

NOAA Coral Reef Watch
Recommendations for Coral Reef Watch
Data and Product Citation

(12 April 2011)

1. Introduction

Best practices and critical research needs in data publication and citation, especially within the Earth and space science data community, are beginning to emerge. Scientists are collectively calling for greater attention to these practices and needs (Parsons *et al.*, 2010). The International Polar Year (IPY) has significantly advanced this effort by recommending data citation in the IPY Data Policy (http://classic.ipy.org/Subcommittees/final_ipy_data_policy.pdf) and by developing guidelines for data citation (<http://ipydis.org/data/citations.html>). The IPY citation guidelines harmonize different approaches and have been adopted by many data centers worldwide (Parsons *et al.*, 2010). Although imperfect, these guidelines serve as a basis for evolving approaches to formally citing data. We (Coral Reef Watch (CRW)) have developed recommendations for citing our data and products, based on the IPY guidelines (<http://ipydis.org/data/citations.html>). The recommendations are expected to evolve over time, as improved citation approaches become available.

In scientific publication, merely acknowledging a data set in the text or in the Acknowledgments section is insufficient. The IPY Data Policy states, 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.” The IPY guidelines can help data users develop appropriate citations for data used in their publications and can help data managers recommend appropriate citation of their holdings. The IPY 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.

According to the IPY guidelines, in general, data sets should be cited like books. When users cite data, they need to use the style dictated by their publishers. However, by providing example citations, data publishers can give users all the important elements they should include in their data set citations.

This CRW document summarizes guidelines and citation examples from the IPY guidelines to provide recommendations for CRW data and product users. Recommended key elements of the citation are, in order: Author(s), Publication Date, Title, Dates Used, Editor/Compiler, Place of Publication, Publisher, Distribution Medium or Location, and Access Date.

Sample 1: If time series data for the CRW Satellite Virtual Station at Midway Atoll are used for analysis that leads to a publication, the data may be cited as:

NOAA Coral Reef Watch. 2000, updated twice-weekly. NOAA Coral Reef Watch 50-km Satellite Virtual Station Time Series Data for Midway Atoll, Jan. 1, 2002–Dec. 31, 2002. Silver Spring, Maryland, USA: NOAA Coral Reef Watch. Data set accessed 2010-05-14 at <http://coralreefwatch.noaa.gov/satellite/vs/index.html>

Sample 2: If Degree Heating Weeks HDF data are used for analysis that leads to a publication, the data may be cited as:

NOAA Coral Reef Watch. 2000, updated twice-weekly. NOAA Coral Reef Watch Operational 50-km Satellite Coral Bleaching Degree Heating Weeks Product, Jan. 1,

2001–Dec. 31, 2010. Silver Spring, Maryland, USA: NOAA Coral Reef Watch. Data set accessed 2011-01-05 at <http://coralreefwatch.noaa.gov/satellite/hdf/index.html>

2. Summarized IPY Guidelines

CRW's recommendations for citing CRW's data and products are based on the IPY guidelines (<http://ipydis.org/data/citations.html>). The IPY guidelines are summarized in the table below, along with their examples shown in the following section. The citation for CRW's data and products should include the following elements (see the table below) as appropriate.

<p>Author or investigator</p>	<p>This is the organization(s) or individual(s), who developed the data set, in a particular field experiment or using a particular algorithm (see Sample 1 below).</p> <p>If the name of editor or compiler is provided, the name must be followed by "(ed. (i.e., editor))" or "(comp. (i.e., compiler))", respectively (see Sample 2).</p> <p>If the data set is a collection of several smaller, independent data sets, the individual data sets should have their own specific citations with author, but the whole collection should not have an author. Instead the collection would like have an editor or compiler (Sample 3).</p>
<p>Publication date</p>	<p>This is the date when the data set was published or otherwise made available for release.</p> <p>1) For a complete data set, publication date is simply the year of release (Sample 4).</p> <p>2) 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 and the data are simply updated (Sample 5).</p> <p>3) For an ongoing data set that is updated on a regular or continual basis, list the first year of publication followed by the frequency of the update. Updates can occur annually or more frequently (Samples 6 and 7).</p> <p>A note on updates vs. new versions:</p> <p>Ongoing updates to a time series 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 (Samples 8 and 9).</p> <p>If a particular version of a time series is discontinued, it is appropriate to indicate when the final update occurred (Sample 10).</p>
<p>Title</p>	<p>The formal title of the data set. It may also include the version or edition information (Sample 11).</p>

Dates used	For a time series, especially a continually updated time series, indicate the dates data were used (Sample 12). Note: this is distinct from the publication date.
Editor or compiler	<p>This is the name of organization, team, or individual 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 but substantial effort was put into compiling the product, the Editor may be more correctly cited as a Compiler. Editors and Compilers are often 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. If the name of the Editor or Compiler is provided, the name must be followed by "(ed. (i.e., editor))" or "(comp. (i.e., compiler))," respectively (Samples 13, 14, and 15).</p> <p>When there is an Editor or Compiler but no author, the Editor is listed first (Sample 16).</p>
Publication place	The name of the city (and state or province and country, if needed to identify the city) where the data set was published or released (Sample 17).
Publisher	This is the name of the individual or organization who published the data set. A publisher often has an implied responsibility for data set stewardship. The publisher is usually a data center and is listed immediately after publication place (Sample 18).
Distributor or associate publisher	This field should be used only when it differs from the publisher, which is rare. 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 (Samples 19 and 20).
Distribution medium or location	<p>If there is one fixed medium, list it. For example, CD-ROM, DVD (Sample 21).</p> <p>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 (Sample 22).</p> <p>Ideally, a persistent identifier such as a Digital Object Identifier (DOI) should be used (Sample 23).</p>
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 (Sample 24).
Data within a larger work	<p>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 (Sample 25).</p> <p>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 (Sample 26).</p>

3. Coral Reef Watch Data Citation Samples

The citation samples provided in this section were copied from the IPY guidelines (<http://ipydis.org/data/citations.html>). The description of these samples is given in the previous section.

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.

Sample 1:

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.

Sample 2:

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 should have their own specific citations with author, but the whole collection should not have an author. Instead the collection would like have an editor or compiler.

Sample 3:

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 complete data set, the publication date is simply the year of release.

Sample 4:

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, and the data are simply updated.

Sample 5:

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 frequency of the update. Updates can occur annually or more frequently.

Sample 6:

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>.

Sample 7:

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 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 version number.

Sample 8:

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>.

Sample 9:

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.

Sample 10:

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.

Sample 11:

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 data were used. Note this is distinct from the publication date.

Sample 12:

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 but substantial effort was put into compiling the product, the Editor may be more correctly cited as a Compiler. Editors and Compilers often may 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 have neither. If the name of the

Editors or Compilers is provided, the name must be followed by "(ed. (i.e., editor))" or "(comp. (i.e., compiler))," respectively.

Sample 13:

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.

Sample 14:

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>.

Sample 15:

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.

Sample 16:

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 Place

This is the city and state or province and country, if needed to identify the city, of the publisher.

Sample 17:

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 directly following the place of publication.

Sample 18:

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, which is rarely. The Distributor or Associate Publisher should be listed in the same manner as a publisher would be.

Sample 19:

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.

Sample 20:

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, such as a CD-ROM or DVD, list as such.

Sample 21:

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.

Sample 22:

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 (DOI) also should be used.

Sample 23:

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](https://doi.org/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.

Sample 24:

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.

Sample 25:

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 a data supplement, one need only cite the parent reference. For example, when using the data at <http://doi.pangaea.de/10.1594/PANGAEA.476007>, the following reference is appropriate.

Sample 26:

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.

References:

Parsons, M. A., R. Duerr, and J.B. Minster. 2010. Data Citation and Peer Review. *Eos* 91(34): 297-298.