



FP7 Project Nr 265124
Acronym: **EGIDA**
Project title: **Coordinating Earth and Environmental cross-disciplinary projects to promote GEOSS**
Theme: ENV.2010.4.1.1-1
Theme title: Supporting the integration of European and international R&D programmes in GEO

Project start date: 01 Sep 2010

D.3.1 Draft GEOSS Citation Standard

Title	D.3.1 Draft GEOSS Citation Standard
Creator	EGIDA WP3 / ST-09-02
Creation date	30/01/2011
Date of last revision	08/07/2011
Subject	GEO S&T, User Requirements, Best Practices,
Status	☐ Draft ☒ Final
Publisher	EGIDA WP3
Type	Text
Description	
Contributor	Mark Parsons, Hans-Peter Plag, Ian McCallum, ST-09-02 Task Team, EGIDA Advisory Board
Format	Doc
Source	
Rights	☐ Restricted ☒ Public
Identifier	EGIDA-Deliverable-template.docx
Language	En
Relation	WP3
Coverage	Not applicable

These are Dublin Core metadata elements. See for more details and examples <http://www.dublincore.org/>



Version history (to be removed before publication)

Version number	Date	Modified by (name and e-mail)	Comments



TABLE OF CONTENTS

Foreword	5
PART 1: Overview of the Process	6
PART 2: Draft Citation Standard Version 1.0 proposed to the STC	8
Preamble	8
1. INTRODUCTION	8
2. DATA CITATION.....	9
Author or Investigator	10
Publication Date	11
Title	12
Editor or Compiler	12
Publication Place	12
Publisher	13
Distributor or Associate Publisher	13
Distribution Medium and Location	13
Access Date	14
Data Within a Larger Work	14
3. IMPLEMENTATION	14
4. RELEVANT LINKS	15
5. REFERENCES.....	15
PART 3: Summary of Reviewers' Comments	17



EGIDA, Coordinating Earth and Environmental cross-disciplinary projects to promote GEOSS
FP7 Project nr 265124



Foreword

A key element of Task 3.1 of EGIDA (Get GEO/GEOSS better acknowledged) is the development of a GEOSS (Data) Citation Standard that will ensure that contributions to GEOSS services and data sets are traceable and fully credited, and that will increase the attractiveness of GEOSS for science and technology communities. The anticipated processes was that EGIDA together with GEO Work Plan Task ST-09-02 would propose to the Science and Technology Committee (STC) a GEOSS data citation standard consistent with the emerging international consensus on data citation rules. Subsequently, together with ST-09-02, EGIDA would support the STC of GEO in the process of getting the standard accepted by GEO and implemented for GEOSS.

The current document has three parts:

Part 1: Overview of the Process

Part 2: Draft Citation Standard Version 1.0 proposed to the STC

Part 3: Summary of two reviews of the Draft Standard

EGIDA Task 3.1 corresponds to the Activity 2a of the Road Map of the STC.



PART 1: Overview of the Process

During the EGIDA Kick-off meeting in September 2010 and a subsequent ST-09-02 Task Team meeting, invited experts in the international discussion of data citation identified a number of complicated issues that are currently under discussion in the international community and that would take considerable time to address in the frame of EGIDA. There was consensus that the GEOSS Citation Standard should be developed in two stages:

- Version 1.0 would reflect the current international consensus with respect to minimum requirements for data citation and fill the current gap of GEOSS not having any citation guidelines or recommendations. The goal was to develop and implement this standard as soon as possible.
- Version 2.0 would be made available after the open issues had been addressed together with relevant groups inside and outside of GEO.

In line with these decisions, a first draft of the GEOSS Citation Standard was prepared by EGIDA partners in cooperation with ST-09-02 within the first four project months. The draft was distributed to the ST-09-02 Task Team with a request for comments. The comments received were taken into account in the finalization of Draft Version 1.0. This version was presented to the STC at its Sixteenth meeting on April 14-15, 2011 in Sydney, Australia. In the discussion, the STC considered the draft a valuable contribution and a good starting point but also emphasized the necessity to develop a Version 2.0 in coordination with all other relevant groups in GEO. Several suggestions were made for improvements of the draft but no formal action on this was agreed.

Subsequently to the STC meeting, the draft was also presented to the participants of the Joint GEO/EGIDA Stakeholder Workshop in Bonn, Germany, on May 9-11, 2011. Furthermore, the draft was sent to two Advisory Board members for review. One of the reviewer is also a co-chair of the STC. For a summary of the reviewers' comments, see Part 3 of this document.

From the discussion at the STC meeting and the comments of the STC co-chair, it is expected that the STC Co-chairs will request changes of the draft before it is presented to the Executive Committee (ExCom) in GEO. EGIDA will provide resources to ST-09-02 to facilitate these changes and to get these changes iterated with the EGIDA partners committed to Task 3.1 and the ST-09-02 Task Team members. EGIDA will also support ST-09-02 in its effort to coordinate the current draft with other relevant groups in GEO, including but not limited to the Data Sharing Task Force and the Architecture and Data Committee.

After acceptance of the draft by the ExCom, EGIDA will provide guidance and support for the implementation of the version 1.0. This implementation will require some addition to the GCI and may also require additions to the meta data requested during registration of data and services in the GEOSS registries.

EGIDA will also provide resources to ST-09-02 for the development of Version 2.0 of the Citation Standard. The development of Version 2.0 will need considerable coordination with other GEO groups in order to ensure that the more demanding requirements will be acceptable within the GEO context and that the effort for implementation of this version will not exceed the limits of what can be requested from voluntary contributions. Issues such as digital object identification, traceability, attribution, tracking of versions, and quality-related aspects may require significant additional



metadata and efforts on the side of data providers. EGIDA together with ST-09-02 will ensure a broad basis for the development of Version 2.0 as a result of a collaborative effort.



PART 2: Draft Citation Standard Version 1.0 proposed to the STC

GEOSS Citation Standard

DRAFT Pending Acceptance by STC
Date: June 16, 2011

Preamble

Data citation is an open issue currently discussed by a number of international organizations, including the International Polar Year (IPY) and the Federation of Earth Science Information Partners (ESIP). Data citation is far more complicated than citation of scientific publication since Data sets differ in many aspects from standard scientific publications, which makes turns data citation far more complex than citation of scientific publications. For example, data sets generally are not locatable and attributable in the same way as scientific publications. Data sets often are not static, and they are mostly not peer-reviewed. All these issues need to be addressed before a comprehensive GEOSS Citation Standard can be developed. However, the IPY and ESIP have developed simple guidelines for data citation that can serve as a temporary GEOSS Citation Standard until a more comprehensive standard is available. The document provided here describes this first version of the GEOSS Citation Standard. The proposed standard focuses on data citation since citation of scientific publication are not considered an open issue. This initial standard includes guidelines for data citation that agree to a very large extent with the IPY and ESIP guidelines, with modifications only made where the specific application to GEOSS required this. This draft has been presented to the GEO Science and Technology Committee (STC) during its 16th Meeting on April 14-15, 2011 in Sydney, Australia. The STC will decide on the further handling of the draft within GEO.

The development of a more comprehensive GEOSS Citation Standard is the next step. This process will have to address questions such as data identification including version tracking, quality control, and attribution. This process needs to be coordinated with the GEO working groups in charge of developing the GEOSS Common Infrastructure (e.g., ADC, SIF, DSTF, GCI-CT), which needs to be able to support the requirements of a future GEOSS Citation Standard. Coordination is also required with similar activities initiated by other groups within GEO, such as the Data Sharing Task Force is necessary in order to avoid duplication of work. Furthermore, it will be important to develop a standard that is in line with the practice emerging in the international discussion led by ESIP.

The STC has asked ST-09-02 to develop a Version 2 of the Citation Standard that addresses most of the currently open issues; is coordinated with the development of the GEOSS Common Infrastructure; and is consistent with the development of data citation by other relevant organizations. The ST-09-02 Task Team will develop this version in close cooperation with the EGIDA project.

Note: The draft Citation Standard has been extracted from a draft deliverable of the EGIDA Project prepared by Ian McCallum, and minor modifications and additions were made by ST-09-02 Task team members and discussed during ST-09-02 Telecons.

1. INTRODUCTION



"Currently a person who publishes really good data is given less credit than someone who publishes a bad paper" (AGU Town Hall Meeting, 2009)

At present, GEOSS is lacking a data citation standard. This is symptomatic of the entire field of geo-spatial data, with few exceptions. A wide variety of organizations and projects support the concept of data citation (e.g., IPY, PANGAEA, NASA DAACS, USGS, NOAA National Data Centers), however no standard has emerged. Users accessing the GEO-Portal are not obliged to cite data from the GEO-Portal – if at all, citation requirements come from the individual data providers. This naturally leads to an at-best non-standard form of data citation or in the worst case, no data citation at all. Poor or no data citation is bad for both the scientist who has produced the data and receives no credit, and for the scientist who wishes to obtain data (i.e., from a scientific publication) but can not locate them owing to a lack of information.

In the scientific community, recognition and renown are important currencies. In order to increase the attractiveness of GEO and GEOSS for scientists, their contributions must be visibly acknowledged when others use it to their benefit. A key element for acknowledgment is a proper GEOSS citation standard, which not only covers published documents but also ensures that data used to create new products is acknowledged and appropriately credited. Therefore, a GEOSS data citation standard needs to be developed and implemented. To be effective, the GEOSS citation standard must aim for compliance/compatibility with international specifications concerning data citation.

GEOSS has a unique opportunity/responsibility to the scientific community that relies on spatial data to help establish such a data citation standard. By encouraging proper citation of data sets, data providers and publishers receive appropriate credit for their efforts, the perception of data management as a discipline improves, and it is easier to track the use and impact of the data ([International Polar Year](#)). The scientific community should recognize the professional value of such activities by endorsing the concept of publication of data, to be credited and cited like the products of any other scientific activity, and encouraging peer-review of such publications ([AGU Position Statement](#)).

The goal of this document is to provide a data citation standard for all GEOSS related datasets consistent with recommendation of relevant international organizations. The guidelines presented herein were adapted from the International Polar Year (IPY) [Data Information Service](#), which was a synthesis of the different approaches agreed to by many international data centers. These organizations acknowledge that the current guidelines still leave a number of issues uncovered (e.g., unique data object identifiers, attribution, quality information). It is recommended that GEO addresses these issues together with the Federation of Earth Science Information Partners (ESIP) and other relevant organizations in a next version of the GEOSS Citation Standard. The development of this more comprehensive Standard will require coordination with the ADC to ensure that the GEOSS Common Infrastructure (GCI) can enable the implementation of the future Standard.

2. DATA CITATION

To recognize the valuable role of data providers (and scientists who collect or prepare data) and to facilitate repeatability of IPY experiments in keeping with the scientific method, users of IPY data must formally acknowledge data authors (contributors) and sources. Where possible, this acknowledgment should take the form of a formal citation, such as when citing a book or journal article. Journals should require the formal citation of data used in articles they publish. Where formal citation is not possible, such as with some medical and social science data, ethical policies for data



collection and data use are encouraged, building upon existing models such as Article 8(j) of the 1992 Convention on Biological Diversity ([IPY Data Policy](#))

The following guidelines were developed by the IPY. These guidelines were adapted from internal guidelines used by the National Snow and Ice Data Center, which has encouraged formal data citation for more than a decade.

In general, data sets should be cited like books. Used here is the author-date system described in Chicago Manual of Style, 15th Edition. When users cite data, they need to use the style dictated by their publishers, but by providing an example, data publishers can give users all the important elements they should include in their citations of data sets. *So long as the mandatory fields are complete, the final publisher style is just semantics.*

An example of a citation in the author-date system is:

Algire, G. H., and F. T. Legallais. 1948. Biology of Melanomas. ed. R. W. Miner. New York: New York Academy of Sciences.

As seen in this example, the elements of the citation in order are:

1. Author(s),
2. Date,
3. Title,
4. Editor,
5. Place of Publication,
6. Publisher.

All these elements are common in data set citations, but other elements, as described below, are commonly used as well. Data publishers (e.g. data centers) have a responsibility to work with data providers and science teams to develop the actual content of the citation.

The citation should include the following elements as appropriate. Although this is shown as a literary citation, most of the elements are captured in standard metadata. The "Citation Information" section of the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) (FGDC-STD-001-1998) is given as reference:

Author or Investigator

This is the individual(s) whose intellectual work, such as a particular field experiment or algorithm, led to the creation of the data set.

Oberbauer, S. 2000. Ecosystem carbon fluxes, Toolik Lake, Alaska 1995. Boulder, Colorado USA: National Snow and Ice Data Center. Data set accessed 2008-05-14 at <http://nsidc.org/data/arcss006.html>.

A particular group or organization may sometimes be the author.

Arctic Climatology Project. 2000. Environmental Working Group Arctic meteorology and climate atlas. Edited by F. Fetterer and V. Radionov. Boulder, Colorado USA: National Snow and Ice Data Center. CD-ROM.



If the data set is a collection of several smaller, independent data sets, the individual data sets would have their own specific citations with author, but the whole collection would not have an author. The collection would likely have an editor or compiler, though.

Cross, M. compiler. 1997. Greenland summit ice cores. Boulder, Colorado USA: National Snow and Ice Data Center in association with the World Data Center A for Paleoclimatology at NOAA-NGDC, and the Institute of Arctic and Alpine Research. CD-ROM.

Publication Date

For a completed data set, the publication date is simply the year of release.

Helmig, D. 2004 Vertical Boundary Layer Profiles for Ozone and Meteorological Parameters at Summit, Greenland, 2000. Boulder, Colorado USA: National Snow and Ice Data Center. Data set accessed 2008-05-14 at <http://nsidc.org/data/arcss100.html>.

For a data set that is updated infrequently or on an irregular basis, list the first year of publication followed by "updated" with the current update information. This is appropriate when the title or version of the data set does not change, the data are simply updated.

Osterkamp, T. 1999, updated 2001 Daily air and active layer temperatures from permafrost observatories in Alaska, 1986-2001. Boulder, Colorado USA: National Snow and Ice Data Center. Data set accessed 2008-05-14 at <http://nsidc.org/data/arcss106.html>.

For an ongoing data set that is updated on a regular or continual basis, list the first year of publication followed by the last update. Updates could occur annually or more frequently.

Maslanik, J. and J. Stroeve. 1999, updated quarterly. DMSP SSM/I daily polar gridded brightness temperatures, Jan. Boulder, Colorado USA: National Snow and Ice Data Center. Data set accessed 2008-05-14 at <http://nsidc.org/data/nsidc-0001.html>.

Hall, Dorothy K., George A. Riggs, and Vincent V. Salomonson. 2007, updated daily. MODIS/Aqua Sea Ice Extent 5-Min L2 swath 1km V005, Oct. 2007–Apr. 2008. Boulder, Colorado USA: National Snow and Ice Data Center. Data set accessed 2008-05-14 at <http://nsidc.org/data/myd29v5.html>.

A note on updates vs. new versions:

Ongoing updates to a time series do change the content of the data set, but they do not typically constitute a new version or edition of a data set. New versions typically reflect changes in sampling protocols, algorithms, quality control processes, etc. Both a new version and an update may be reflected in the publication date. The title should indicate the new version.

Hall, Dorothy K., George A. Riggs, and Vincent V. Salomonson. 2006, updated daily. MODIS/Terra snow cover Extent 5-Min L2 swath 1km V005, Oct. 2007–Apr. 2008. Boulder, Colorado USA: National Snow and Ice Data Center. Data set accessed 2008-05-14 at <http://nsidc.org/data/myd29v5.html>.

Hall, D. K., G. A. Riggs, and V. V. Salomonson. 2000, updated daily. MODIS/Terra snow cover 5-Min L2 swath 500m V004, Oct. 2007–Apr. 2008. Boulder, Colorado USA: National Snow and Ice Data Center. Data set accessed 2008-05-14 at <http://nsidc.org/data/myd29v5.html>.



If a particular version of a time series is discontinued, it is appropriate to indicate when the final update occurred.

Hall, D. K., G. A. Riggs, and V. V. Salomonson. 2000, updated 2002. MODIS/Terra snow cover 5-Min L2 swath 500m V003, Jan. 2001–Apr. 2001. Boulder, Colorado USA: National Snow and Ice Data Center. Data set accessed 2008-05-14 at <http://nsidc.org/data/myd29v5.html>.

Title

This is the formal title of the data set. It may also include version or edition information.

Liu, H., K. Jezek, B. Li, and Z. Zhao. 2001. Radarsat Antarctic Mapping Project digital elevation model version 2. Boulder, CO: National Snow and Ice Data Center. Data set accessed 2008-05-14 at <http://nsidc.org/data/nsidc-0082.html>.

Dates Used

For time series, especially continually updated time series, indicate which dates of data were used. Note this is distinct from the publication date.

Hall, Dorothy K., George A. Riggs, and Vincent V. Salomonson. 2006, updated daily. MODIS/Terra snow cover Extent 5-Min L2 swath 1km V005, Oct. 2007–Apr. 2008. Boulder, Colorado USA: National Snow and Ice Data Center. Data set accessed 2008-05-14 at <http://nsidc.org/data/myd29v5.html>.

Editor or Compiler

An editor is the person or team who is responsible for creating a value-added and possibly quality-controlled product from the data. In cases where there is minimal scientific or technical input, yet still substantial effort in compiling the product, the person may be more correctly cited as a compiler. Editors and compilers may often be responsible for a larger work that includes an individual author's data set. Occasionally, there may be both a compiler and editor. Some products will have neither.

Armstrong, R., J. Francis, J. Key, J. Maslanik, T. Scambos, and A. Schweiger. 1998. Polar Pathfinder sampler: Combined AVHRR, SMMR-SSM/I, and TOVS time series and full-resolution samples. Compiled by S. Khalsa. Boulder, CO, USA: National Snow and Ice Data Center. CD-ROM.

Cline, D., R. Armstrong, R. Davis, K. Elder, and G. Liston. 2002, Updated July 2004. CLPX-Ground: ISA snow pit measurements. Edited by M. Parsons and M. J. Brodzik. Boulder, CO: National Snow and Ice Data Center. Data set accessed 2008-05-14 at <http://nsidc.org/data/nsidc-0176.html>.

Bockheim, J. 2003. "University of Wisconsin Antarctic Soils Database". In International Permafrost Association Standing Committee on Data Information and Communication (comp.). 2003. Circumpolar Active-Layer Permafrost System, Version 2.0. Edited by M. Parsons and T. Zhang. Boulder, CO: National Snow and Ice Data Center/World Data Center for Glaciology. CD-ROM.

When there is an editor or compiler but no author, the editor is listed first.

Publication Place

This is the city, state (when necessary), and country of the publisher.



Cavalieri, D., C. Parkinson, P. Gloersen, and H. J. Zwally. 1996, updated 2006. Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I passive microwave data, March 2002–Sept. 2003. Boulder, Colorado USA: National Snow and Ice Data Center. Data set accessed 2008-05-14 at <http://nsidc.org/data/nsidc-0051.html>.

Publisher

The publisher is whoever published the data set. A publisher often has an implied responsibility for stewardship of the data set. This is usually a data center and is written immediately after the place.

Cavalieri, D., C. Parkinson, P. Gloersen, and H. J. Zwally. 1996, updated 2006. Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I passive microwave data, March 2002–Sept. 2003. Boulder, Colorado USA: National Snow and Ice Data Center. Data set accessed 2008-05-14 at <http://nsidc.org/data/nsidc-0051.html>.

Distributor or Associate Publisher

This field should be used only when it differs from the publisher, i.e. rarely. Its listing should be written in the same manner as that of publisher. Sometimes NSIDC acts as a simple distributor; sometimes we are an associate publisher; sometimes others are associate publishers.

Environmental Working Group. 2000. Environmental Working Group: Joint U.S.-Russian Arctic sea ice atlas. Ann Arbor, MI: Environmental Research Institute of Michigan; distributed by the National Snow and Ice Data Center. CD-ROM.

Cross, M. compiler. 1997. Greenland summit ice cores. Boulder, CO: National Snow and Ice Data Center in association with the World Data Center A for Paleoclimatology at NOAA-NGDC, and the Institute of Arctic and Alpine Research. CD-ROM.

Distribution Medium and Location

If there is one fixed medium, list it. For example, CD-ROM, DVD.

International Permafrost Association Standing Committee on Data Information and Communication (comp.). 2003. Circumpolar Active-Layer Permafrost System, Version 2.0. Edited by M. Parsons and T. Zhang. Boulder, CO: National Snow and Ice Data Center/World Data Center for Glaciology. CD-ROM.

If data are available over the internet or through multiple digital media options it is best to include a reference to the location of the data. Often this is through a standard URL.

Cavalieri, D., C. Parkinson, P. Gloersen, and H. J. Zwally. 1996, updated 2006. Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I passive microwave data, March 2002–Sept. 2003. Boulder, Colorado USA: National Snow and Ice Data Center. Data set accessed 2008-05-14 at <http://nsidc.org/data/nsidc-0051.html>.

Ideally, a persistent identifier such as a Digital Object Identifier should be used.

König-Langlo, Gert and Hatwig Gernandt. 2006. Compilation of radiosonde data from the Antarctic Georg-Forster station of the German Democratic Republic from 1985 to 1992. Bremerhaven,



Germany: Alfred Wegener Institute for Polar and Marine Research Data set accessed 2008-05-22.
doi:10.1594/PANGAEA.547983

Access Date

Because data can be dynamic and changeable in ways that are not always reflected in publication dates and versions, it is important to indicate when on-line data were accessed. It is not necessary to indicate an access date for a fixed medium like a DVD.

Cavalieri, D., C. Parkinson, P. Gloersen, and H. J. Zwally. 1996, updated 2006. Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I passive microwave data, March 2002–Sept. 2003. Boulder, Colorado USA: National Snow and Ice Data Center. Data set accessed 2008-05-14 at <http://nsidc.org/data/nsidc-0051.html>.

Data Within a Larger Work

A particular data set may be part of a compilation, in which case it is appropriate to cite the data set somewhat like a chapter in an edited volume.

Bockheim, J. 2003. "University of Wisconsin Antarctic Soils Database". In International Permafrost Association Standing Committee on Data Information and Communication (comp.). 2003. Circumpolar Active-Layer Permafrost System, Version 2.0. Edited by M. Parsons and T. Zhang. Boulder, CO: National Snow and Ice Data Center/World Data Center for Glaciology. CD-ROM.

Increasingly, publishers are allowing data supplements to be published along with peer-reviewed research papers. When using the data supplement one need only cite the parent reference. For example, when using the data at doi:10.1594/PANGAEA.476007, the following reference is appropriate.

Stein, Ruediger, Bettina Boucsein, and Hanno Meyer. 2006. "Anoxia and high primary production in the Paleogene central Arctic Ocean: first detailed records from Lomonosov Ridge." *Geophysical Research Letters*, 33: L18606. doi:10.1029/2006GL026776.

3. IMPLEMENTATION

It is recommended that the implementation of the GEOSS Citation Standard should proceed along the following lines:

1. Use IPY data citation Guidelines to implement a first version of the GEOSS Citation Standard.
2. Address issues such as the unique identification of digital objects, versions of datasets, data stability, attribution, etc. in a next version of the GEOSS Citation Standard.
3. Implement the concepts in the GEOSS Common Infrastructure.
4. Lobby publishers to require data citation.

1. Owing to the fact that the basic IPY Guidelines generally reflect the current consensus on data citation and are relatively simple in their structure, it is proposed that GEOSS applies these guidelines in their current form (as described in the previous section). It is understood that these guidelines are dynamic and would be modified or improved over time to suite the specific needs of GEOSS.



In addition, it would be beneficial if GEOSS could automatically generate a standard data citation for each GEOSS dataset as per literature citation databases - i.e. export the citation in Text, RefWorks, RIS format (e.g. EndNote), BibTeX, comma separated. This would likely facilitate improved referencing. It is worth noting that e.g. EndNote does not contain a default reference type for data – thus the recommendation to cite data using the format of a book avoids this problem.

2. Options for data identification include Digital Object Identifiers (DOI), which are used to provide current information, including where data (or information about data) can be found on the Internet. Information about a digital object may change over time, including where to find it, but its DOI name will not change. Unfortunately, at this point no one identifier scheme has emerged to meet all the needs of scientific data publications (Parsons et al. 2010). This needs to be explored further. A UUID (Universal Unique Identifier) is a 128-bit number used to uniquely identify some object or entity on the Internet. A UUID addresses the need to uniquely and unambiguously identify a particular data set or subset no matter which copy a user has (Parsons et al. 2010). This seems promising in terms of uniquely identifying data, but should also be explored further.

3. The GCI will have to be designed to provide the attributes required in a future GEOSS Citation Standard. This will require a coordination between the development of this standard and the implementation of, e.g., object identifiers, UUID, and other relevant attributes in the GCI.

4. A long term goal of GEOSS should be to actively lobby publishers to require data citation. Ultimately journal editors and reviewers need to be more rigorous in demanding that authors accurately cite the data they use in their research (Parsons et al. 2010). Along these lines of thinking, one journal already exists which is dedicated to publishing data: [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.

In summary, it is proposed here that GEOSS proceed as a first step with establishing a data citation standard along the lines of the IPY citation standard (Parsons 2010). This is in part possible as the standards are assumed to be dynamic, allowing for improvements over time. The majority of open issues surrounding data citation, e.g. persistent identifiers, should be dealt with in the near future.

4. RELEVANT LINKS

<http://libraries.mit.edu/guides/subjects/data/access/citing.html>

<http://www.dlib.org/dlib/march07/altman/03altman.html>

<http://ands.org.au/guides/data-citation-awareness.html>

http://wiki.esipfed.org/index.php/Interagency_Data_Stewardship/Citations

5. REFERENCES

AGU Town Hall meeting, 2009. See

http://wiki.esipfed.org/index.php/Interagency_Data_Stewardship/2009AGUTownHall.

Parsons, M.A., 2010. Presentation at ST-09-02 Task Team Meeting, Rome, Italy, October 30, 2010. See <http://www.geo-tasks.org/st0902/meetings> and http://www.geo-tasks.org/st0902/meetings/20100930_Rome/parsons_citation_geoss.pdf.



Parsons, M.A., Duerr, R., & Minster, J.B., 2010. Data citation and peer review. *Eos , Trans. Am. Geophys. Union*, **91**, 297-298



PART 3: Summary of Reviewers' Comments

Foreword

Ver 1.0 of the draft citation standard was sent to two members of the EGIDA advisory board. They each reviewed the document, providing comments for consideration. Those comments are presented below largely in their original format. Minor editorial comments have been removed.

Joern Hoffmann (Reviewer 1)

Thank you for the opportunity to comment on the draft version 1.0 of the citation standard document.

I find that it is a good start and I support the notion that this should move forward even when we expect there to be further evolutions in the future. I do want to recommend that you consider strengthening the document in a number of ways before presenting it to the Joint Committees:

General

First, the document does not motivate why GEO is a good framework to implement this citation standard in. Nor does it reference the S&T Roadmap, which would help to explain this. Both should be done: The document already explains why data citation is important, mainly quoting from IPY texts. In most places where "GEOSS citation standard" is mentioned, you are really just talking about a citation standard. So, I strongly recommend that the Preamble focuses on why GEO is the right framework for this development and how a GEOSS Citation Standard is pursuing its objectives.

Second, surprisingly (to me and several others I have talked to), the proposed standard does not include DOIs. I merely notes in passing, that the DOI may not be sufficient. This appraisal is not shared by several experts I have talked to. I have not been able to understand the arguments of both sides at this point, but would suggest that the document takes some room to explain the reasons for not including DOIs. While the Parsons reference is cited to support the inadequacy of the DOI, this question is central enough to the proposal to reward spending some words on the issue.

Finally, the document would also be strengthened by explaining, very pragmatically, what is meant by "implementing" the citation standard. My understanding is very simple indeed:

1. put this document on the web and make it known
 2. add functionality to the GEO Portal that (a) reminds people accessing GEOSS Data to cite it, (b) propose how to cite it, based on metadata provided (c) allow exporting metadata files to be used in citation software (d) collect the metadata needed for citation during the registration process and (e) possibly indicate along with the GEOSS Components if they have all necessary metadata fields filled.
- Is this correct? I think it would be useful to put a picture in people's heads.

Editorial Comments:

1. The bulk of the document is rather unsurprising explanations of what an author is, what a date is, etc. I suggest to separate this part from the rest to allow readers to quickly assess the main point of the proposal. The explanations could be placed in an annex.

Herbert Haubold (Reviewer 2)

General

I think this is a good and valuable piece of work, and I agree that proper citation of data sets is essential. In Environmental institutions we are often faced with the problem that we do not know exactly where particular data sets come from. This is especially true for data that are (sometimes



repeatedly) reworked, value added so to say. Here data users need to know each step along the chain and especially the original data source.

Exact tracing of data sets is essential to enable a quality check or to know that such a check is not needed because the data are trustworthy. Moreover, when data users (e.g. in policy) receive information drawn from data (we keep saying users want information not data, which is not exactly correct by the way), it would be essential to enable them to exactly trace all data sources which is possible only when cited correctly. So I see a double benefit of good data citation, for the data providers to be recognized as you say below, and for data users as described above. That latter reason might be added to the paper.

More specific comments below

Maybe we should rather speak of data citation requirements than of a standard? See below.

Even if a publication contains data, it is a very different story anyway. Reading this, to me the issues is that a paper by itself gives you all you need to do a proper citation, it is in there, and with data sets this is not always the case, so the first step would be for all data sets to provide the meta data that the citation will draw from, right?

Another issue is that without this credit, the person (for data in most cases this will not be a scientist, though) has not much of an incentive to put their data in the system.

As I just said I agree that it is essential to enable proper recognition of data providers, but from this I do not see why there must be a particular GEOSS standard.

Because GEOSS is a huge network and a large variety of data sets in the GCI, it is an ideal place to start developing THE citation standard, which would be good for any data citation, no matter whether within GEOSS or not. This would make your approach even more important because also actors not involved in GEOSS would potentially benefit.

Data cannot just be peer-reviewed such as a publication for two reasons: most environmental, statistical etc. data are gathered owing to legal requirements, conventions and the like, so there are given standards and you either fulfill them or not. Data in scientific publications cannot be reviewed without detailed information on how they were gathered and other data sets to compare them with. Especially the latter you do not have in many cases.

This is very good and it is, sorry, a good reason why we do not need a particular GEOSS data citation standard. Papers are cited in many different ways, yet you always (well almost) find them. The issue is particular bits of information you need, that is why I think we may not speak of a data citation standard, but of data citation requirements.

I know I am biased here, but for priority environmental data you also need the legal frame, such as a directive to which the data gathering responds. Just an example:

http://rod.eionet.europa.eu/obligations;jsessionid=A629F92FA344AEC90AB77C6840C53984?country=3&id=&filter=GO&issue=15&client=-1&_sourcePage=jgx-1UqSh7oBvly8cPRxXJoVXeDNhenyPzrrjzfXHko%3D&_fp=O2dNdyoTtnw3YT8-iFUgxXI-r4ZwDmhXbL_PO7oyM-a760Mo9vIMLkq2uSNjSmu

For the future, also real-time (or near-real time) should be considered, such as the upcoming SEIS data.



In such a case the original data sources would be important, see remark at beginning of document. Unless all of us behave well and the value adder/compiler clearly cites his/her sources.

I would add another issue, namely making certain that all data sets uploaded to the GCI provide the meta data needed for the citation standard, because if these are incomplete, the citation will not work in spite of all else.