



# Data Citation Standard: A Means to Support Data Sharing, Attribution, and Traceability

Ian McCallum, [Hans-Peter Plag](#), Steffen Fritz, & Stefano Nativi

*A Contribution of the EC-Funded EGIDA Project to the  
GEO Work Plan Task ID-03: Science and Technology in GEOSS*

Why is data citation important?

Who is involved?

Where do we stand?

What is GEO's Contribution?

# **Why is data citation important?**

# In an evidence-based society ...

Data have an increasing societal relevance.

More data, increased interoperability, increased integration of different sources

Data repositories are also expanding, offering long-term storage and access to archived data (i.e. PANGAEA, Dryad, Dataverse)

Dedicated data journals are appearing, which facilitate publishing of data (e.g. Earth System Science Data)

In post-normal science, high uncertainties are paired with high decision stakes (e.g., climate change, impacts of pollution, sustainability, ...)



A screenshot of a web browser displaying the homepage of Earth System Science Data (ESSD). The browser tab is titled "ESSD - Home" and the address bar shows "www.earth-system-science-data.net/home.html". The page has a blue header with the text "Earth System Science Data" and "The Data Publishing Journal". On the left, there is a navigation menu with links: Home, Online Library ESSD, Online Library ESSDD, Alerts &amp; RSS Feeds, General Information, Submission, Review, Production, and Subscription. On the right, there is a section titled "Earth System Science Data (ESSD)" with a "Scheduled Special Issues" button. Below this, it lists "Chief Editors: David Carlson &amp; Hans Pfeiffenberger" and describes the journal's open access and peer-review policies. At the bottom, it lists indexing services: ADS, Directory of Open Access Journals (DOAJ), Bodleian Library (UK), Deutsche Digitale Bibliothek (D), Library of Congress (USA), and Portico.

# Traceability, Sharing, Attribution, ...

Data citation is becoming highly relevant for traceability (IPCC)

Getting credit is an important incentive for researchers and data providers

Lack of data citation hampers data sharing

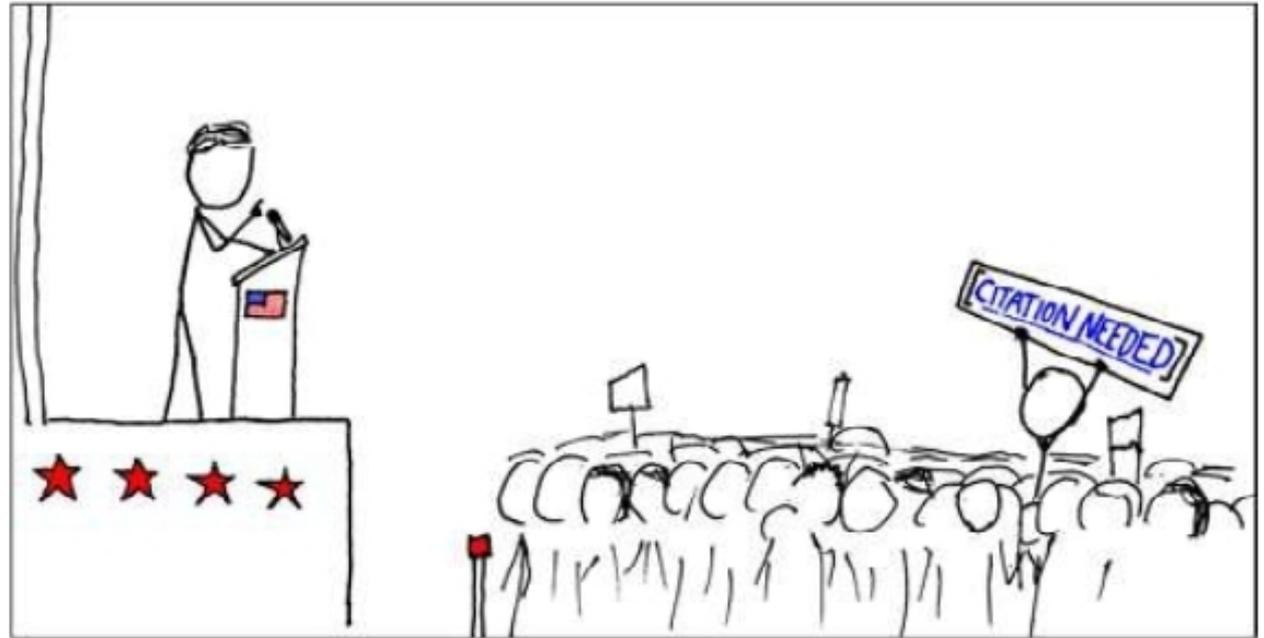


Fig. 1: There is a need for citation of data sources, not only for scientific papers. From John Wilbanks, Creative Commons, DataCite Annual Meeting 2011, Berkeley, August 25, 2011, see [www.slideshare.net/wilbanks/datacite-wilbanks](http://www.slideshare.net/wilbanks/datacite-wilbanks).

**BUT:**

While citation rules are well developed for papers, very few rules exist for data citation.

Why is data citation important?

**Who is involved?**

# An on-going international discussion ...

Organization	Contribution
International Polar Year (IPY)	The IPY developed a set of rules for data citation, which have been used both within and outside of the IPY context
ICSU/CODATA	Acknowledges the need for robust data citation and identified key issues to be addressed by citation rules. See <a href="http://www.codata.org/taskgroups/TGdatacitation/index.html">http://www.codata.org/taskgroups/TGdatacitation/index.html</a>
DATA-PASS	The Data Preservation Alliance for Social Sciences addresses data citation in the context of social sciences. See <a href="http://www.data-pass.org/">http://www.data-pass.org/</a>
ESIP Federation	Adapted the IPY rules and modified them into data citation rules for the ESIP Federation
DATA-CITE	Aims to support data access and re-usability through citation rules. See <a href="http://www.datacite.org/">http://www.datacite.org/</a>
U.S. National Academies	The Board on Research Data and Information of the U.S. National Academies is preparing a report on data citation. See <a href="http://sites.nationalacademies.org/PGA/brdi/PGA_063656">http://sites.nationalacademies.org/PGA/brdi/PGA_063656</a>
EGIDA	The project provided important input to the GEOSS Citation Standard V1.0. See <a href="http://www.egida-project.eu/">http://www.egida-project.eu/</a>

Recently:

- Thomson Reuters: Data Citation Index for Discovery of Global Data Sets (June 2012)
- GEO Data Sharing Group focuses on data citation

Why is data citation important?

Who is involved?

**Where do we stand?**

# Towards data citation ...

Data citation is far more complicated than citation of scientific publications as additional factors must be considered (i.e. versions, etc.).

Ball, A. & Duke, M. (2012). 'How to Cite Datasets and Link to Publications'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: [www.dcc.ac.uk/resources/how-guides](http://www.dcc.ac.uk/resources/how-guides).

## In this section

Briefing Papers

How-to Guides

Appraise & Select Data

**Cite Datasets**

Develop a Data Plan

License Research Data

Curation Reference Manual

Curation Lifecycle Model

Policy and legal

Data Management Plans

Tools

Case studies

Repository audit and assessment

Standards

Publications and presentations

Roles

Curation journals

## Cite Datasets and Link to Publications

This guide will help you create links between your academic publications and the underlying datasets, so that anyone viewing the publication will be able to locate the dataset and vice versa. It provides a working knowledge of the issues and challenges involved, and of how current approaches seek to address them. This guide should interest researchers and principal investigators working on data-led research, as well as the data repositories with which they work.

By Alex Ball (DCC) and Monica Duke (DCC)

Published: 18 October 2011

Last updated: 20 June 2012

Browse the guide below or [download the PDF](#)

Please cite as: Ball, A. & Duke, M. (2012). 'How to Cite Datasets and Link to Publications'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides>

### IDCC



The 8th International Digital Curation Conference (IDCC) will take place on 14-16 January in Amsterdam, The Netherlands. Its theme will be "Infrastructure, Intelligence, Innovation: driving the Data Science agenda".

[Read more](#)

## Related Information

## Useful links

# Towards data citation ...

Minimal set of parameters:

Field	Description
<b>Author</b>	The creator of the dataset
<b>Publication Date</b>	Whichever is the later of: the date the dataset was made available, the date all quality assurance procedures were completed, and the date the embargo period (if applicable) expired
<b>Title</b>	As well as the name of the cited resource itself, this may also include the name of a facility and the titles of the top collection and main parent sub-collection (if any) of which the dataset is a part
Edition	The level or stage of processing of the data, indicating how raw or refined the dataset is
Version	A number increased when the data changes, as the result of adding more data points or re-running a derivation process, for example
Feature name and URI	The name of an ISO 19101:2002 'feature' (e.g. GridSeries, ProfileSeries) and the URI identifying its standard definition, used to pick out a subset of the data.
Resource type	Examples: 'database', 'dataset'.
Publisher	The organisation either hosting the data or performing quality assurance.
Unique numeric fingerprint	A cryptographic hash of the data, used to ensure no changes have occurred since the citation
Identifier	An identifier for the data, according to a persistent scheme
<b>Location</b>	A persistent URL from which the dataset is available. Some identifier schemes provide these via an identifier resolver service

# Towards data citation ...

ICSU CODATA provides a comprehensive overview of **issues to be considered** in developing data citation rules:

[www.codata.org/taskgroups/TGdatacitation](http://www.codata.org/taskgroups/TGdatacitation).

## **A. Technical**

- Interoperability and Facilitation of Re-use
- Citation Formats
- Metadata
- Database Versioning

## **B. Scientific**

- Different disciplines may have disparate needs for granularity;
- Differences among disciplines that need to be addressed distinctly?

## **C. Institutional**

- What are the roles of the respective stakeholders?
- What are the implications for these stakeholders?
- Does this vary by discipline?

## **D. Financial**

- Lot of granularity can be cost-prohibitive.
- Must be accessible and its costs affordable by all necessary user communities.

# Towards data citation ...

ICSU CODATA provides a comprehensive overview of issues to be considered in developing data citation rules:

[www.codata.org/taskgroups/TGdatacitation](http://www.codata.org/taskgroups/TGdatacitation).

## **E. Sustainability**

## **F. Persistent Identifiers**

- e.g., use of the DOI (Digital Object Identifier) System.
- Use of DOI names for datasets is promoted by the not-for-profit DataCite consortium, which has registered over 600,000 datasets;
- However, significant differences between data and documents, that may make some aspects of the DOI system less attractive.

## **G. Legal Issues/Intellectual Property Rights**

- Any registry system must accommodate emerging intellectual property rights mechanisms, e.g. Creative Commons and Science Commons licensing, as well as traditional copyright law.

## **H. Socio-cultural and Community Norms**

- Develop a common basis and community of practice for recognizing and rewarding data work;

## **I. Other Issues will arise ...**

Why is data citation important?

Who is involved?

Where do we stand?

**What is GEO's Contribution?**

# The Group on Earth Observations (GEO) ...



- ... is an intergovernmental group;
- Initiated in 2003 by G8 in response to the World Summit on Sustainable Development, Johannesburg 2002;
- Has now more than 90 Member Countries;
- Has more than 60 Participating Organizations; ...

# ... addresses nine Societal Benefit Areas ...

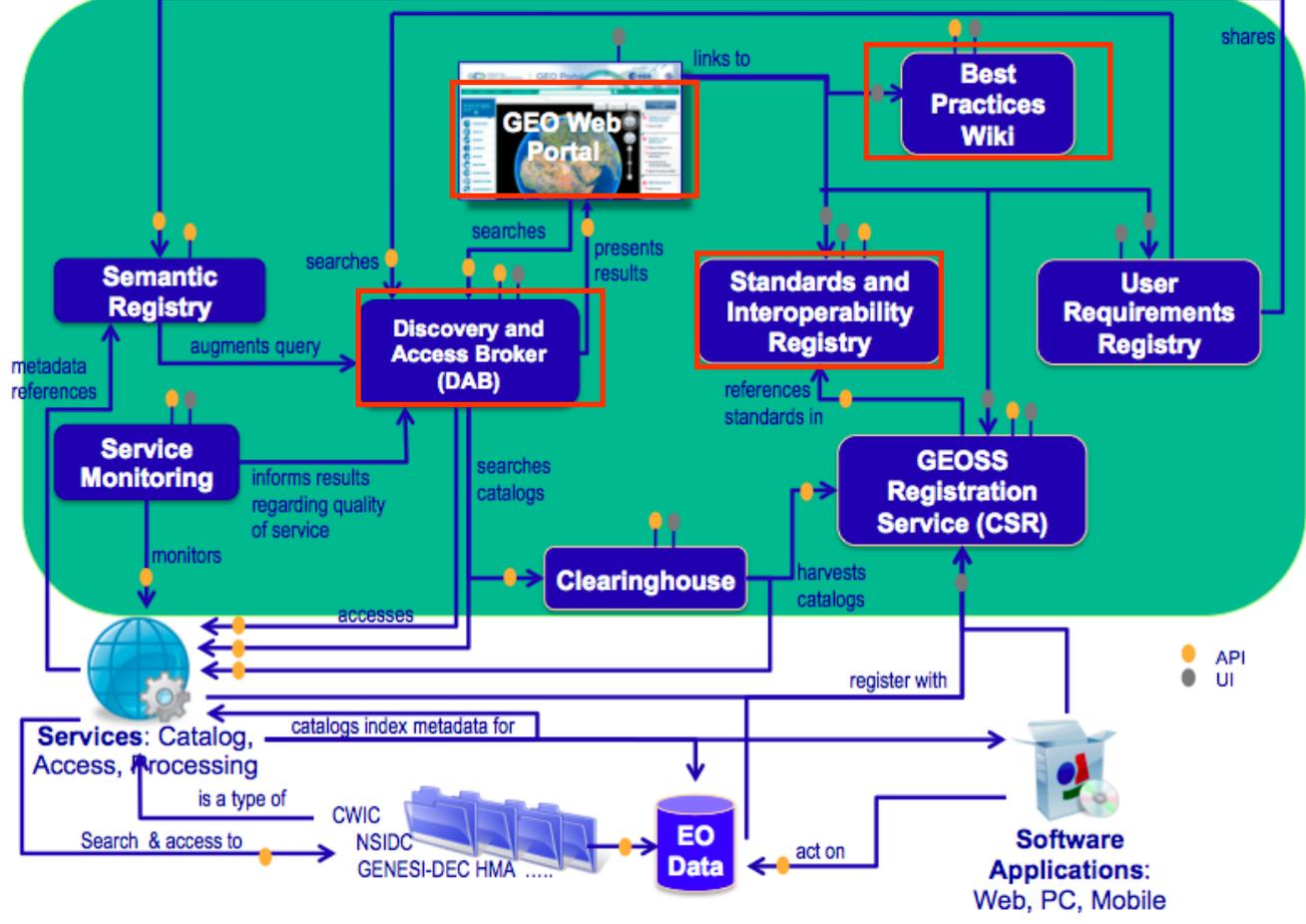
## THE GLOBAL EARTH OBSERVATION SYSTEM OF SYSTEMS



# ... implements GEOSS...



*GEOSS Infrastructure interactions VERSION GCI2-4B*



# ... and supports data citation ...

## GEOSS DATA CITATION GUIDELINES

The former GEO Work Plan Task ST-09-02 with support of the EGIDA Project has developed draft data citation guidelines.

This draft is based on guidelines developed by international groups.

This draft is under constant improvement as new aspects are addressed.



## GEOSS Data Citation Guidelines: Version 2.0

---

August 13, 2012

I. McCallum<sup>1</sup>, H.-P. Plag<sup>2</sup>, S. Fritz<sup>1</sup>

<sup>1</sup> International Institute for Applied Systems Analysis, ESM Program, Laxenburg, Austria

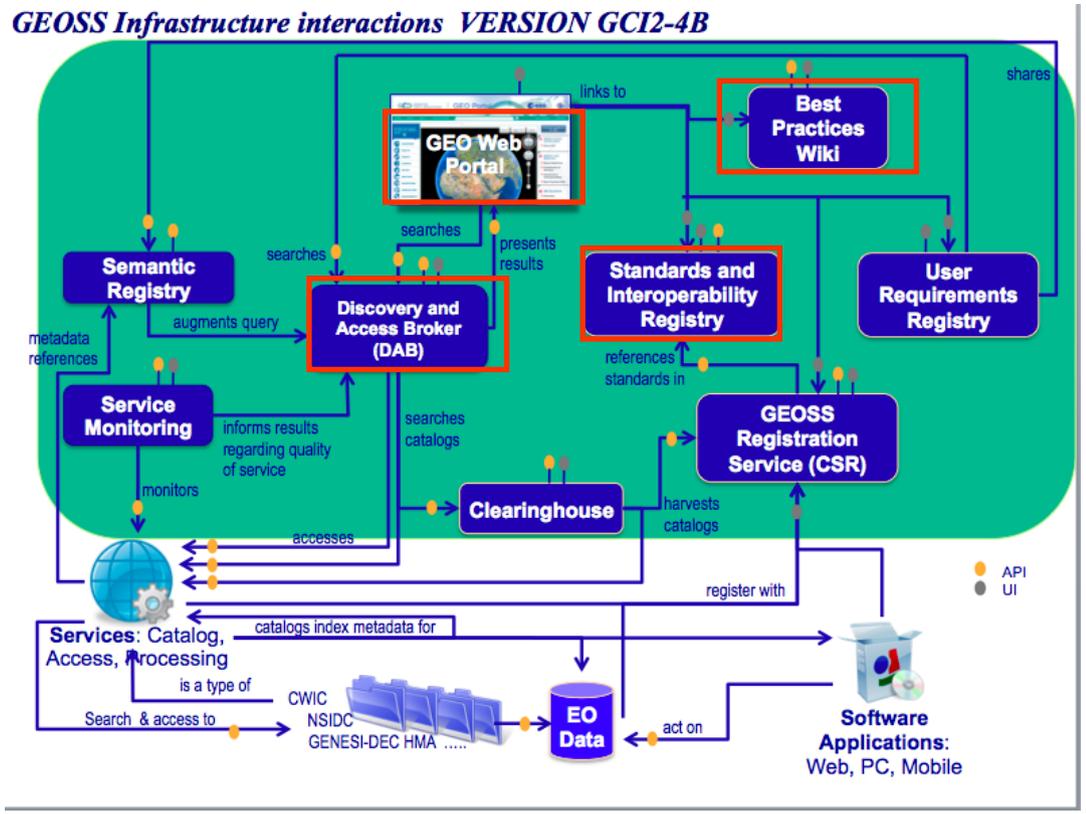
<sup>2</sup> Nevada Bureau of Mines and Geology and Seismological Laboratory, University of Nevada, Reno, Mail Stop 178, Reno, NV 89557, USA

# ... and supports data citation ...

The GEO Plenary supports a testbed implementation of the draft GEOSS Data Citation Guidelines.

Implementation of a data citation guidelines:

- Registration of the guidelines in the SIR;
- Best Practice in data citation in the BPW;
- Request for data citation on the GEO Portal and DAB;
- Provision of recommended citation.



Testbed will give experience with data citation and reveal issues currently not identified.