



# TOWARDS A DATA CITATION STANDARD FOR GEOSS

---

I. McCallum, H.-P. Plag & S. Fritz

# GIVING CREDIT TO DATA PRODUCERS

- An important incentive for researchers is recognition
- Citation rules well developed for papers, few rules exist for data citation
- Data citation however becoming highly relevant (IPCC)
- Supporting data citation a strong incentive for GEOSS
- Lack of data citation hampers data sharing

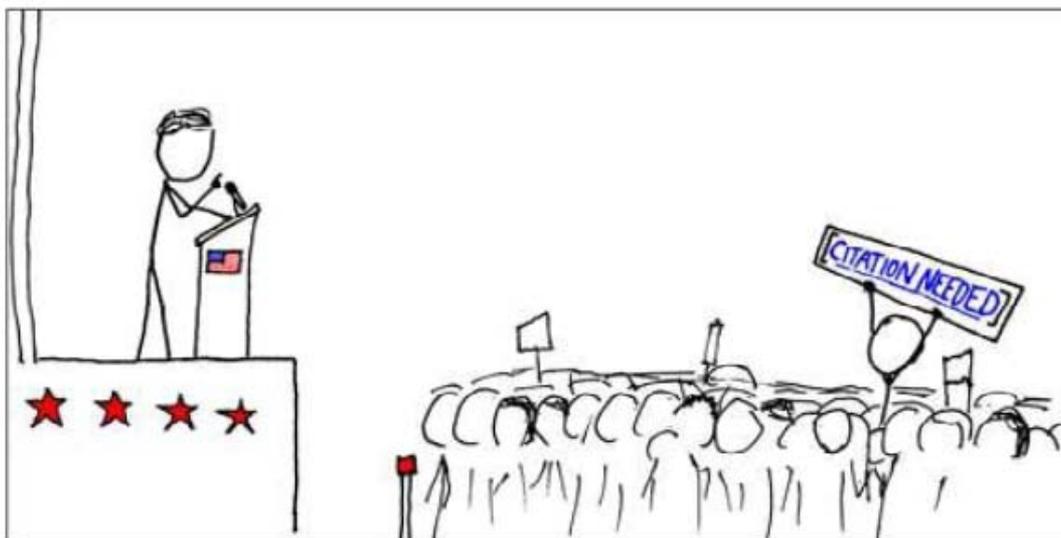


Fig. 1: There is a need for citation of data sources, not only for scientific papers. From John Wilbanks, Creative Commons, DataCite Annual Meeting 2011, Berkeley, August 25, 2011, see <http://www.slideshare.net/wilbanks/datacite-wilbanks>



# TOWARDS DATA CITATION

- The need for data citation rules is increasingly acknowledged and addressed by leading scientific organizations
- Several proposals for guidelines have emerged and a better understanding of the many issues at hand is evolving, but to date, no standard has been accepted.

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>



# TOWARDS DATA CITATION

---

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



# TOWARDS DATA CITATION

- Furthermore, dedicated data journals are appearing which aim to help researchers specifically publish data (e.g. Earth System Science Data).
- This journal aims to establish a new subject of publication: to publish data according to the conventional fashion of publishing articles, applying the established principles of quality assessment through peer-review to datasets.



# TOWARDS DATA CITATION

---

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

- Elements of a data citation: Ball, A. & Duke, M. 2011:

[www.dcc.ac.uk/resources/how-guides](http://www.dcc.ac.uk/resources/how-guides)



Field	Description
Author	The creator of the dataset
Publication Date	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
Title	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
Location	A persistent URL from which the dataset is available. Some identifier schemes provide these via an identifier resolver service

# GEOSS DATA CITATION STANDARD

---

- The former Science and Technology Committee (STC) of GEO with support of the EGIDA Project has developed a draft standard.
- This draft is based on guidelines developed by international groups.
- This draft standard is under constant improvement as new aspects are addressed.



# A TESTBED FOR DATA CITATION: GEOSS

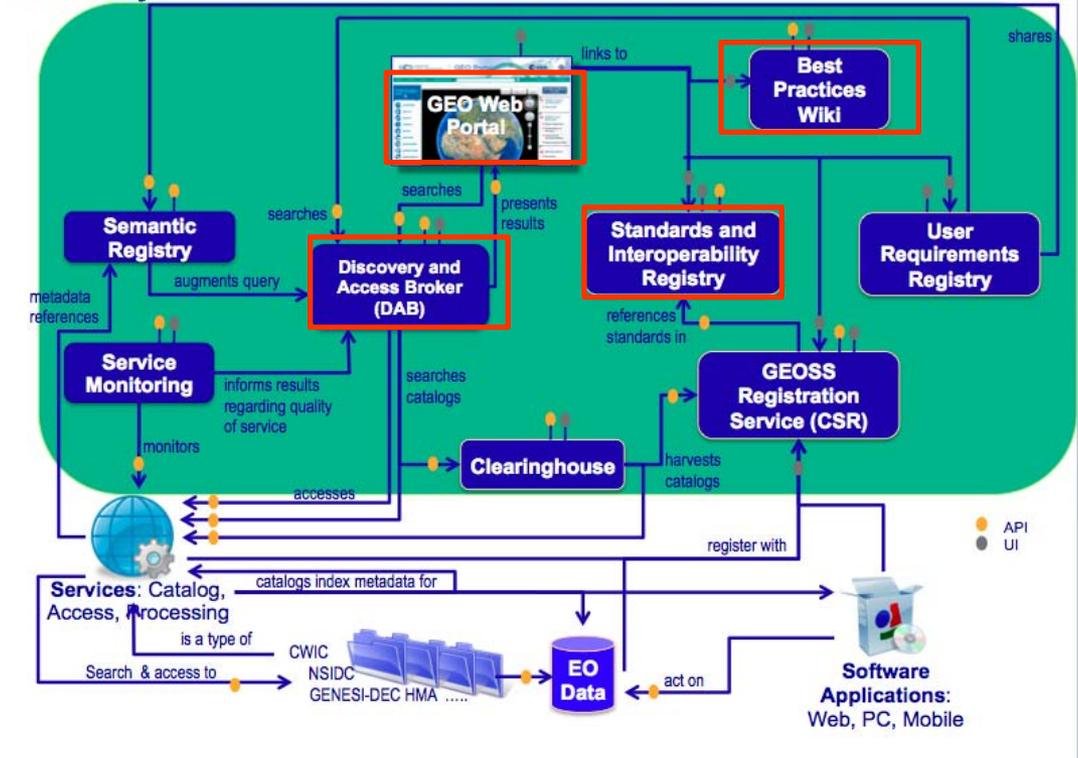
- The GEO Plenary supports a testbed implementation of the draft GEOSS Data Citation Standard.
- Currently, users of the GEO-Portal are not obliged or encouraged to cite data accessed through GEOSS – if at all, citation requirements come from the individual data providers.
- The testbed implementation of the draft GEOSS Data Citation Standard will rectify this situation.



# A TESTBED FOR DATA CITATION: GEOSS

- Implementation of a data citation standard and request will impact several components: the standard will be registered in the Standards and Interoperability Registry; a Best Practice in data citation will be developed and included in the Best Practices Wiki; a request for data citation will be included in the GEO Portal, and the Data Discovery and Access Broker will transmit information required for proper data citation.

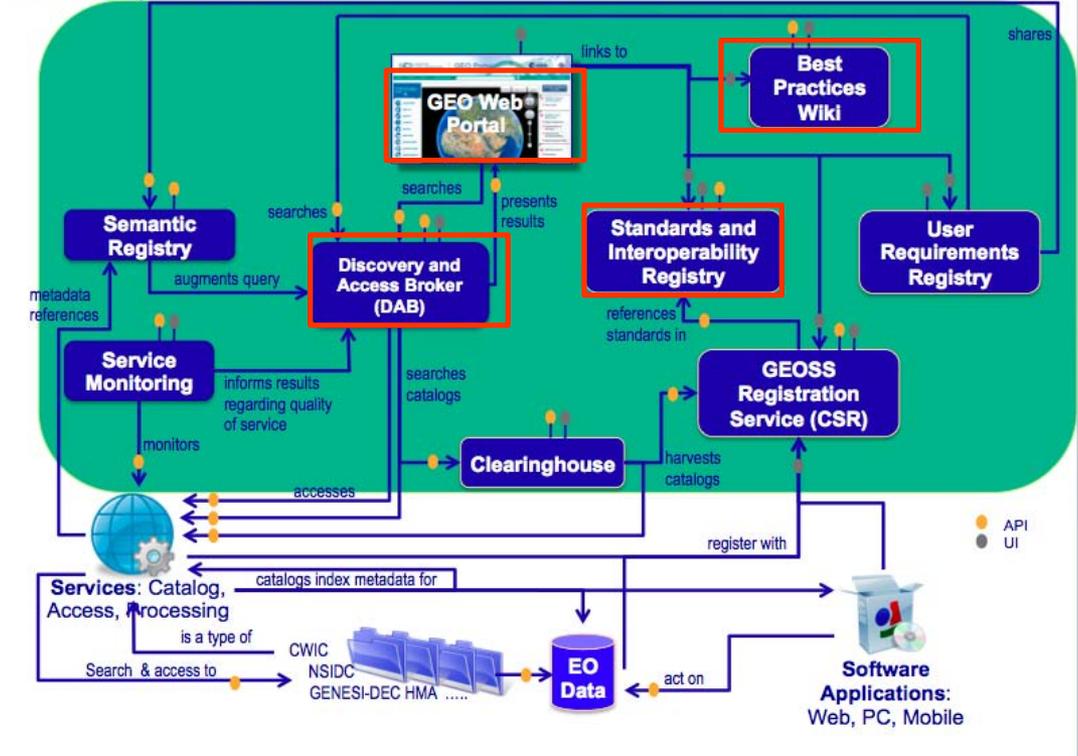
*GEOSS Infrastructure interactions VERSION GCI2-4B*



# A TESTBED FOR DATA CITATION: GEOSS

- The process of implementing the draft and iteratively improving it is led by the GEO Work Plan Task ID-03 "Science and Technology in GEOSS", and coordinated with the GEO working groups in charge of developing the GEOSS Common Infrastructure; other relevant GEO components (e.g., the GEO Data Sharing Task Force); and is aligned with the emerging international specifications concerning data citation. The experience with the testbed will be infused into the international discussion on data citation.

*GEOSS Infrastructure interactions VERSION GCI2-4B*



# ISSUES TO BE ADDRESSED

---

## 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.

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

<http://www.codata.org/taskgroups/TGdatacitation/index.html>



# ISSUES TO BE ADDRESSED

---

## 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 ...

GEO will consider most of the above issues that are currently not taken into account by V1.0 of the GEOSS data citation standard in V2.0, currently being developed by the GEO Task ID-03.



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

<http://www.codata.org/taskgroups/TGdatacitation/index.html>

# EGIDA Project

