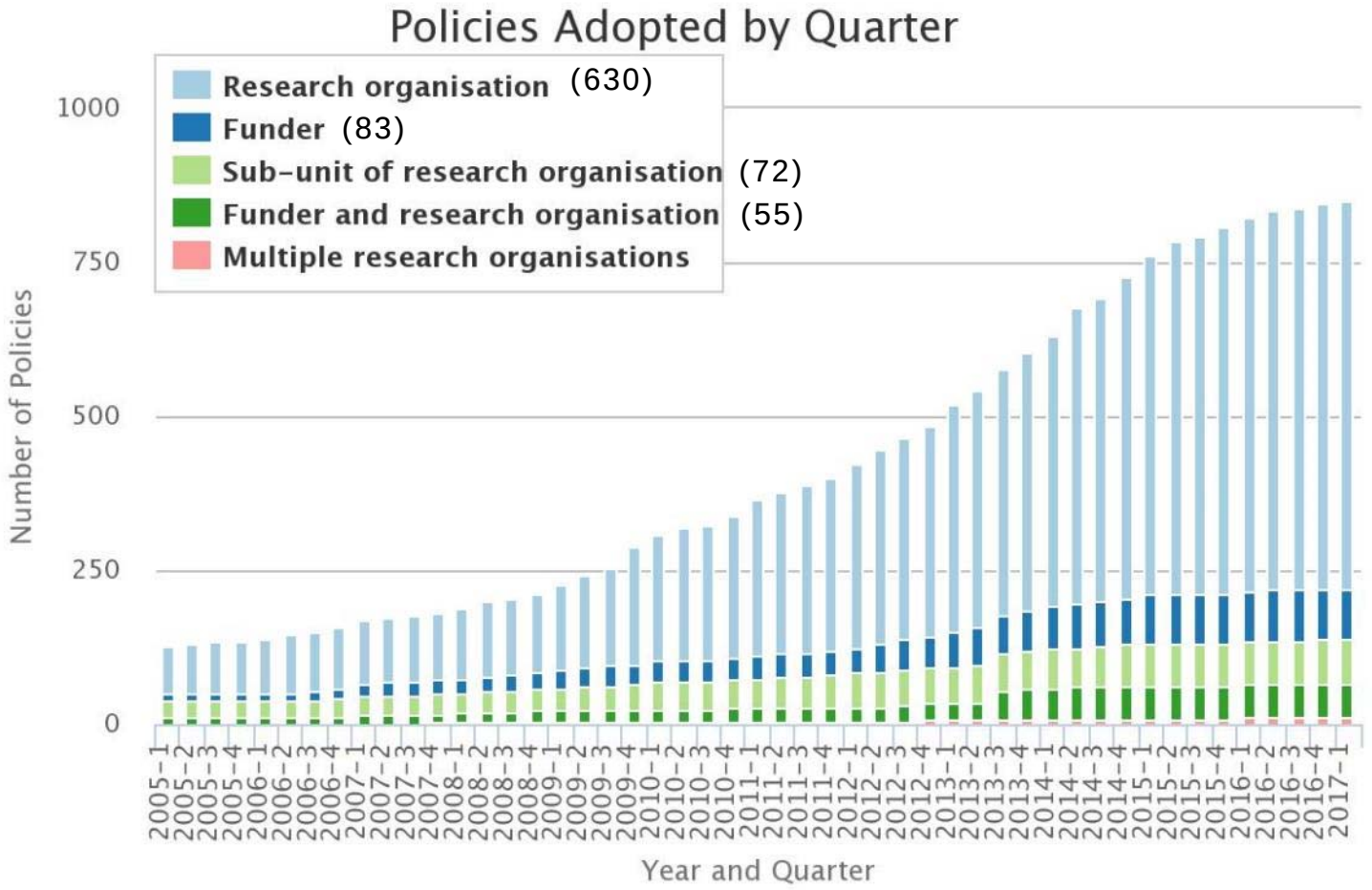
- Compact for Open-Access Publishing Equity (COPE)

☐ Adopt OA mandate policy

- ✓ Each faculty member grants to the institution nonexclusive permission to make available his or her articles and to exercise the copyright in those articles for the purpose of open dissemination.
- ✓ The institution make the scholarly article available to the public in an open-access repository.

Source: Compact for Open-Access Publishing Equity (http://www.oacompact.org/)
Model Open Access Policy (https://osc.hul.harvard.edu/modelpolicy)

Number of OA policies adopted



840 institutions

Source: Registry of Open Access Repository Mandates and Policies (ROARMAP)
http://roarmap.eprints.org/

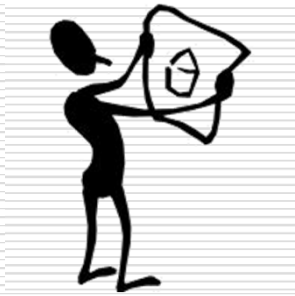
2-2. Policy Developments: Open Access to Research Data

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

Declaration on Access to Research Data from Public Funding

30 January 2004 - C(2004)31/REV1

- 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

Data Sharing Policy, NIH (2003-)

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.

Data Management Plan (DMP) required!

Source: NIH Data Sharing Policy and Implementation Guidance
https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

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.*

http://www.nature.com/sdata/ https://www.journals.elsevier.com/data-in-brief

Data Repositories

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

- General

- Disciplinary Data Repositories
  - Numerous

When did the lead change from academics to other actors?

- At times of open access movement for publications, academics were taking the lead.
- The lead of open access of research data is taken by the funding agencies, and other actors.

3-1. Drivers

Data-Intensive Scientific Discovery

The Fourth Paradigm: Data-Intensive Scientific Discovery

Tony Hey

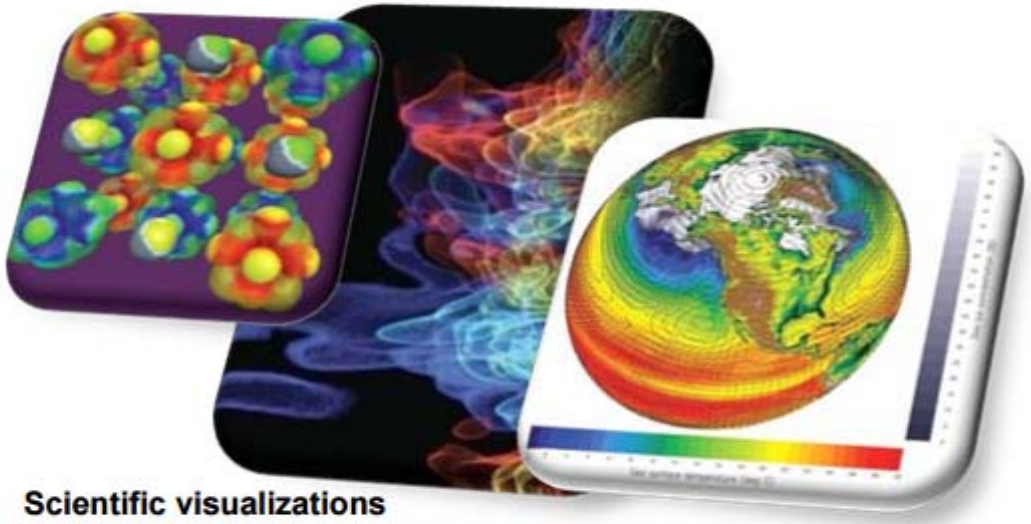
Corporate Vice President

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, …

SensorMap
Functionality: Map navigation
Data: sensor-generated temperature, video camera feed, traffic feeds, etc.



Scientific visualizations
NSF Cyberinfrastructure report, March 2007

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…

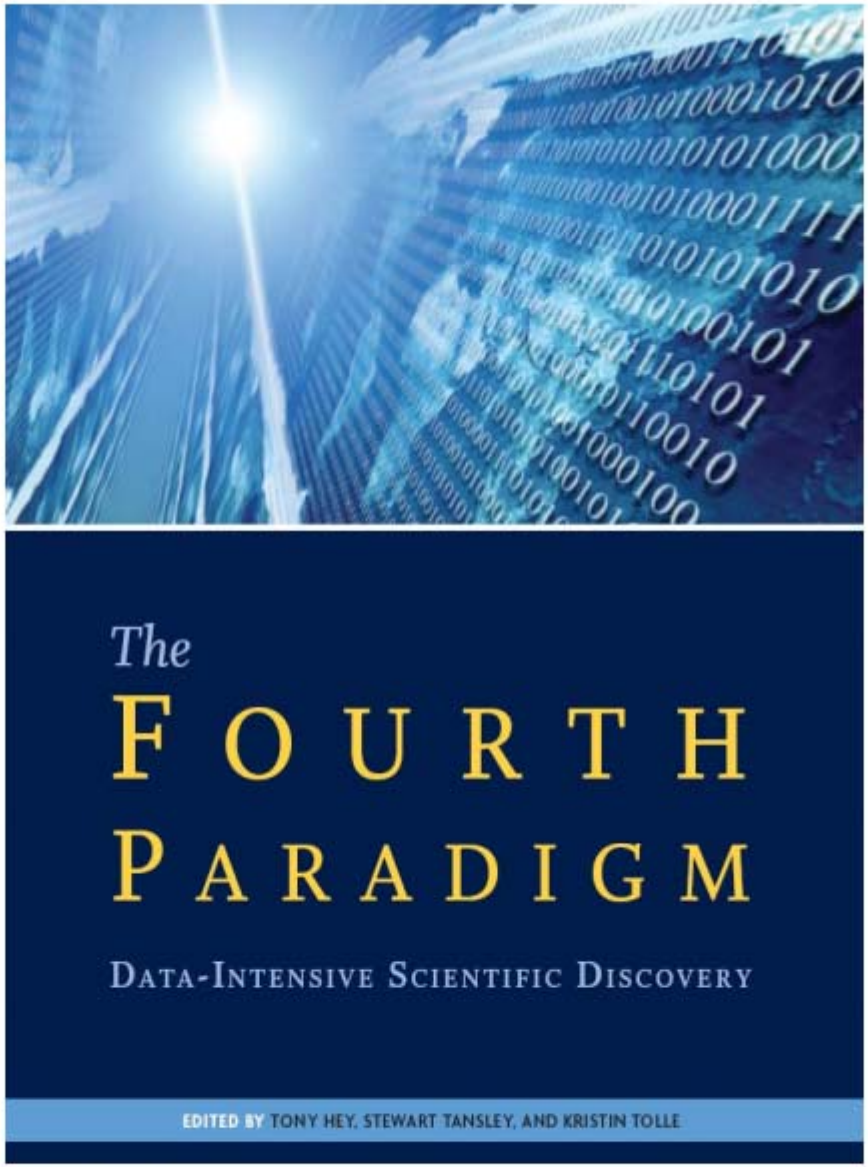
3. Last few decades – **Computational Science**
   - Simulation of complex phenomena

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

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 form the Digitized Sky Survey combined with an image of the moon, synthesized within the WorldWide Telescope service.

Science must move from data to information to knowledge

With thanks to Jim Gray

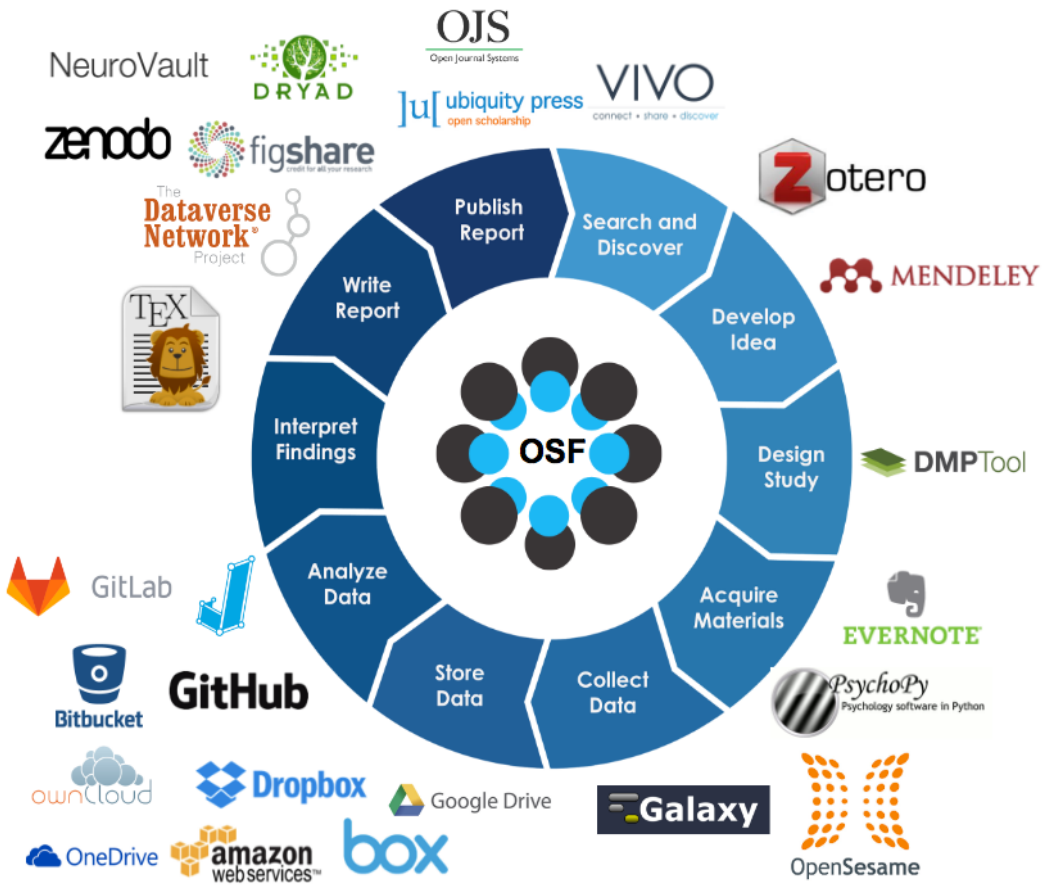


3-2. Drivers

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

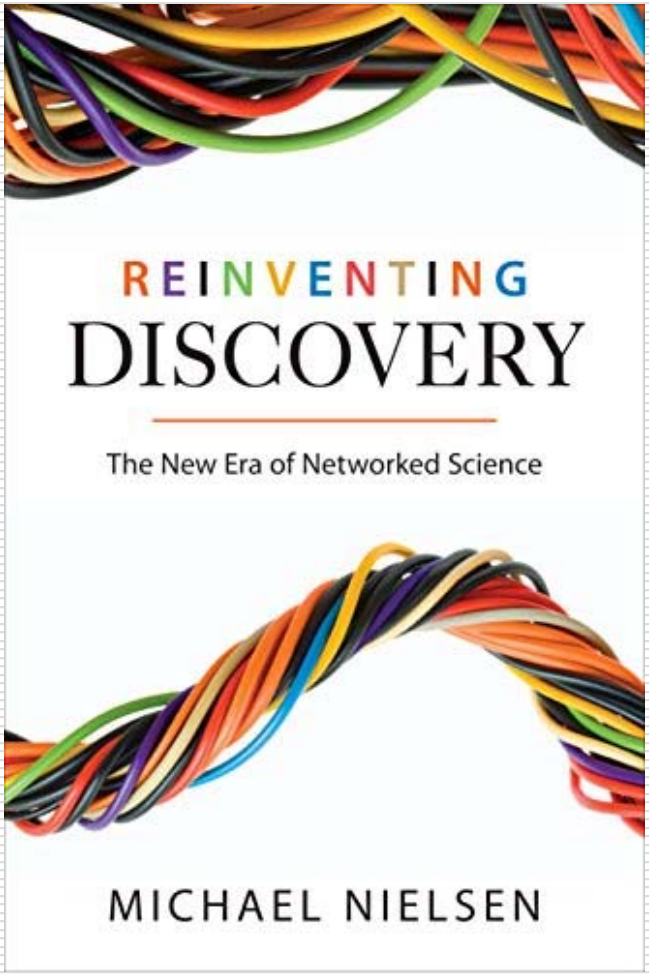
- Global challenges and innovations requiring:
  - Combining data from various discipline
  - Multi-disciplinary collaboration
- Data sharing could:
  - Enable data combination from various discipline
  - Enable research data to be used for industrial purposes and problem solving

Accountability and Efficiency

- Publicly-funded research should be held accountable
  - Open access to scientific results
- The investment return to publicly-funded research should be maximized
  - Reuse of research data by other researcher and by industries
  - Reuse of research data in multidisciplinary projects

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/B005OQGZ54
http://sparc.arl.org/news/sparc-honors-michael-nielsen-innovator-bringing-open-science-mainstream

Citizen Science...examples

- Galaxy Zoo 
  - Crowdsourced astronomy project where people classify galaxies
- Foldit 
  - Online puzzle video game about protein folding
- eBird eBird
  - Online database of bird observations

3-3. Drivers

Call for Research Transparency

Retraction Watch

Tracking retractions as a window into the scientific process

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
9. Hendrik Schön (36) Sources: PubMed and Thomson Scientific
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)

Pages

Source: The Retraction Watch Leaderboard
http://retractionwatch.com/the-retraction-watch-leaderboard/ (Accessed 2017/03/19)

ClimateGate Scandal

This is the worst scientific scandal of our generation

Telegraph

http://blog.heartland.org/2011/11/heartland-institute-mentioned-in-climategate-emails-rounds-1-and-2-part-1/
https://wattsupwiththat.com/2009/11/28/telegraphs-booker-on-the-climategate-scandal/
https://thinkprogress.org/climategate-hacked-emails-reveal-global-warming-deniers-are-crazed-conspiracy-theorists-ea8dfeb792b3#.ke1ie5d3v

HOME | CODATA BLOG | EVENTS | NEWS

CODATA

International Council for Science : Committee on Data for Science and

ABOUT | EVENTS | MEMBERSHIP | INITIATIVES | TASK GROUPS | WORKING GROUPS

Message from President Geoffrey Boulton

- 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.

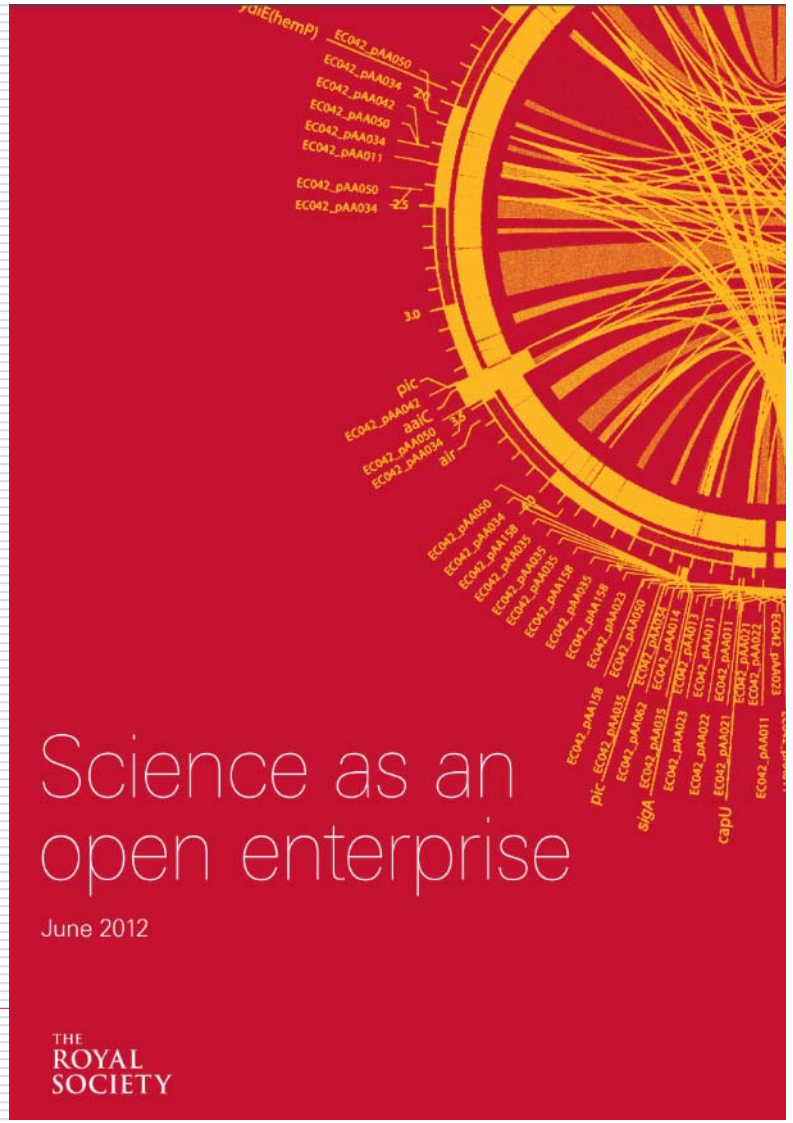
Source: CODATA: Message form President Geoffrey Boulton
http://www.codata.org/message-from-president-geoffrey-boulton

The Royal Society: Science as an open enterprise (2012)

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
https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/



Science International: Open Data in a Big Data World

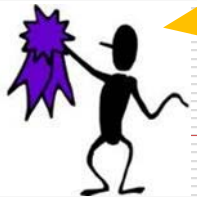
- The accord identifies the opportunities and challenges of the data revolution as today's predominant issue for global science policy. It proposes fundamental principles that should be adopted in responding to them. It adds the distinctive voice of the scientific community to those of governments and 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

ICSU iap ISSC twas

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

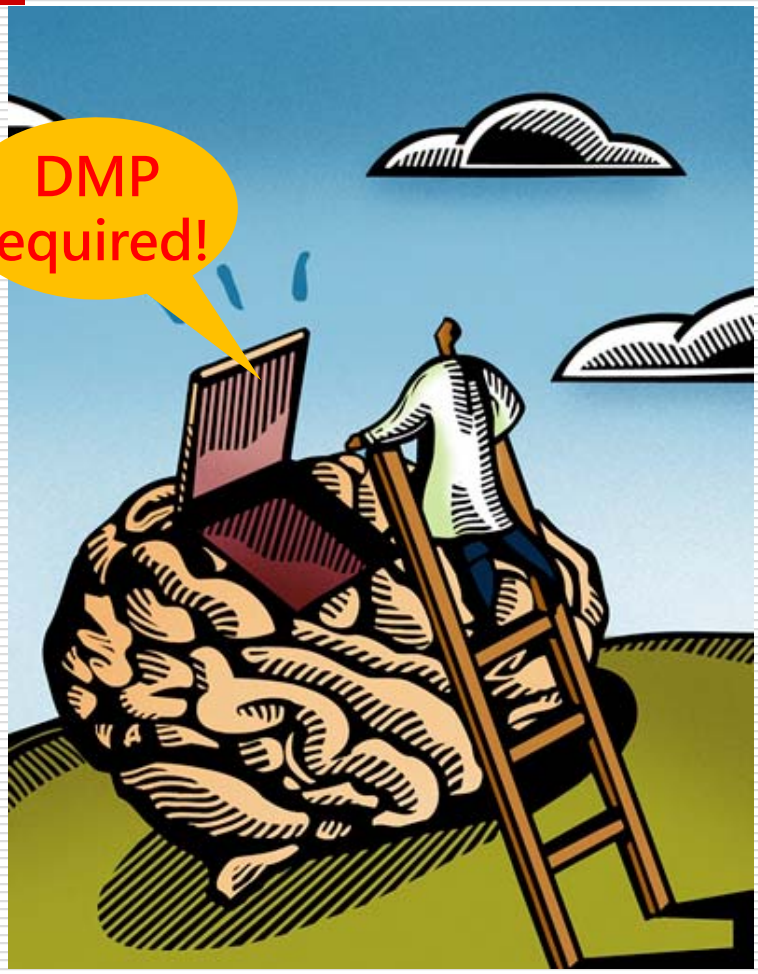
What drivers appeal to academics?

- Researchers working on specific research topics are not necessarily interested in how to perform research or changing research topic to societal challenges.
- Research transparency issue is a nightmare for most of the honest researchers.

4. Developments of Open Science and its Actors

Mandates for Research Data Management for Publicly-funded research

- Funding agencies requiring data management plans (DMPs) from researchers
  - EPSRC in UK requiring institutions to be responsible.

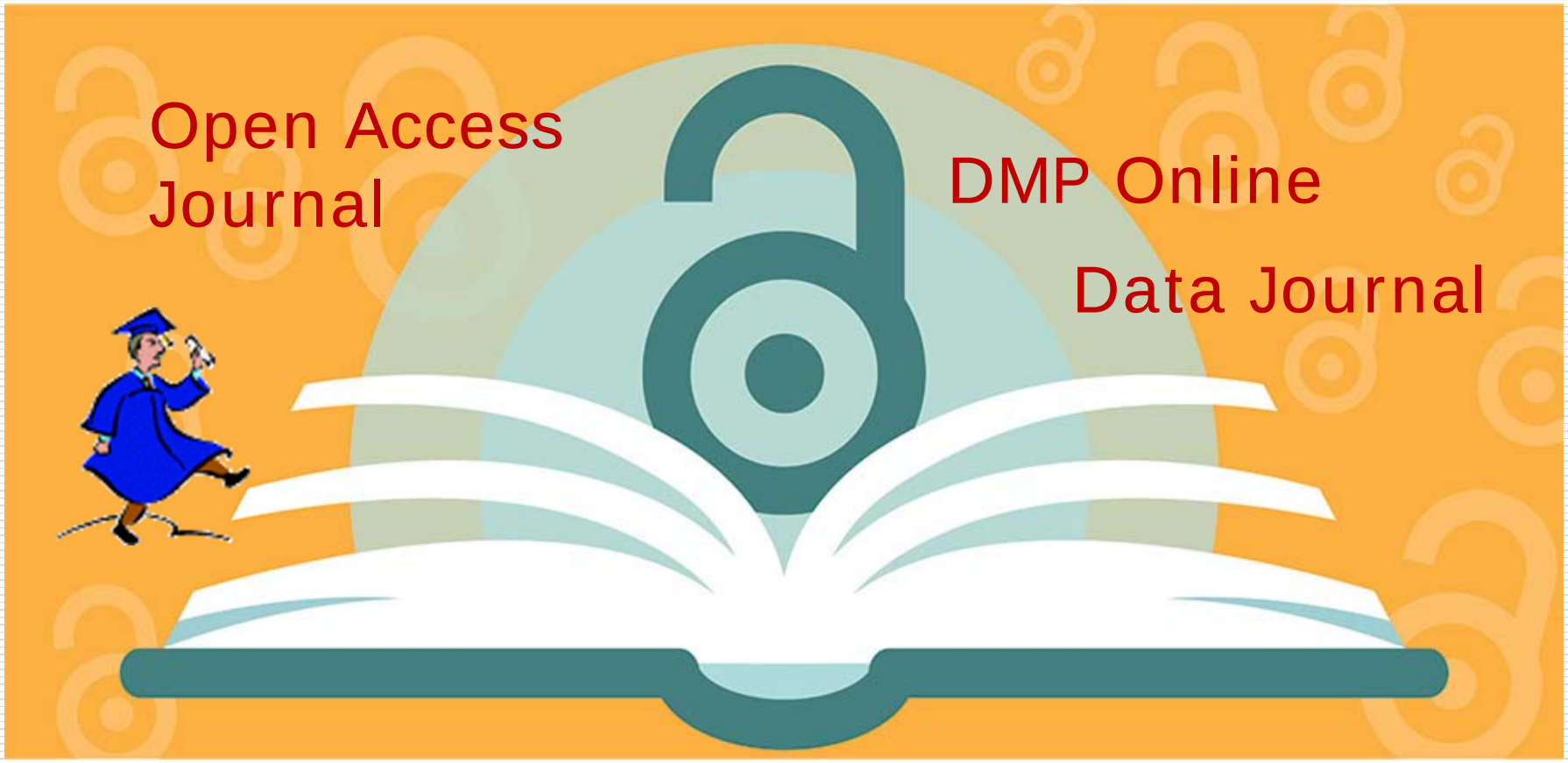


Provision of data repositories and data analysis tools

- Provision of research infrastructure for research competitiveness and research transparency purposes.

Provision by Academic Institutions and Governments

Open Accessing of Research Publications and RDM



University Libraries Publishers

Research Evaluation Reform

Scholarly Communication Community

Publishers

Enhancing Transparency and Reproducibility

Governments
Funding agencies

Citizen Science

Actors of Open Science

Funding Agency

We have kept the balance by keeping the **tradition!**

University Administration

Publishers

Researchers

Government

Citizen

University Libraries

ICT centers

...Researchers are not anymore the main actor in research!

5. 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

Changing Academic Reward System

☐ Recognizing data

- ➢ Data journals
- ➢ Data citations

☐ Recognizing social impact

- ➢ Altmetrics

http://www.nature.com/sdata/ https://www.journals.elsevier.com/data-in-brief https://www.datacite.org/
http://altmetrics.org/manifesto/ https://impactstory.org/ https://www.altmetric.com/ http://readermeter.org/

Making data available increases citations

- Alter, Pienta, Lyle
  - 240%, social sciences *
- Piwowar, Vision
  - 9% (microarray data)†
- Henneken, Accomazzi
  - 20% (astronomy) #

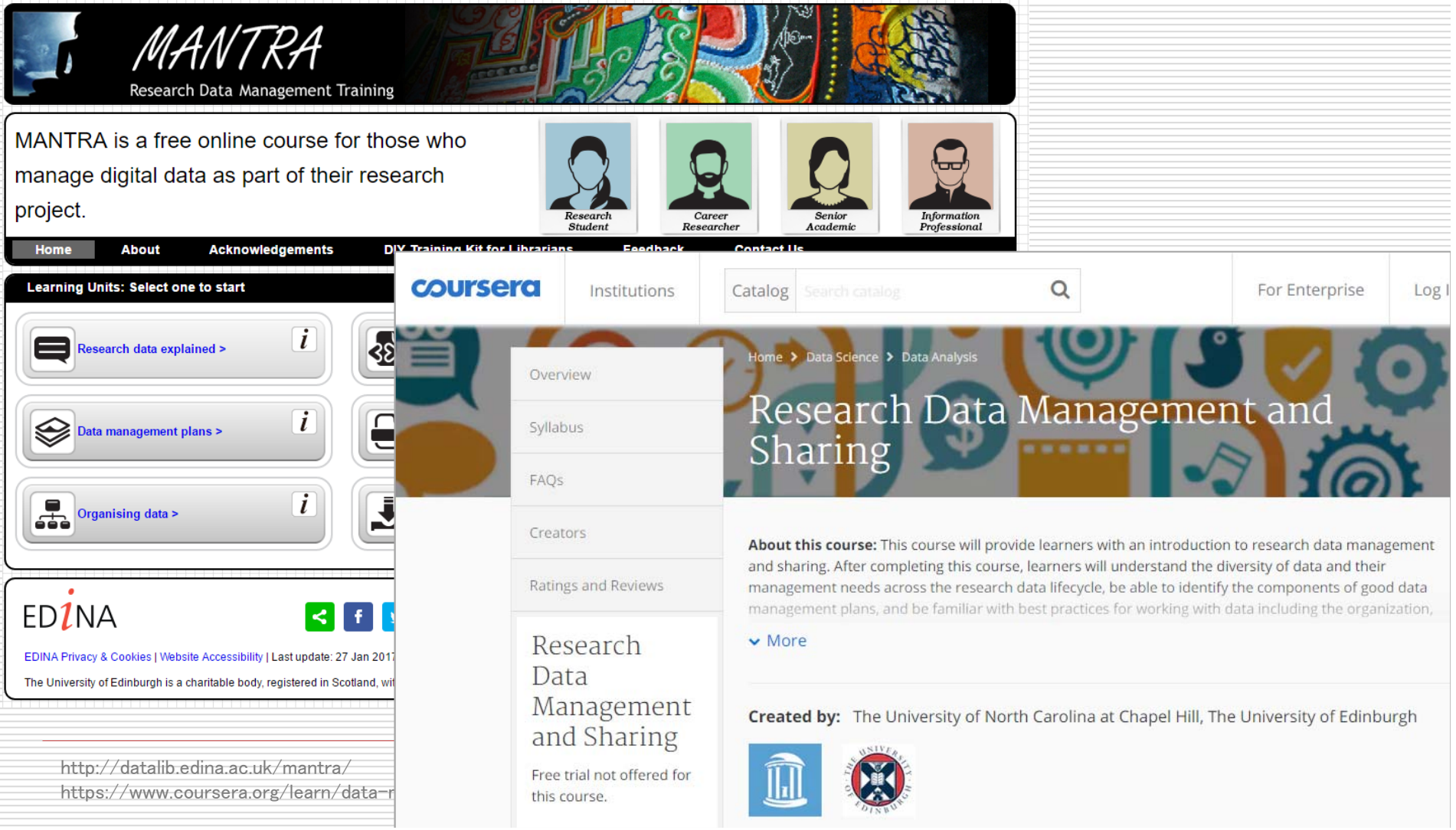
Edwin Henneken, Alberto Accomazzi, (2011) Linking to Data – Effect on Citation Rates in Astronomy. http://arxiv.org/abs/1111.3618

* Amy Pienta, George Alter, Jared Lyle, (2010) The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data. http://hdl.handle.net/2027.42/78307

† Piwowar H, Vision TJ. (2013) Data reuse & the open data citation advantage. PeerJ PrePrints 1:e1v1 http://dx.doi.org/10.7287/peerj.preprints.1v1

Source: Kevin Ashley, Digital Curation Center, UK
Porto research data conference

Providing Training for Research Data Management



http://datalib.edina.ac.uk/mantra/
https://www.coursera.org/learn/data-r

Changing Scholarly Communication ...Peer Review System

Open Peer Review

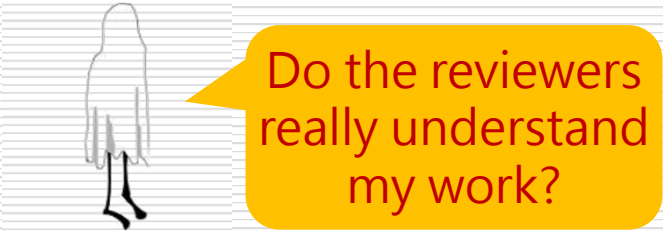
- Reviewer's comments are open to public with/without the name of reviewer
- Enabling transparent peer review

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



Can you convince the academics?

- The researchers have built up the academic reward system to be respected for the articles they produce.
- If this should change, the change has to be initiated by the researchers, and not from outside.

6. Discussion: Possibility of Universities Adopting the Idea of Open Science

Misalignment of values in pursuing Open Science

Aspiration for excellent research

Attempts to align researchers' values and the purposes of Open Science

- Research evaluation reform
  - Not accepted by academics.
- Evidence that sharing data leads to enhanced citations.
  - Not compelling enough as academics know that they are respected on the quality of research articles and not by research data which are the basis for their theoretical thoughts.

Different actors in a university

- University administration and research office
  - Try to follow the mandates of funding agencies for institutional compliance.
  - Some pursue OS as a way to be research competitive.
- University libraries and IT centers
  - Try to provide research support.
  - Greatly excited as OS gives them greater role.
- Researchers
  - Follow the mandates by funding agencies but do not see the point…

The needs for Open Science on the side of academics

1. Research data is getting out of control even for the researchers.
2. Researchers working on societal problems are not adequately evaluated and need alternative reward system.
3. There are researchers who try to make research more transparent systematically.
4. Academics favor openness in general.

Possible Future of Open Science

- Trying to pursue Open Science in top-down manner is hardly worked out.
- However, the proliferation of digital technology is adopted by researchers, and will eventually transform how science is conducted.
- Also, the strong attention towards solving societal issue will gradually be incorporated by the academics.

The idea of Open Science will be adopted without forces.

Accelerating the adoption of Open Science

- If pursuing Open Science in a rather forceful way…
  - Approaching the university administration where the hiring and promotions occur might work.
  - University administration might favor the idea to get competitive advantage over peer institutions.