



RECODE Conference, 'Policies for open access to research data'
National Documentation Centre, Athens, Greece
Thursday 15 January 2015

Policies for open access to research data: international and scientific community dimension

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

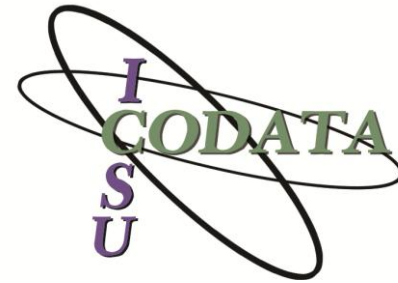




CODATA: Committee on Data for Science
and Technology



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

Open Data Policies

Current Best Practice for Research Data Management Policies
A Memo for the Danish e-Infrastructure Cooperation and the Danish Digital Library

Simon Nielsen and Lars Møller
May 2014



COUNCIL OF FINNISH ACADEMIES

GEO GROUP ON EARTH OBSERVATIONS

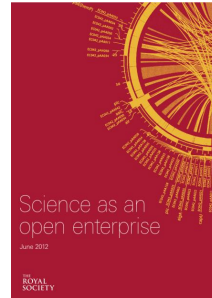
GEO-X Planning & Science Ministerial Summit
Side Event

Implementation of Data Sharing Principles

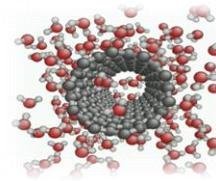
Monday, 13 January 2014, 14:30-18:00
Salle 404, Level 3, OSG



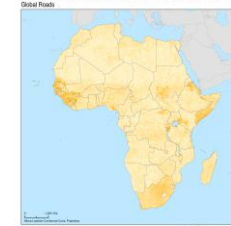
ICSU
International Council for Science



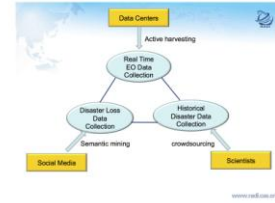
Challenges in Data Science



Global Roads Open Access Data Set, Version 1 (gROADSv1) Africa



The Global Roads Open Access Data Set (gROADSv1) was developed under the auspices of the CODATA Global Open Access Data Set (GOADS) project. The GOADS project is a joint effort of the International Council for Science (ICSU) and the International Union of Pure and Applied Chemistry (IUPAC) to create a global open access data set for the world's roads. The data set is available for public use and is intended to support research in transportation, urban planning, and other fields. The data set is available for public use and is intended to support research in transportation, urban planning, and other fields. The data set is available for public use and is intended to support research in transportation, urban planning, and other fields.



DC¹
Data Citation Principles

DSJ

Capacity Building



International Collaboration

SciDataCon 2014
International Conference on Data Sharing and Integration for Global Sustainability



SciDataCon 2014, 2-5 Nov, New Delhi





Open Data Policies

Current Best Practice for Research Data Management Policies

A Memo for the Danish e-Infrastructure Cooperation and the Danish Digital Library

Simon Hodson and Laura Molloy
May 2014

Expert Report on data policies for Danish e-Infrastructure Group

Workshop on data policies for ICSU Europe planned for Spring 2015



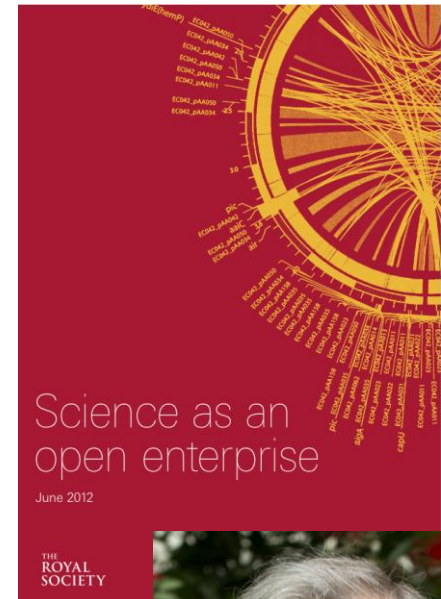
COUNCIL OF FINNISH ACADEMIES



Substantial input to ICSU Report on Statement on Open Access and Metrics
<http://bit.ly/icsu-OA-statement>



ICSU
International Council for Science



Implementation of Data Sharing Principles

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

Leading role in GEO DSWG and DMP TF
http://bit.ly/GEO_DSPs



New CODATA President
Geoffrey Boulton, FRS
Chair of *Science as an Open Enterprise* Report





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**
- **Agenda Setting:** advocacy, strategy, policy and practice for Open Science.
- **Advisory Body:** advise international programmes and other initiatives.
- **Benefit CODATA members:** engage with National Committees and Unions where appropriate.
 - Encourage development and implementation of data policies to benefit Open Science.
 - Develop policy and guidance with National CODATA Committees (and academies, funders, transnational bodies ...), with International Scientific Unions (and Learned Societies, publishers, journal editorial boards).



Open Data Policies: international and research discipline dimensions



- Some observations on Data Availability Policies:
 - Not bureaucratic interference. Embody important principles of good science.
 - **Should aim to ensure that researchers benefit from ‘doing the right thing’ – e.g. data citation.**
 - Most effective when they come from and are endorsed by research communities: in this field of research we conduct our practice in this way.
 - Should also provide guidelines on implementation.
 - Necessarily have an international dimension: research is an international collaborative/competitive enterprise par excellence and this makes policy frameworks important.
- Role of ICSU, the International Council of Science.
- Importance of international data sharing initiatives like GEO, Group on Earth Observations.
- Data policy and sharing issues in international science programmes.
- We now have many statements of principle. National and EC funder policies, some journal and institutional policies.
- **Important that various funders and community organisations work together to establish implementation guidelines for high level policies.**
- **How does this research community implement these important principles of good research practice?**

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



Data Sharing in Genomics

- Since 1991 a series of policies and agreements favouring **pre-publication data release**.
- Require deposit to a public nucleotide sequence database (GenBank, EMBL or DDBJ) within a short time of the sequence data being created.
- 1991 National Human Genome Research Institute (NHGRI) and the Department of Energy (DOE) data release policy: 6 months of generation.
- **1996 Bermuda Principles** (International Human Genome Sequencing Consortium), within 24 hours for certain data types, but allowed some exceptions and de facto limitations on reuse of whole sequences (publication priority).
- **2003 Fort Lauderdale Agreement** recommended extending Principles to all data types and removing all restrictions (24 hours or 1 week depending on data type, not publication priority for whole sequences).
 - Convened by Wellcome Trust, included an international group of data producers, users, database personnel, journal editors and funding agency representatives...
 - Recognised that each of the three **stakeholders** in the system - data producers, data users and funding agencies - has an active role to play in promulgating and helping implement the practice of open and rapid data release.



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, CICC



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

Data Management Principles Task Force preparing further guidance for data contributors.



ICSU Statement on Open Access and Metrics

- ICSU, *International Council for Science*, Statement on 'Open access to scientific data and literature and the assessment of research by metrics', Sept 2014 <http://bit.ly/icsu-OA-statement>
- Endorses the OECD Principles and Guidelines
- Recommendation 4: '**Science publishers and chief editors of scientific publications should require authors to provide explicit references to the datasets underlying published papers, using unique persistent identifiers.** They also should require clear assurances that these datasets are deposited and available in trusted and sustainable digital repositories. **Citing datasets in reference lists using an accepted standard format should be considered the norm.**'



ICSU

International Council for Science



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.



Dryad Joint Data Archiving Policy

- 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>
- Joint Data Archiving Policy: <http://datadryad.org/jdap>
- **This journal requires, as a condition for publication, that data supporting the results in the paper should be archived in an appropriate public archive, such as GenBank, TreeBASE, Dryad, or the Knowledge Network for Biocomplexity.**
- Dryad Data Repository: <http://datadryad.org/>
- Not-for-profit organisation whose mission is to maintain a data repository providing a home for data underpinning peer-reviewed research articles.
- Many other journals have developed policies...



How 'might' the data policy landscape evolve?



- We have high level principles and many funder, institutional and journal policies.
 - Which data are in scope?
 - What standards and repositories to use?
 - Timescales and embargoes?
- **Next step is one of making specific community agreements and guidelines around implementation:**
 - Guidelines on which data are in scope, with examples.
 - Guidelines on standards and repositories.
 - Agreements on timescales and embargoes.
 - Agreements on implementation and monitoring:
 - Editorial boards and review panels.
 - Publisher editors and checklists (e.g. for citations).
- Important that funders and representative bodies (e.g. ISUs) convene stakeholder meetings to establish agreements and guidelines:



Earth and Space Science Statement of Commitment



- **Statement of Commitment from Earth and Space Science Publishers and Data Facilities: A Coalition for the Publication of Data in the Earth and Space Sciences.**
- Small funding from NSF to encourage work; Town Hall Meeting at American Geophysical Union, Dec.
 - *The Coalition primarily aims to*
 - a) ***foster consensus and consistency*** among publishers, editors, funders, and data repositories on how data that are part of scholarly publications should be curated and published;
 - b) *help data repositories collect the data that are within their scope easier and more comprehensively and raise their visibility;*
 - c) ***help authors properly submit their data to repositories*** up front by simplifying the data submission process and providing sufficient information about it, and
 - d) *help journals and funding agencies by doing all of the above.*
 - *Collectively, all of these help identify best practices and promulgate them, which will ultimately improve the integrity of data associated with publication.*
- Promote adoption and implementation of data policies by journals and societies; promote accredited repositories; training; data citation; identifier and standards.



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 level of development – and appetite – in different countries for an ‘Open Data regime’?
- **Encourage ICSU Members (National Academies of Science, funders, International Scientific Unions) to articulate policies and guidance for implementation.**
- 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

POLICIES



A catalogue of data preservation, management and sharing policies from international funding agencies and regulators.



STANDARDS

FORMATS TERMINOLOGIES CHECKLISTS



A catalogue of reporting standards (minimum reporting guidelines, exchange formats and terminologies) and organizations that develop these.

DATABASES



A catalogue of databases, described according to the BioDBcore guidelines, along with the standards used within them; compiled in collaboration with 2012 NAR Database.



- Works to be a *coherent, curated and searchable registry* of **standards** for describing, reporting and sharing **experiments in life science, environmental, biomedical and biotechnological domains** – **collaborate with existing efforts and registries to**
 - **Develop** assessment **criteria** for usability and popularity of standards
 - **Associate** standards to **journals' and funders' data policies** and **databases**
- With the objective to help stakeholders to make informed decisions *e.g.*
 - funders: which standards or databases to use or recommend in their policies; identify efforts they have been or should be funded, or require integration
 - developers: facilitate selection, use and harmonization of standards, avoiding duplications



Thank you for your attention!

Simon Hodson

Executive Director CODATA

www.codata.org/blog

Email: execdir@codata.org

Twitter: [@simonhodson99](https://twitter.com/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



CODATA-ICSTI Task Group Data Citation, Standards and Practices

For Attribution

Workshop and Report:

http://bit.ly/for_attribution

Out of Cite, Out of Mind

http://bit.ly/out_of_cite

Joint Declaration of Data Citation Principles:

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

Background and Developments:

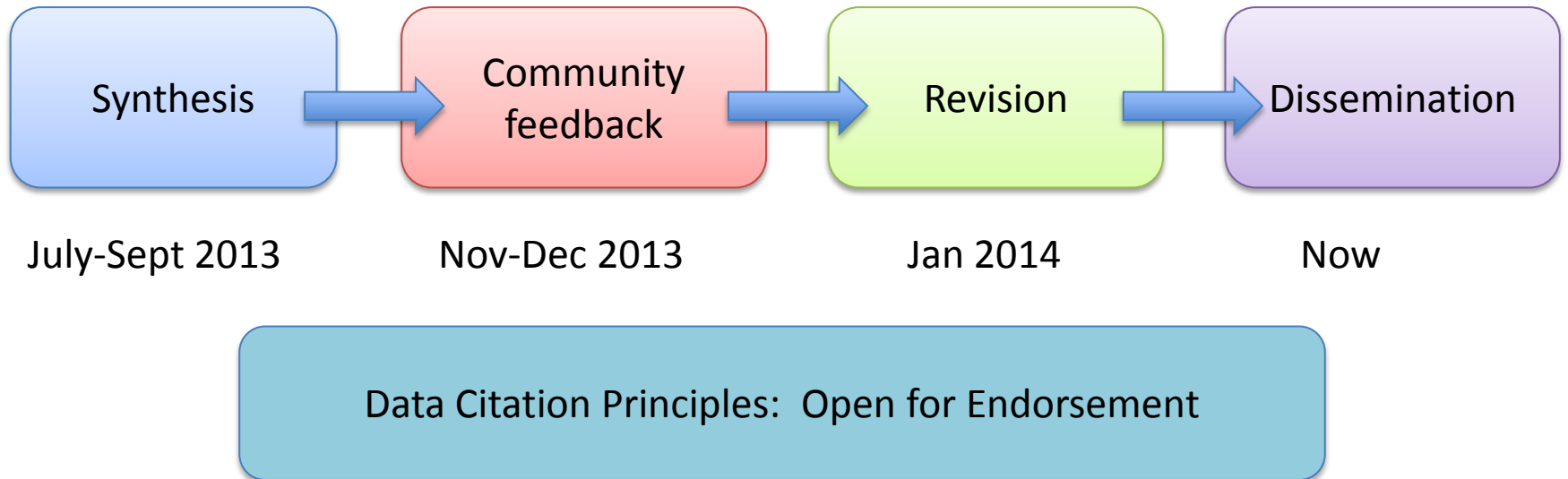
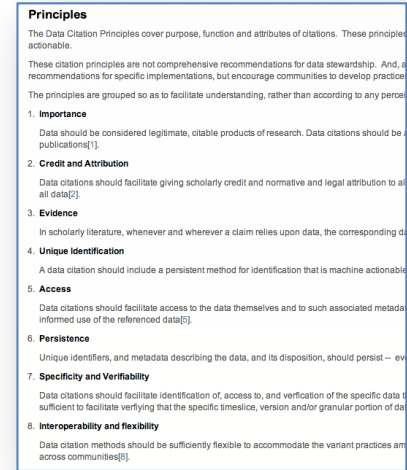
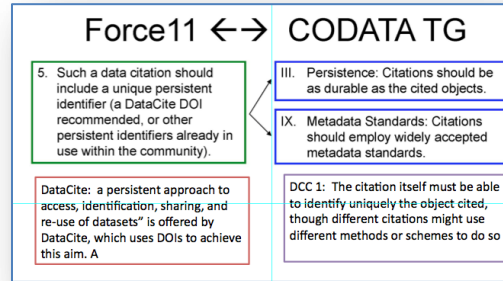
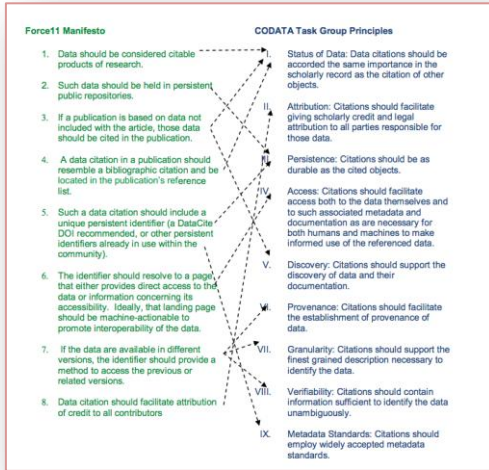
http://bit.ly/data_citation_principles



DC¹

Data Citation Principles

Process





Joint Data Citation Principles

Purpose

- 1. Importance.** Data should be considered legitimate, citable products of research. **Data citations should be accorded the same importance in the scholarly record as citations of other research objects, such as publications.**
- 2. Credit and attribution.** Data citations should facilitate giving scholarly **credit** and normative and legal **attribution** to all contributors to the data, recognizing that a single style or mechanism of attribution may not be applicable to all data.
- 3. Evidence.** In scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited.

Function

- 4. Unique Identification.** **A data citation should include a persistent method for identification that is machine-actionable, globally unique, and widely used by a community.**
- 5. Access.** Data citations should facilitate access to the data themselves and to such associated metadata, documentation, code, and other materials, as are necessary for both humans and machines to make informed use of the referenced data.



Joint Data Citation Principles

Attributes

- 6. Persistence.** Unique identifiers, and metadata describing the data and its disposition, should persist -- even beyond the lifespan of the data they describe.
- 7. Specificity and verifiability.** Data citations should facilitate identification of, access to, and verification of the specific data that support a claim. Citations or citation metadata should include information about provenance and fixity sufficient to facilitate verifying that the specific timeslice, version and/or granular portion of data retrieved subsequently is the same as was originally cited.
- 8. Interoperability and flexibility.** Data citation methods should be sufficiently flexible to accommodate the variant practices among communities, but should not differ so much that they compromise interoperability of data citation practices across communities [8].

Endorse the Data Citation Principles

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



Joint Declaration of Data Citation Principles (Overview); Slide: Sarah Callaghan

[Add Endorsement](#)

Joint Declaration of Data Citation Principles Post to Twitter

Individual Endorsements

120 Endorsements

First Name	Last Name	Affiliation	Endorsement Date
Alberto	Accomazz	NASA Astrophysics Data System	2014-02-27 16:21
Donat	Agosti	Plazi	2014-03-04 13:25
Micah	Altman	MIT	2014-02-27 09:56
Martin	Alvarez Espinar		2014-02-27 03:52
Eva	Amsen	F1000Research	2014-03-10 11:04
Roger	Barry	NSIDC/CIRES, Univ. of Colorado	2014-02-28 08:07
Rob	Baxter	EPCC, University of Edinburgh	2014-02-28

Organization Endorsements

40 Endorsements

Organization	Endorsement Date
BioMed Central	2014-03-21 01:30
CODATA	2014-02-27 09:59
CODATA-ICSTI Task Group on Data Citation Standards and Practices	2014-02-27 07:59
DANS	2014-03-07 05:07
DataCite	2014-03-26 04:50