



DatCite Annual Conference
Inist-CNRS, Vandoeuvre-lès-Nancy, France
Thursday 7 August 2014

CODATA and international dimensions of data policy: advocacy and impact on practice

Simon Hodson
Executive Director CODATA
www.codata.org/blog
execdir@codata.org
[@simonhodson99](https://twitter.com/simonhodson99)



International Dimensions of Data Policies

- International data policy development, and CODATA's role.
- Types and evolution of data policies.
- Some observations on Data Availability Policies:
 - Not bureaucratic interference.
 - Embody important principles of good science.
 - More effective when they come from and are endorsed by research communities.
 - **Should aim to ensure that researchers benefit from 'doing the right thing' – e.g. data citation.**
 - Necessarily have an international dimension: research is an international collaborative/competitive enterprise par excellence and this makes policy frameworks important.
- Strong engagement of many countries with data sharing initiatives like GEO, Group on Earth Observations.
- Data policy and sharing issues in international science programmes.
- Role of CODATA with national members and National Committees.
- **CODATA International Data Policy Committee and other initiatives:**
 - better understanding of international policy landscape.
 - guidelines and other activities to encourage adoption and implementation of data policies.



Types of Data Policies

International Quasi-Governmental Principles

- e.g. OECD
- High level principles, charting a direction of travel

Research Funder Policies

- Expectations of professional conduct for grant recipients
- Expressions of principles of research integrity
- Developed in consultation but 'feel' more prescriptive

Research Institution Policies or Codes

- Codes of conduct
- Sometimes aspirational
- Response to funders
- Responsibilities of researchers and institutions

International Data Sharing or Research Initiative Agreements or Principles

- e.g. Genome Agreements, IPY, GEOSS, Future Earth?
- Broad principles of engagement
- Community norms

Community Norms in Journal Data Availability Policies

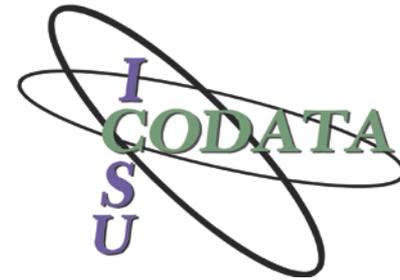
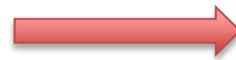
- E.g. Dryad JDAP, IUCr Journals,
- Expression of research community norms, practice



CODATA's Mission



ICSU's Mission



CODATA's Mission

*“Strengthen international science for the benefit of society **by promoting improved scientific and technical data management and use.**”*



CODATA: international collaboration for Open Scientific Data

Data Policy

- International and national aspects of data policy.
- Data policy committee: setting an international agenda for data policy, expert forum, advice and consultancy.
- Coordinating with national committees.

Data Science

- Long-standing activities: fundamental constants.
- Strategic working groups; community-driven task groups.
- Disciplinary and interdisciplinary data challenges, Big Data

Capacity Building

- Longstanding work on data preservation and access with developing countries.
- Executive Committee Task Force on Capacity Building: setting an international agenda for capacity building; Early Career WG

Data for International Science

- Support for ICSU Mission.
- Data issues and challenges in international, interdisciplinary science programmes.



GEO Data Sharing Principles, 2005



Intergovernmental organisation that promotes the availability and sharing of earth observation data (satellite and in situ).

1. There will be **full and open** exchange of data, metadata and products shared within GEOSS, recognizing relevant international instruments and national policies and legislation;
2. All shared data, metadata and products will be made available with **minimum time delay and at minimum cost**;
3. All shared data, metadata and products being free of charge or no more than cost of reproduction will be encouraged for research and education.

GEO Data Sharing Principles: http://bit.ly/GEO_DSPs

CODATA and GEO DSWG: http://bit.ly/CODATA_GEO_DSWG



OECD Data Principles, 2007



OECD Principles and Guidelines for Access to Research Data from Public Funding

http://bit.ly/oecd_principle

S

OECD Principles and Guidelines released in 2007, built on the OECD Declaration on Access to Research Data from Public Funding http://bit.ly/OECD_Declaration_2004

Findings of OECD Follow-Up Group, CODATA Data Science Journal: <http://dx.doi.org/10.2481/dsj.3.135>

US National Research Council study, *Bits of Power*: pointed out:

The value of data lies in their use. Full and open access to scientific data should be adopted as the international norm for the exchange of scientific data derived from publicly funded research.

OECD Data Principles, 2007

OECD Principles:



OECD Principles and Guidelines for Access to Research Data from Public Funding

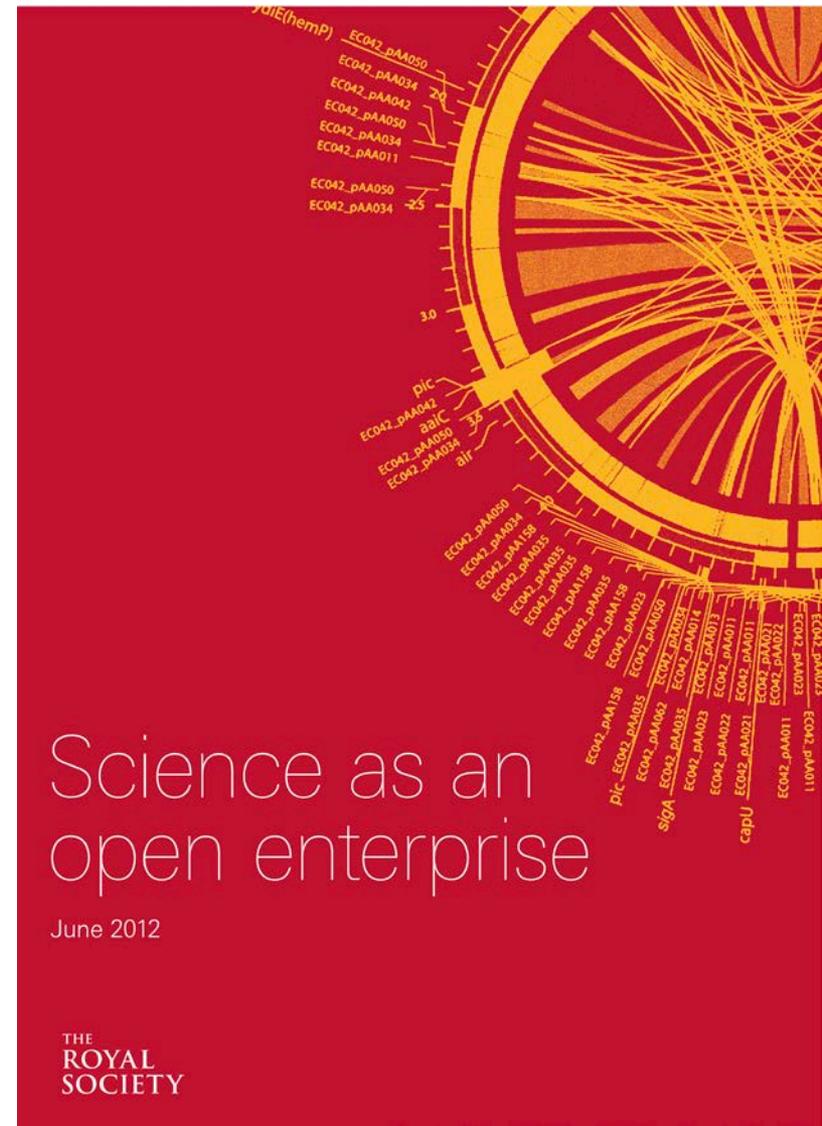
http://bit.ly/oecd_principle

S

- A. **Openness:** 'Openness means access on equal terms for the international research community at the lowest possible cost, preferably at no more than the marginal cost of dissemination. Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based.'
- B. **Flexibility:** take into account the needs of each research area and different funding and legal regimes.
- C. **Transparency:** importance of the transparency and visibility of data sources.
- D. **Legal Conformity:** respect for legal rights, security, privacy etc.
- E. **Protection of Intellectual Property**
- H. **Interoperability:** due attention to semantic and technical interoperability, documentation standards.
- I. **Quality:** pay attention to particular attention to ensuring compliance with explicit quality standards.
- F. Responsibility, G. Professionalism, K. Efficiency, L. Accountability and M. Sustainability in running of data infrastructures.

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

- Opportunities for to advance science through using data are being delayed or lost.
- Science is not properly conforming to its principles of verification, reproducibility and collective self-correction.
- Data should be open by default.
- Sometimes there are good reasons why data should not be made open: but unless such reasons apply they should be open.
- **Intelligent openness: data should be accessible, assessable, intelligible, usable.**
 1. Data must be **accessible** and readily located.
 2. Data must be **intelligible** to those who wish to scrutinise them.
 3. Data must be **assessable** so that judgements can be made about their reliability and the competence of those who created them.
 4. Data must be **usable** by others.





G8 Science Ministers' Statement

Open Scientific Research Data

1. To the greatest extent and **with the fewest constraints possible** publicly funded scientific research data should be open, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests, whilst acknowledging the legitimate concerns of private partners.
 2. **Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards.**
 3. To maximise the value that can be realised from data, the mechanisms for delivering open scientific research data should be efficient and cost effective, and consistent with the potential benefits.
 4. To ensure successful adoption by scientific communities, open scientific research data principles will need to be **underpinned by an appropriate policy environment, including recognition of researchers fulfilling these principles, and appropriate digital infrastructure.**
- G8 Science Ministers' Statement, 18 June 2013:
<https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>



G8 Open Data Charter, 2013

- Applies governmental/administrative, statistical and research data produced by public funding.
- 1. **Open by Default:** along with other recent statements, stresses the principle of 'Open by Default' as a way of minimising *unnecessary* restrictions on the availability of research data.
 - **Advances a specific definition of 'Open', which covers:**
 - **Open Data:** Open data should be available **free of charge** in order to encourage their most widespread use.
 - **Open Formats:** non-proprietary and interoperable formats should be used.
 - **Open Licenses:** use Open licenses where possible, so that no restrictions or charges are placed on the re-use of the information for non-commercial or commercial purposes, save for exceptional circumstances.
- 2. **Quality and Quantity:** strike a balance between quality and availability; release to improve quality.
- 3. **Usable by All:** available to all citizens.
- 4. **Releasing Data for Improved Governance:** to promote transparency and accountability.
- 5. **Releasing Data for Innovation:** reuse of data for research and enterprise.
- G8 Open Data Charter and Technical Annex, 18 June 2013:
<https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>



European Union Horizon 2020 Data Policy

- New requirement for projects in many parts of Horizon 2020 Programme (exceptions are sections of programme with particularly strong commercial collaborations).
- **All written outputs must be Open Access:** http://bit.ly/H2020_OA_Guide
- **All significant data produced by projects must be made available unless there are appropriate reasons not to:** http://bit.ly/H2020_DMP_Guide

Principles for Data Released Under Horizon 2020

- **Discoverable** and **accessible**; **assessable**, **intelligible** and **usable** beyond the original purpose for which it was collected; **interoperable** to specific quality standards

Data Management Plan: proposals must include a brief data management plan, which must be expanded if funded.

- Data must be deposited in an appropriate repository or archive:
 - data underpinning publications should be released immediately;
 - other significant data should be deposited within a reasonable timescale.
- Open licence; **permanent identifier**, **data citation**; **allows period of exclusive access**; accompanied by software/code necessary for reproduction.



HORIZON 2020

The EU Framework Programme for Research and Innovation



GEO Data Sharing Principles, post 2015?



1. Data, metadata and products will be shared through GEOSS as **Open Data by default**, by making them available as part of the GEOSS Data-CORE without charge, without restrictions on reuse, subject to the conditions of registration and attribution when the data are reused;
2. Where international instruments, national policies or legislation preclude the sharing of data as Open Data they should be made available through GEOSS with minimal restrictions on use and at no more than the cost of reproduction and distribution;
3. All shared data, products and metadata will be made available through GEOSS with minimum time delay.



GEO-X Plenary & Geneva Ministerial Summit
Side Event

Implementation of Data Sharing Principles

Monday, 13 January 2014, 14:30-18:00
Salle 5+6, Level 3, CIGG



GEO Data Sharing Principles, post-2015: http://bit.ly/GEO_DSPs



Data Principles for Future Earth

- Data principles proposed by CODATA and WDS.
 1. **Excellence in Data Management:** data generated and modified in Future Earth, and associated research products—such as code—will be managed throughout the research lifecycle in accordance with good practice.
 2. **Openness and Protection:** data, and other research products, generated and modified in Future Earth will be made as openly available as possible, with minimum delay and at minimum cost, while respecting relevant international agreements, national policies and legislation for the protection of personal, sensitive and commercial data;
 3. **Integrity and Legacy:** data, and other research products, generated and modified in Future earth will be discoverable, accessible, intelligible and reusable, in the short and long term, and will therefore be selected appropriately, quality-assessed, furnished with appropriate metadata, machine readable licences, and maintained in trusted digital repositories.



CODATA-ICSTI Task Group Data Citation, Standards and Practices

For Attribution

Workshop and Report:

[http://bit.ly/for attribution](http://bit.ly/for_attrIBUTION)

Out of Cite, Out of Mind

[http://bit.ly/out of cite](http://bit.ly/out_of_cite)

Joint Declaration of Data Citation Principles:

<https://www.force11.org/datacitation>

Background and Developments:

[http://bit.ly/data citation principles](http://bit.ly/data_citation_principles)

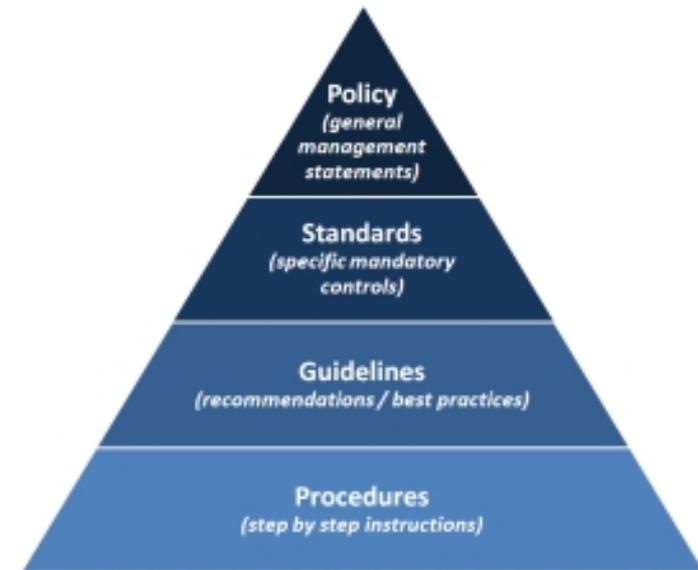
DC¹

Data Citation Principles



CODATA Data Policy Committee: Vision

- Designed to help CODATA fulfill mission and lead on Data Policy Issues internationally.
- **Genuinely international and diverse in membership and perspective: c. 12 members from**
- **Advisory Body:** advise international programmes and other initiatives.
- **Agenda Setting:** position statements (short statements or white papers) on data policy issues.
- **Benefit CODATA members:** engage with National Committees and Unions where appropriate.
 - Promote greatest possible availability of data assets.
 - Engage in a process of validation and iteration with National CODATA Committees (and academies, funders, transnational bodies ...), with International Scientific Unions (and Learned Societies, publishers, journal editorial boards).



CODATA National Committees

- **What are the benefits of having a CODATA National Committee?**
 - a point of contact and engagement with CODATA;
 - a group to build and coordinate collaboration on CODATA agendas and activities (including data policies and standards, Task Groups, capacity building and training, data issues in developing countries, early career data professionals);
 - an entity to collaborate with other National Committees, bilaterally or in groups;
 - a forum by which national stakeholders (research funders, National Academies, research institutions, data centres, learned societies, research libraries, etc) may raise and advance various agendas particularly those with an international dimension.





Indian National Data Sharing and Accessibility Policy

- *Global experience has demonstrated convincingly that access to data leads to breakthroughs in scientific understanding as well as to economic and public good, in addition to several benefits to civil society. **Given the deployment of substantial level of investment of public funds in collection of data and the untapped potentials of benefits to social society, it has become important to make available non-sensitive data for legitimate and registered use.***
- National Data Sharing and Accessibility Policy (NDSAP), March 2012:
<http://www.dst.gov.in/NDSAP.pdf>
- Places emphasis on a **negative list** of sensitive data types, rather than a positive list of data to be released: i.e. the default is open, unless the data is on the 'stop' list.
- Allows for data to be Open, accessible to registered users and under restricted access.





Indian National Data Sharing and Accessibility Policy

- Implementation Guidelines, Feb 2014:
http://data.gov.in/sites/default/files/NDSAP_Implementation_Guidelines_2.2.pdf
- Deal mostly with public data, but research data produced by government institutes of funded by the government is in scope.
- *For data to be reused, it needs to be adequately described and linked to services that disseminate the data to other researchers and stakeholders. **The current methods of storing data are as diverse as the disciplines that generate it. It is necessary to develop institutional repositories, data centers on domain and national levels that all methods of storing and sharing have to exist within the specific infrastructure to enable all users to access and use it.***

The screenshot shows the official website of the Department of Science & Technology, Ministry of Science & Technology, Government of India. The page is titled "National Data Sharing and Accessibility Policy (NSDAP)". The header includes the department's name in English and Hindi, along with navigation links for Home, Sitemap, and Search. A navigation menu below the header lists: About DST, Scientific Programmes, S&T System In India, Scientific Services, and Administration. The main content area begins with a paragraph explaining the policy: "Large volumes and different types of data, including some of scientific and technical relevance are generated and compiled by various arms of the Government of India and various State Governments for meeting their specific requirements. Scientific organizations generate data and develop scientific data bases deploying huge public funds. Since such data are not generated under any standardized format, inter-operability of both scientific and technical data poses a serious challenge. Global experience has demonstrated convincingly that access to data leads to breakthroughs in scientific understanding as well as to economic and public good, in addition to several benefits to civil society. Given the deployment of substantial level of investment of public funds in collection of data and the untapped potentials of benefits to social society, it has become important to make available non-sensitive data for legitimate and registered use."

Kenya Research Data Policy?



The screenshot shows the Kenya Open Data Portal website. At the top left is the 'KENYA openData' logo. A green navigation bar contains links for Home, About Kenya Open Data, Data Catalogue, Suggest a Dataset, Community Apps, Outreach, Help, Sign Up, and Sign In. Below the navigation bar is a section titled 'Kenya Open Data Survey 2014'. The main content area features a 'COUNTRY COMMITMENTS' section with a map of Africa and a callout box for Kenya detailing its 'Open Government Partnership' commitments. To the right of the map is a 'Latest Datasets and Views' list with items like 'Poverty Rate, by District', 'County Urbanization: Nairobi', and 'Local Authority Expenditures by Year: Nairobi'. Below this are five smaller tiles: 'Health Facility Chart' (a pie chart), 'KCPE Exam Scores' (a bar chart), 'GoK shared services' (a flowchart), 'Counties' (a map of Kenya's counties), and 'Developer? Click here!' (a button with a code icon).

- Kenya Open Data Portal: <https://opendata.go.ke> ; second public data portal in Africa ; includes a number of research data sets.
- What about research data, data produced in universities, research institutes?
- Kenyan funders and research institutions do not currently have policies relating to data availability.



International Workshop on Open Data for Science and Sustainability in Developing Countries

- Convened by CODATA Task Group, *Preservation of and Access to Scientific and Technical Data in/for/with Developing Countries*.
- Partners included UNESCO; Kenyan Ministry of Information, Communications and Technology; Jomo Kenyatta University of Agriculture and Technology; GEO, ICSU-WDS, RDA.
- Strong endorsement for the workshop from Kenyan Cabinet Secretary and from local universities and research institutes.
- Discussed Open Data Principles, the application of such principles in developing countries and their relationship to scientific objectives supporting SDGs.
- Developed a set of ***Principles of Preservation of and Open Access to Research Data in Developing Counties*** and ***Guidelines*** for their implementation.
- Particular concerns for credit and periods of privileged access; concerns about giving away intellectual assets for free...





Community Norms and Journal DAPs

- Government or funders data policies increasingly prevalent (OSTP, H2020, various national funders).
- How will these policies be implemented? Importance of research community participation.
- Community Norms have been particularly effective (Bermuda and Fort Lauderdale agreements in genomics; CIF and IUCr etc.)
- **Challenges: how to...**
 - develop research community input on policy implementation?
 - improving understanding of implementation challenges and requirements?
 - advocate proportional and coordinated response to policy directives?
- Dryad Joint Data Archiving Policy: <http://datadryad.org/jdap>
- Joint declarations from journal editorial boards, Feb 2010, in American Naturalist, Evolution, the Journal of Evolutionary Biology, Molecular Ecology, Heredity, and other key journals in evolution and ecology: <http://www.journals.uchicago.edu/doi/full/10.1086/650340>
- PLoS convened workshop at IDCC. Two clear recommendations: journal DAPs, journal data citation policy.
- **Requires engagement and involvement of journal editorial boards.**





ICSU Europe-CODATA Workshop on Data Policy Development and Implementation

- Premise that despite EC H2020 Pilot Policy, there remains considerable national diversity.
- Important role can be played by ICSU Europe members in promoting policies and developing proportionate implementation guidelines.
 - Which data? When (i.e. period of privileged access)? Will I be credited?
- What is the policy landscape internationally, in Europe and above all in ICSU Europe countries?
- What are ICSU Europe members current views and activities on data policy developments?
- What is the level of development – and appetite – in different countries for an ‘Open Data regime’?
- What opportunities exist for ICSU Europe members to play a role in:
 - articulating scientific community requirements?
 - advocating proportional and coordinated response to policy directives?
 - developing scientific community input on policy implementation?
 - improving understanding of implementation challenges and requirements?
- Output: short landscape report and implementation recommendations to be published by ICSU Europe members.
- Support from CODATA and CODATA International Data Policy Committee and other partners.



COUNCIL OF FINNISH ACADEMIES



www.SciDataCon2014.org

New Delhi 2-5 Nov 2014

SciDataCon 2014

International Conference on
Data Sharing and Integration for Global Sustainability



Home	Scientific Programme	Abstract Submission	Registration	Venue	Contact	News
----------------------	--------------------------------------	-------------------------------------	------------------------------	-----------------------	-------------------------	----------------------



USGS/NASA Landsat

Get Involved!

Propose topics for sessions of the SciDataCon2014 Scientific Programme on our **Conference Forum!**

Useful Information

- About Delhi
- Accommodation
- Tours
- Visas
- Invitation Letter

Welcome to SciDataCon 2014

Research data are essential to all scientific endeavours. The emerging cultures of data sharing and publication, open access to, and reuse of data are the positive signs of an evolving research environment. Nevertheless, several cultural and technological challenges are still preventing the research community from realizing the full benefits of these tendencies.

The **Committee on Data for Science and Technology (CODATA)** and the **World Data System (WDS)**, interdisciplinary committees of the **International Council for Science (ICSU)** are supporting and encouraging these positive changes by actively promoting effective data policies and good data management practices in the research community, to produce better science, which ultimately benefits society. Likewise, the challenges and opportunities of 'Big Data' may have ramifications for the conduct of science and for the whole of society: the value and importance of data are being recognized more than ever before.

As a major contribution to this effort, the two organizations are co-sponsoring and organizing a high profile international biennial conference on scientific research data. SciDataCon will provide a unique platform bringing together international experts and practitioners in data sciences, technologies and management; researchers from the natural, social, health, and computer sciences; research funders and sponsors; and policy makers and advisors.



Thanks for your attention!

CODATA Website: <http://www.codata.org/>

CODATA Blog: <http://www.codata.org/blog/>

SciDataCon 2014: <http://www.scidatacon2014.org/>

CODATA General Assembly 2014: <http://www.codata.org/general-assembly>

Simon Hodson

Executive Director CODATA

www.codata.org/blog

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