

GoMODP Metadata Editorial Committee

Quality-control of metadata citation information available from
GoMODP partners

Daniel Ricard^{1 2} Robert M. Branton^{1 3}
Lou Van Guelpen^{1 4}

March 31st 2009 - GoMODP AGM - Portland, ME

¹Centre for Marine Biodiversity

²Dept. of Biology, Dalhousie University, Halifax, NS, Canada

³Ocean Tracking Network, Dalhousie University, Halifax, NS, Canada

⁴Huntsman Marine Science Centre, St. Andrews, NB, Canada

1 Background

2 Methods

3 Results

4 Discussion

Metadata editorial committee

Recent work:

- Encourage use of the Global Change Master Directory (GCMD) for handling metadata records
- Use of R to summarise, quality-control and visualise metadata on GCMD portal
- Document the flow of authority from Data Providers to Data Centres, Partners and Partnership

Current work:

- Improve citation information in metadata records
- Empower partners to effectively guide individual data providers

GoMODP portal on the Global Change Master Directory (GCMD)

The GoMODP portal on GCMD was used to identify the different GoMODP partners and to determine what data centres were associated with each partner.

<http://gcmd.nasa.gov/KeywordSearch/gomodp/partners.html>

The screenshot shows a web browser window displaying the GoMODP portal. The browser's address bar shows the URL <http://gcmd.nasa.gov/KeywordSearch/gomodp/partners.html>. The page title is "GoMODP Partners - Mozilla Firefox". The browser's search bar contains the text "Search Web". The page content includes a navigation menu with links for HOME, DATA SEARCH, DATA SERVICES, AUTHORIZING TOOLS, PARTNERS, ABOUT US, and GCMD HOME. A sidebar on the left contains a list of links: About Portals, GCMD Portal Listings, Add to the GoMODP portal, GoMODP Data Sets, GOMMP Portal, and GoMODP Website. The main content area is titled "Find Data Sets at these Nodes:" and lists several data centers with their logos and names: Bedford Institute of Oceanography, Centre for Marine Biodiversity, Coastal Ocean Observation and Analysis, Environmental Protection Agency, NOAA - Coastal Services Center, NOAA - Northeast Fisheries Science Center, NOAA - National Environmental Satellite, Data and Information Service, National Oceanographic Data Center, and St. Andrews Biological Station.

Why is citation information important?

One recurring theme at Ocean Biodiversity Informatics 2007 conference was: "*How can we ensure that all data providers receive their due when their data are used?*"

- Gives much-needed credit to data providers
- Used by data users for citation in manuscripts and reports
- Provides authority to the metadata record and the underlying data
- Ensures reproducibility of analyses (i.e. other people can find the same data)

Example of proper citation information

Metadata citation information was revised during the OBIS Canada makeover project. Here are two of the 19 metadata citations that were reviewed and revised:

- Van Guelpen, L. *Atlantic Reference Centre Museum of Canadian Atlantic Organisms - Invertebrates and Fishes Data*. In: *Atlantic Reference Centre Digital Collections*. OBIS Canada, Bedford Institute of Oceanography, Dartmouth, Nova Scotia, Canada, 2006, Version 2, Digital.
- K. Smedbol, J. Gale, L. Bajona *Canada Maritimes Regional Cetacean Sightings*. In: *Fisheries and Oceans, Canada - Digital Collections*. OBIS Canada, Dartmouth, Nova Scotia, 2005, Version 1, Digital. Originating database created/maintained at St. Andrews Biological Station, New Brunswick.

What is the citation information and where is it in the metadata record?

Concept	GCMD/DIF	Geoconnections/FGDC
1 Creator	Data_Set Citation: Dataset_Creator	8.1 Originator
2 Title	Data_Set Citation: Dataset_Title	8.4 Title
3 Series	Data_Set Citation: Dataset_Series_Name	8.7.1 Series Name
4 Publisher	Data_Set Citation: Dataset_Publisher	8.8.2 Publisher
5 Place	Data_Set Citation: Dataset_Release_Place	8.8.1 Publication Place
6 Date	Data_Set Citation: Dataset_Release_Date	8.2 Publication Date
7 Version	Data_Set Citation: Version	8.5 Edition
8 Form	Data_Set Citation: Data_Presentation_Form	8.6 Geospatial Data Presentation Form

How is GoMODP doing?

Overall results for GoMODP, broken down by originating metadata node:

Status	GCMD/DIF	Geoconnections/FGDC	Total
No citation information	72	0	72
Warnings and errors	12	18	30
Warnings only	66	8	74
Proper citation	0	1*	1
Total	150	27	177

*Tufts University Center for Conservation Medicine, SEANET Beached Bird Survey Data, SEANET Data, Tufts University Center for Conservation Medicine, North Grafton, Massachusetts, United States of America, 2005-03-25, March 2005, Map > Thematic Map, Spreadsheet, Tabular Digital Data, Vector Digital Data

The GoMODP metadata from GCMD as a single dataset

The basic idea of the analysis presented here is to treat all metadata records associated with the GoMODP as a single dataset.

This dataset is analysed to identify missing citation information.

Errors and missing information are compiled for each metadata record and summarised for each Data Centre and Partner.

- GCMD provides an Application Programming Interface (API) to access DIF metadata record
- This API provides an alternative to the web-based interface available

Information about the API:

http://gcmd.gsfc.nasa.gov/Aboutus/collaborations/tools/archive_api.html

R script to analyse metadata records from GoMODP partners

Using the GCMD API and the R Software Environment for Statistics and Graphics, we wrote a script that performs the following tasks:

- For each GoMODP partner
 - For each data centre
 - 1 Retrieve the citation information contained in each metadata record
 - 2 Create a hyperlink to the metadata record in its originating context
 - 3 Analyzes the citation information available
 - 4 Write results to HTML file

Metadata citation - quality control report

The full report is available from the GoMODP site.

This page is the primary tool for this initiative and we encourage all partners to make use of it.

<http://www.gomodp.org/Members/branton/>

[metadata-editorial-committee/gcmd-gomodp-citations-informed/](http://www.gomodp.org/Members/branton/metadata-editorial-committee/gcmd-gomodp-citations-informed/)

Metadata Citation quality control report

A quality control report of data citations on the GCHD portal. Click partner name to list of data centers and citations for the partner. Citations are linked to the original source portal (e.g. GCMD, CGDI) as well as includes a list of missing elements (e.g. Place). Partners, data center staff and providers click the link to edit the metadata records and provide missing elements. See "HowTo Videos" also in this folder for help. In some cases, missing elements may be found elsewhere in the particular metadata record (e.g. CitationsDatasetTitle may be copied from EntryTitle). See the "Xwalks between Metadata Standards" page, also in this folder for suggestions on where to find the missing elements. Summary counts of TotalRecords by Portal, MissingCitations, GoodRecords, WithWarningsOnly, WithWarningsAndOrErrors, Errors, Warnings are also given.

Quality-control of GCMD metadata records from GoMODP partners

This file was generated by an R script invoked on Mon Jan 19 22:47:59 2009

As found on this GCHD URL, there are currently 18 partners with DIF records in the GoMODP:

1. Bedford Institute of Oceanography (5 data centre(s))
2. Centre for Marine Biodiversity (3 data centre(s))
3. Coastal Ocean Observation and Analysis (1 data centre(s))
4. Environmental Protection Agency (2 data centre(s))
5. Gulf of Maine Area Program of the Census of Marine Life (1 data centre(s))
6. Gulf of Maine Ocean Observing System (1 data centre(s))
7. Gulf of Maine Research Institute (2 data centre(s))
8. Hurlstman Marine Science Centre/Atlantic Reference Centre (1 data centre(s))
9. Maine Department of Marine Resources (1 data centre(s))
10. New Hampshire Department of Environmental Services (1 data centre(s))
11. NOAA - Coastal Services Center (1 data centre(s))
12. NOAA - Northeast Fisheries Science Center (3 data centre(s))
13. Stellwagen Bank National Marine Sanctuary (1 data centre(s))
14. Tufts University School of Veterinary Medicine (1 data centre(s))
15. USGS - Woods Hole Science Center (5 data centre(s))
16. US Fish and Wildlife (2 data centre(s))
17. Wells National Estuarine Research Reserve (1 data centre(s))
18. Woods Hole Oceanographic Institution (1 data centre(s))

Example of report output

Partner: Huntsman Marine Science Centre/Atlantic Reference Center

Huntsman Marine Science Centre/Atlantic Reference Center (1 data centre(s))		
Dataset ID	Dataset title	Citation
Data centre: HUNTSMAN-MSC/ARC, with 1 DIF record(s)		
Canada_ARC_InvertebratesFishes originating from CANADA-CGDI URL to DIF record in its originating context	Atlantic Reference Centre Museum of Canadian Atlantic Organisms - Invertebrates and Fishes Data	Atlantic Reference Centre Museum of Canadian Atlantic Organisms, Atlantic Reference Centre Museum of Canadian Atlantic Organisms - Invertebrates and Fishes Data, W02 No Series , Bedford Institute of Oceanography, Dartmouth, Nova Scotia, Canada, 200403, W03 No Version, W04 No Form , Record Processing Summary: 0 Errors, 3 Warnings.
HUNTSMAN-MSC/ARC: 1 TotalRecords, 0 GCMD, 1 CGDI, 0 MissingCitations, 0 GoodRecords, 1 WithWarningsOnly, 0 WithWarningsAndOrErrors, 0 Errors, 3 Warnings.		

- 1 metadata record from 1 data centre
 - Record originates from Geoconnections/FGDC
 - 3 warnings for that record
 - W02 - No Series
 - W03 - No Version
 - W04 - No Form

Not all warnings are relevant (e.g. Series and Version do not necessarily apply here)

Where to from here?

- Review and revise citation information
 - Option 1: a group effort, where each partner and their provider(s) fix their citation information
 - Option 2a: a one-person, one-off manual effort to review and revise all GoMODP-related citation information, subsequently approved by partners and providers
 - Option 2b: like 2a, using an R script to systematically harvest missing citation information from elsewhere in the metadata record

Where is this missing information in the metadata record?

<http://www.gomodp.org/Members/branton/metadata-editorial-committee/xwalk/>

Where to from here (cont.)?

- Introduction of a limited set of standardized keywords to facilitate discovery by different end-user groups
- Apply the methodology developed here to other portal communities within GCMD (e.g. OBIS) and institutions using the ISO-19115 metadata standard

Thank you for your attention

Robert Branton

rmbranto@dal.ca



Daniel Ricard

ricardd@mathstat.dal.ca

Lou Van Guelpen

**ARC@mar.dfo-
mpo.gc.ca**

**with funding from the Centre for Marine Biodiversity and
Fisheries and Oceans Canada**



Fisheries and Oceans Canada Pêches et Océans Canada

