Crystallographic Information and Data Management Symposium
University of Warwick, Sun 25 August 2013

Research data management and UK funding policies

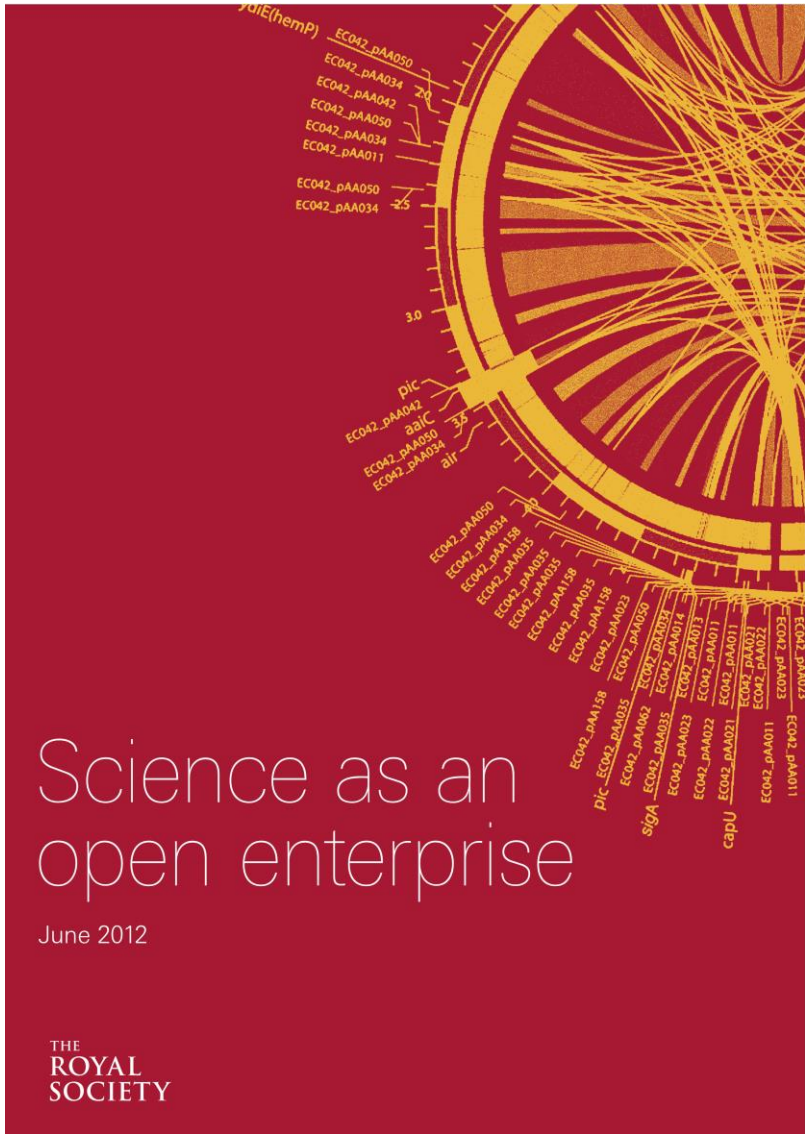
Simon Hodson
Executive Director CODATA
www.codata.org
execdir@codata.org
@simonhodson99

Research Data Policy Directions

- Research funder policies, legislative frameworks, good practice, open data agenda
  - The outputs of publicly funded research should be publicly available.
  - The evidence underpinning research findings should be available for validation
  - Greater return on investment through reuse of research data
- Good data management is good for research
  - More efficient research process, avoidance of data loss, research benefits of data reuse
- Aligns with university mission to provide excellent research infrastructure.
- Pressure on funders and universities to have better oversight of research outputs and impacts.
- Some moves from journals and learned towards policies for availability of underlying research data.

Royal Society *Science as an Open Enterprise* Report, 2012

- 'how the conduct and communication of science needs to adapt to this new era of information technology'.
- **Intelligent Openness:** data should be accessible, assessable, intelligible, usable.
- '**As a first step towards this intelligent openness, data that underpin a journal article should be made concurrently available in an accessible database**. We are now on the brink of an achievable aim: for all science literature to be online, for all of the data to be online and for the two to be interoperable.'
- Royal Society June 2012, Science as an Open Enterprise, http://royalsociety.org/policy/projects/science-public-enterprise/report/



Science as an Open Enterprise Report: six key changes

- **a shift away from a research culture where data is viewed as a private preserve;**
- expanding the criteria used to evaluate research to give **credit for useful data communication** and novel ways of collaborating;
- the development of **common standards** for communicating data;
- **mandating intelligent openness for data relevant to published scientific papers;**
- strengthening the cohort of **data scientists** needed to manage and support the use of digital data (which will also be crucial to the success of private sector data analysis and the government's Open Data strategy);
- the development and use **of new software tools** to automate and simplify the creation and exploitation of datasets.
- Royal Society 2012, Science as an Open Enterprise, http://royalsociety.org/policy/projects/science-public-enterprise/report/

International Dimensions to Data Policy Directions

- Build on **OECD principles**: http://www.oecd.org/sti/sci-tech/38500813.pdf
- Related to OA: the outputs of publicly funded research should be publicly available.
- US funders require DMPs (NSF) and promote data sharing.
- **US Office of Science and Technology Policy** Memorandum 'Expanding Public Access to the Results of Federally Funded Research': http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research
  - Funders required 'to develop a plan to support increased public access to the results of research', including data. Draft plans were due 22 Aug 2013.
- **EC Horizon 2020** will require Data Management Plans: see discussion on OKF Blog http://blog.okfn.org/2013/07/16/ec-consultation-on-open-research-data/
- **G8 Science Ministers Statement:** https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/206801/G8_Science_Meeting_Statement_12_June_2013.pdf
  - Addressing global challenges requires coordination, open data;
  - Need for Global Research Infrastructure;
  - Need for Open Scientific Research Data;
  - Need to expand access to scientific research results.

RCUK: Research Funder Principles

- RCUK Common Principles on Data Policy: http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx
- **Public good:** Publicly funded research data are produced in the public interest should be made openly available with as few restrictions as possible
- **Planning for preservation:** Institutional and project specific data management policies and plans needed to ensure valued data remains usable
- **Discovery:** Metadata should be available and discoverable; Published results should indicate how to access supporting data
- **Confidentiality:** Research organisation policies and practices to ensure legal, ethical and commercial constraints assessed; research process should not be damaged by inappropriate release
- **First use:** Provision for a period of exclusive use, to enable research teams to publish results
- **Recognition:** Data users should acknowledge data sources and terms & conditions of access
- **Public funding:** Use of public funds for RDM infrastructure is appropriate and must be efficient and cost-effective.

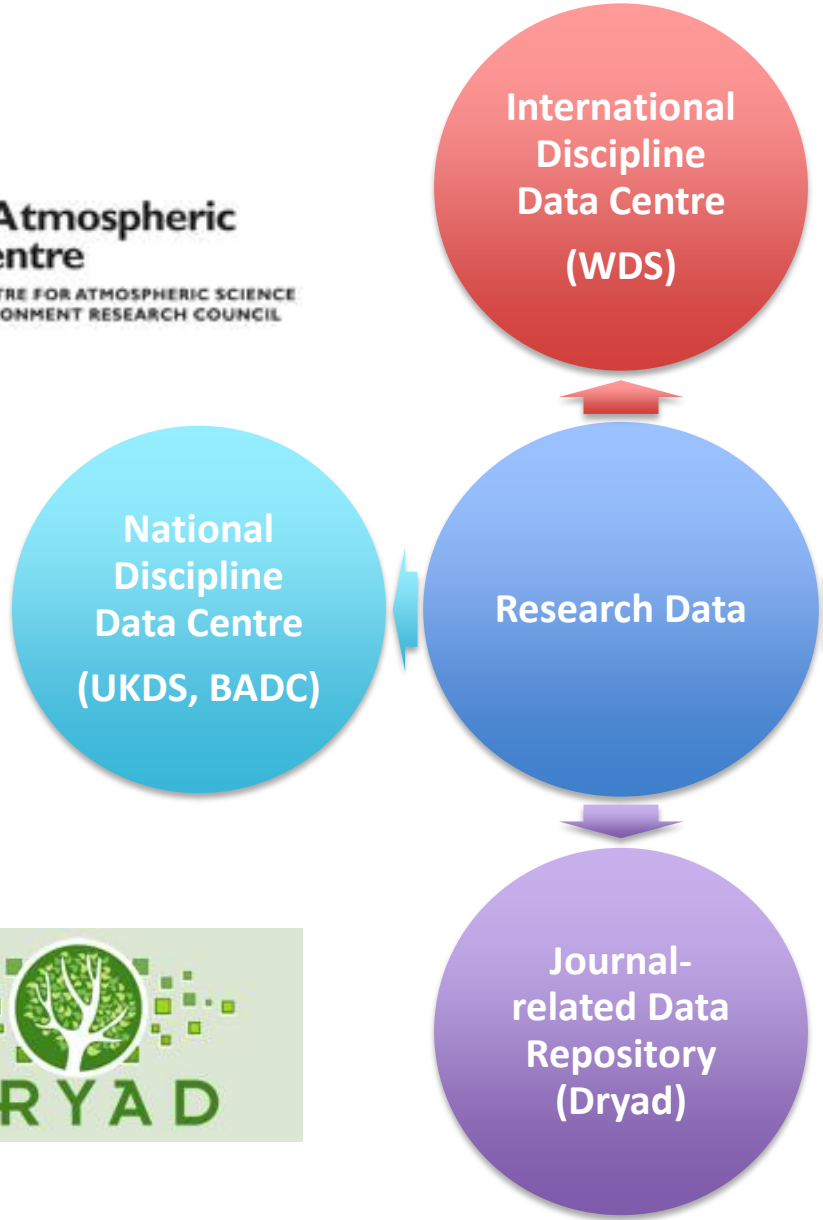
EPSRC Research Data Policy Expectations

- Policy and expectations: http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx
- Research organisations to have RDM policy, advocacy and support functions. (i, iii)
- Research data to be effectively managed and curated throughout the life-cycle (viii)
- **Research organisations to maintain public catalogue of research data holdings, adequate metadata and permanent identifier (v)**
- **Publications to indicate how research data can be accessed (ii)**
- Data to be retained for 10 years from last access (vii)
- Research data management to be adequately resourced from appropriate funding streams (ix)
- Roadmap in place by 1 May 2012
- Compliance by 1 May 2015

EPSRC

Pioneering research and skills

Where should research data go?



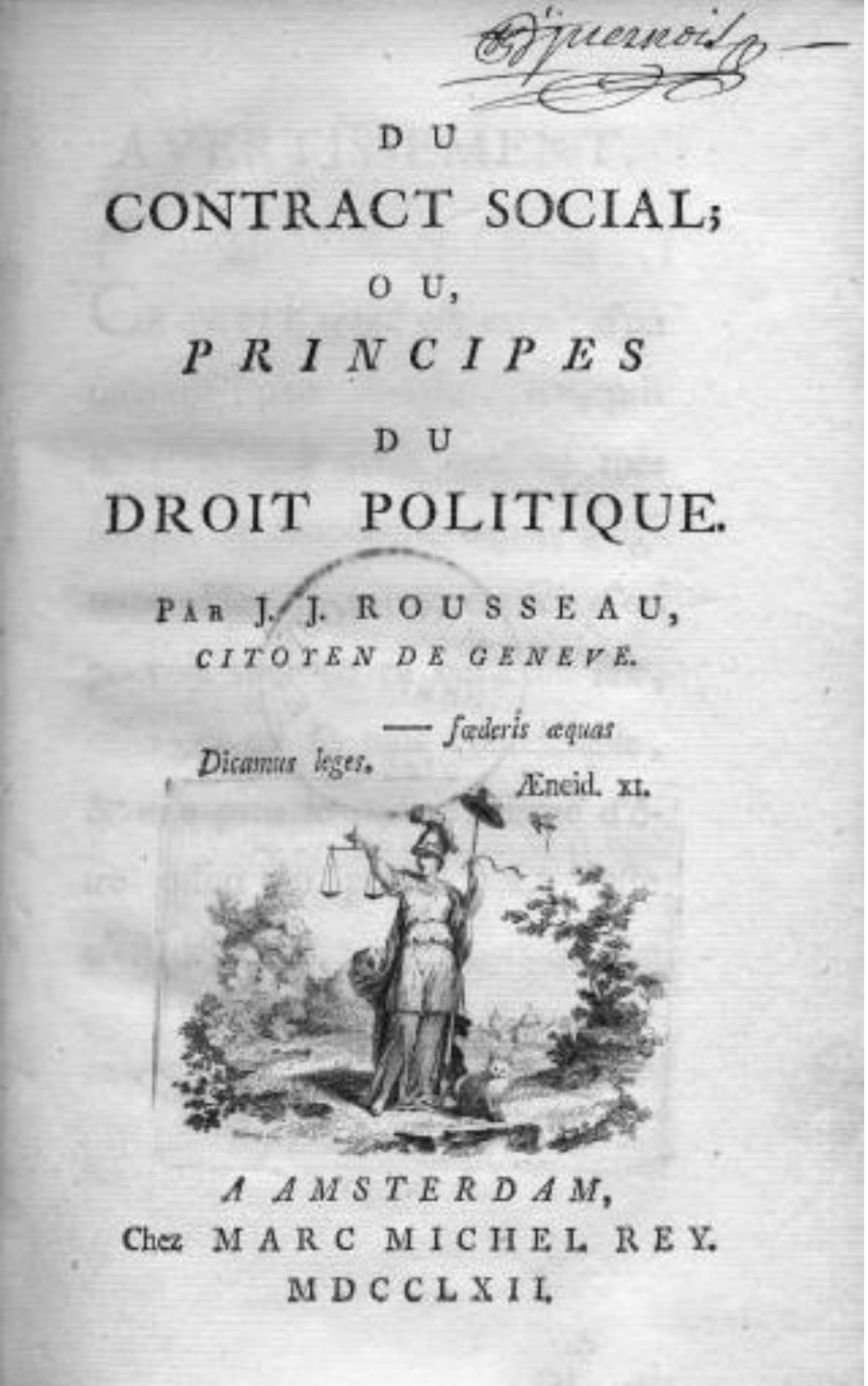
British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL

ICSU WORLD DATA SYSTEM

Jisc

DRYAD

DCC



The credibility and effectiveness of the research enterprise is due in large part to **the social contract behind scholarly publishing**. Researchers disclose their work to their peers in return for professional credit. In so doing, they also expose their findings to be confirmed or refuted, and enable other researchers to build upon their results. **Dryad seeks to extend this social contract to research data by providing a model for how a disciplinary repository can motivate researchers to disclose the data that is of the greatest value for scientific reuse, that associated with publications**, and realize the manifold benefits of free access to scientific data in perpetuity.

http://datadryad.org/pages/about

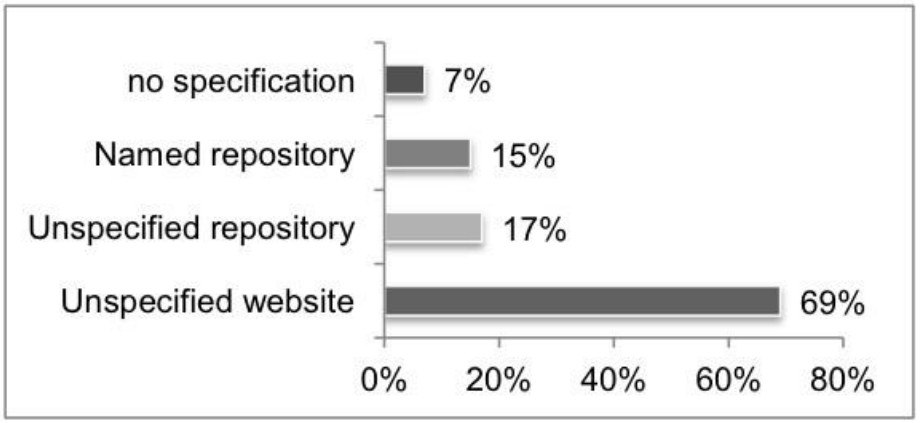
DRYAD

Vision 'Open Data and the Social Contract of Scientific Publishing'
http://www.bioone.org/doi/full/10.1525/bio.2010.60.5.2

Other 'Policy' Directions: Journal Data Policies

- **Journal data availability policies:**
- JorD project http://jordproject.wordpress.com/ reports that nearly 50% of journals sampled have a data availability policy of some sort (though only 25% of these can be characterised as 'strong').
- **Journal policies generally not clear or specific about repository, standards etc.**
- Policies > Standards > Data repositories : http://www.biosharing.org/

FIGURE 5: LOCATIONS OF DATA SUGGESTED BY JOURNAL DATA SHARING POLICIES



Understanding Data Publication Processes: PREPARDe Project

- Examined and modeled a number of **workflows** for data publication (publishing data associated with research publications).
  - Report on publication processes.
- Data repository **accreditation**.
  - White paper of principles of repository accreditation to be released.
- Scientific **review** of datasets.
  - White paper of principles and recommendations on data peer review.
- **Cross-linking** between repositories and data publishers.
  - Requirements for a third party broker to facilitate multi-directional linking between datasets and literature.
- See http://proj.badc.rl.ac.uk/preparde/wiki/DeliverablesList and http://www2.le.ac.uk/projects/preparde

Key Research Data Challenges

- Open by default, but develop clarity about the limits of openness;
- Developing and **sustaining** the data infrastructure;
- More clarity in policies about standards, data resources;
- Standards for discovery and reuse;
- Effective linking of data to publications;
- Citation of data;
- Incorporation of data availability / publication in research credit;
- Developing data science as a discipline and data scientists…

CODATA's Mission

To strengthen international science for the benefit of society by promoting improved scientific and technical data management and use.

CODATA's Strategic Plan

Maintain an international **leadership** role in the field of scientific data and information.

Provide and **influential and authoritative voice** in national and international policy regarding scientific data management.

Provide a **focal point** for international, cross- disciplinary collaboration and communication on key scientific data issues.

Unique position comprises national members' committees, International Union members, Task Groups, strategic initiatives, close relationship to ICSU.

Elements of Strategic Plan

1. **Policy frameworks for data:** take the lead in defining a policy agenda for scientific data.
   - First step is to establish **Data Policy Committee.** Provide focus and expertise.
2. **Frontiers in data science and technology:** coordinate work in key frontiers of data science and interdisciplinary application areas.
   - Current activities: nanotechnology, data for sustainable development, approaches to data recovery.
   - International **Science Data Conference**, with WDS, New Delhi 2-5 Nov 2014.
   - Reinvigorate the **Data Science Journal**.
   - **Task Groups** http://www.codata.org/taskgroups/index.html **and Working Groups (e.g. Young Data Scientists).**
3. **Data strategies for international science:** support major ICSU scientific programmes to address data management needs (including infrastructure, policies, processes, standards).
   - Integrated Research on Disaster Risk
   - Future Earth

Data Citation, Standards and Practices

- Co-Chairs: Christine Borgman, Jan Brase, Sarah Callaghan; Consultant: Paul Uhlir; see http://www.codata.org/taskgroups/TGdatacitation/index.html
- Involvement of a range of key organisations and experts.
- Major Report ***Out of Cite, Out of Mind*** to be released on **3 September 2013**
- Forceful set of 'First Principles' for data citation:

1. **Status of Data:** Data citations should be accorded the same importance in the scholarly record as the citation of other objects.
2. **Attribution:** Citations should facilitate giving scholarly credit and legal attribution to all parties responsible for those data.
3. **Persistence:** Citations should be as durable as the cited objects.
4. **Access:** Citations should facilitate access to data by humans and by machines.
5. **Discovery:** Citations should support the discovery of data and their documentation.
6. **Provenance:** Citations should facilitate the establishment of provenance of data.
7. **Granularity:** Citations should support the finest grained description necessary to identify the data.
8. **Verifiability:** Citations should contain information sufficient to identify the data unambiguously.
9. **Metadata Standards:** Citations should employ widely accepted metadata standards.
10. **Flexibility:** Citation methods should be sufficiently flexible to accommodate the variant practices among communities.

CODATA Geo Sharing

- CODATA involved in implementation of the **Group on Earth Observations (GEO) Data Sharing Principles since 2008**
  - Data Sharing Implementation White Paper published in 2009 by the Journal of Space Law and CODATA Data Science Journal
  - Helped draft data sharing guidelines and organize numerous side events at GEO and other events
- CODATA represented ICSU as one of six co-chairs of the GEO Data Sharing Task Force, 2009-11 (P. Uhlir and R. Chen along with J. Gabrynowicz)
  - Data Sharing Action Plan accepted by GEO-VII Plenary in 2010; included establishment of the GEOSS DataCORE
- CODATA again representing ICSU as one of the co-chairs of new GEO Data Sharing Working Group (P. Uhlir and R. Chen)
  - Addressing GEOSS Data-CORE, licensing, metrics, user authentication, data documentation and quality

What can CODATA do for you?

1. A membership organisation, so members get out what they put it: what can you do with CODATA?
2. CODATA offers a voice and community for data scientists and those interested in data.
3. **Isn't the data space a crowded area** (CODATA, WDS, RDA)?
    - Working closely with WDS and RDS to avoid overlap and ensure collaboration. In some areas we will collaborate on joint activities; in others ensure complementary activities.
    - CODATA has a (distinct, but not exclusive) emphasis towards policy, procedure and standards issues; also has a unique and strong membership (national committee model and engagement with International Unions); support for ICSU programmes.
4. Very keen to reinvigorate engagement with international scientific unions.
5. ***Crystallography seems a very good example: what can other disciplines learn – what lessons/case studies can be developed?***
6. ***What are the ongoing data issues in crystallography?***

Contact

Simon Hodson
Executive Director CODATA
www.codata.org
Email: execdir@codata.org
Twitter: @simonhodson99
Tel (Office): +33 1 45 25 04 96 | Tel (Cell): +33 6 86 30 42 59

CODATA (ICSU Committee on Data for Science and Technology),
5 rue Auguste Vacquerie, 75016 Paris, FRANCE