Enabling Scientific Data Sharing and Re-use

R. Darby, S. Lambert, B. Matthews, Michael Wilson
Science and Technology Facilities Council, Didcot, UK

K. Gitmans
Alfred Wegener Institute for Polar and Marine Research, Germany

S. Dallmeier-Tiessen, S. Mele
CERN, Geneva, Switzerland

J. Suhonen
CSC - IT Center for Science, Espoo, Finland
“We must give taxpayers more bang for their buck. Open access to ... data is an important means of achieving this.”

Máire Geoghegan-Quinn, European commissioner for research, innovation and science

July 2012.
The study

1) collect 21 examples of data sharing
2) create baseline model of sharing;
3) test model completeness in workshop;
4) identify drivers, barriers and enablers
5) prioritise through interviews with 55 experts;
6) derive recommendations for stakeholders.
Stakeholders

- Policy-makers: National, Regional
- Funders: Research, Infrastructure
- Researchers: Data producers, Data consumers
- Research and education organisations
- Data management and infrastructure service providers (librarians, publishers)
Data Sharing Process

- Research Strategy
- Project Funding
- Data collection/simulation
- Data Analysis
- Scientific Publication
- Preservation Business Case
- Preservation Planning
- Pre-archive phase
- Data Preservation
- Discover data
- Access data

Social & Economic Impact Creation
Context Factors

- Academic discipline:
  - Source of data, Cost of data collection, Possibility to collect data again, Complexity of data analysis

- Country:
  - Legislation, Infrastructure, Funding

- Age of researcher:
  - Willingness to invest effort for possible long-term benefit

- Data Re-use Sector:
  - Non-commercial research, Commercial research, Education
Drivers

• a) Societal benefits
• b) Academic Benefits
• c) Research Benefits
• d) Organisational Incentives
• e) Individual Contributor Incentives
Barriers

f) Individual Contributor Incentives
g) Availability of a Sustainable Preservation Infrastructure
h) Trustworthiness of the data, Data Usability, Pre-archive activities
i) Data Discovery
j) Academic Defensiveness
k) Finance
l) Subject Anonymity and Personal Data Confidentiality
m) Legislation/Regulation
Enablers

n) Individual Contributor barriers
o) Availability of a Sustainable Preservation Infrastructure
p) Trustworthiness of the data, Data usability, Pre-archive activities
q) Data Discovery
r) Academic Defensiveness
s) Finance
t) Subject Anonymity and Personal Data Confidentiality
u) Legislation/Regulation
Data Publication Practice
Themes

• The role of publishers in data sharing
• Data citation and description for discovery and use
  – Granularity of data to be cited
How big is the Data Problem for journals?

Too big for the Jnl of Neuroscience and Cell:

**Jnl of NeuroScience:**
The Graph depicts the average size of a Journal of Neuroscience article and supplemental material in megabytes.

As a consequence, the Journal no longer accepts supplementary files to manuscripts, soon the supplementary material would outgrow the article volume. The burden on the peer review process became simply to large.

**Journal Cell:**
Editors suspect researchers to treat supplements as data dumping grounds (Emily Markus, Cell)

**General:**
Publishers cannot guarantee proper preservation and future accessibility of supp files.
Data Management Infrastructure

Themes

- Finance: funding infrastructure and data services;
- Quality assurance of research data
- Standards and interoperability
Interdisciplinary data sharing standards: common semantics
The Ideal Data Pyramid

1. More integration of text and data, viewers and seamless links to interactive datasets
2. Only if data cannot be integrated in article, and only relevant extra explanations
3. Seamless links (bi-directional) between publications and data, interactive viewers within the articles
4. More Data Journals that describe datasets, data mgt plans and data methods
Culture and Policy Themes

- Data sharing culture
- Public visibility of research data
- National and regional policy and legal frameworks
- Incentives in the academic reward system for good data practice
Incentives in the academic reward system

• “data sharing could work against individual scientists' need for recognition” - Gerrit Hirschfeld, Nature

• a common system of credit and recognition for data production and sharing is needed

• provide researchers with clear instructions on how to cite data

• Include data publication & citation metrics in researcher appraisal
Questions?