

Project Report to the Executive Committee of the
Gulf of Maine Ocean Data Partnership

for

A Pilot Project Toward Registration of Metadata for the
Gulf of Maine Ocean Data Partnership

By the

Huntsman Marine Science Centre

Lou Van Guelpen
22 June 2005

The first annual meeting for the Governing Board of the Gulf of Maine Ocean Data Partnership (GoMODP) was held 7 December 2004. At that meeting, the member for the Huntsman Marine Science Centre was requested to provide a proposal for a pilot project to register metadata for five GoMODP partners: NOAA-Northeast Fisheries Science Center (trawl surveys), Maine Department of Marine Resources (trawl surveys), Stellwagen Bank National Marine Sanctuary (habitat use by humans and marine mammals), Tufts University SEANET (seabird mortality), and Wells National Estuary Research Reserve (benthic invertebrates).

This pilot project is a “Product and Milestone” in the GoMODP Work Plan Project 2: Data Discovery (page 9). The goals were 1) to demonstrate, through the registrations of metadata for the five partners, how discovery of data sets/collections can be automated by having data providers “register” their data with the GeoConnections Discovery Portal, 2) to report on lessons learned in the process, 3) to provide recommendations how partners should proceed with the registration of their metadata, and 4) a goal added by the Technical Committee – to report on the experiences of pilot partners regarding “data authority”.

1) Automation of data discovery through the GeoConnections Discovery Portal

The GeoConnections Discovery Portal (<http://www.geoconnections.org/CGDI.cfm/fuseaction/home.welcome/lang/E/gcs.cfm>), following the FGDC metadata standard, is the principal clearing house for geospatial metadata in Canada. The portal issued a new release on 16 June 2005. From the home page metadata are discoverable by clicking **Quick Search**. This application seems to work poorly, as it returned no GoMODP partner metadata for “Gulf of Maine”, “Maine”, or “Gulf of Maine Ocean Data Partnership” which were keywords entered by each partner. Metadata also may be discovered from the home page by consecutively clicking Sitemap, GeoConnections Discovery Portal (far down on the Sitemap page), and Search for Data. This procedure leads to **Advanced Search Options** with location, subject, keywords, time period, product type, an alphabetical list of database names,

and other lists (all databases searchable via the Internet). **Location**, selected through a dropdown list, is ambiguous (does it mean location of metadata owner, data coverage, or portal where metadata were registered?), and spurious (searching United States returns metadata that include Canadian content by Canadian owners which were not registered on a US metadata portal). **Subject**, also selected via a dropdown list, is very general (e.g. “Oceans” is the closest for biological data, returning thousands of entries). **Keywords**, the most important search parameter, seems to work quite well, e.g. “Gulf of Maine” and “Gulf of Maine Ocean Data Partnership” returned the partners in this pilot project. **Time Period** is a dropdown list ranging from “1 day past” to “50 years past”, and would seem to be of limited help unless used in conjunction with other options. **Product Type** is a dropdown list of general topics (e.g. airborne imagery, nautical maps and charts, satellite imagery) which may be useful to some, especially in conjunction with other options. Interestingly, the choices for product type don’t match those available when entering metadata content. The **Alphabetical list of database names** works well if the first word of the database title is known, but when not apparent (e.g. “Standardized Survey Sightings Data for the Stellwagen Bank National Marine Sanctuary”), searching can be tedious or unsuccessful. **Other Lists** (all databases searchable via the Internet) returned 258 items.

2) Lessons learned

- a) Metadata registration during this project generally seems to have been given a low priority by partners, with delays to late in the project period, or afterward, requiring a no cost extension to the contract. This may be the case when full scale partnership data registration is implemented.
- b) The intricacies of GeoConnections metadata entry can be confusing, and much jargon is involved. However, entry becomes easier with practice. Each GeoConnections metadata entry blank contains a clickable guidance button, but the guidance can be confusing because of jargon or ambiguity. It seems this guidance was not always consulted by partners. The concept of entities and their attributes and attribute values, which comprise the fields and rows of a dataset, are restrictive. Only one entity, with only one attribute, is allowed, requiring creativity when completing these entries. For all these reasons, much advice was required for partner metadata entry.
- c) A “getting started” package was provided to partners, but this was not always consulted. Occasionally advice was overlooked.
- d) GeoConnections has some idiosyncrasies. For instance, changing the “Descriptive Name” of a dataset by one partner resulted in a new metadata product listed on GeoConnections. On another occasion, a partner copied and pasted a metadata string from my account, open in one web browser, to his account open in another browser. Upon saving his session, nearly all his metadata were copied over with mine. Only GeoConnections portal personnel could resolve these problems.
- e) Partner metadata entry sometimes contained inaccuracies, e.g. Bounding Box coordinates.

- f) After metadata are drafted, they should be reviewed by the senior responsible person so the view, accuracy, and level of sophistication of the institution are properly represented.
- g) GeoConnections mirrors metadata registered through NASA's Global Change Master Directory (GCMD). GeoConnections shares its metadata (Canadian or US) with the GCMD. This means GoMODP partners will be able to register their metadata with either portal, and be discoverable through both. GeoConnections has no sharing agreements with NASA's Geospatial One-Stop. The Technical Committee also is investigating relationships among the various metadata portals to make recommendations to the partnership.
- h) One partner raised some important questions about the GeoConnections Discovery Portal which were answered by the Geospatial Metadata Coordinator:
 - i. **Q** - Is it possible to save our metadata files locally as txt, xml, or html? Can we edit outside of the Geoconnections interface and then upload again? **A** - Another colleague and I have tested the XML import/export functionality on our system and we have experienced many problems. Records can only be uploaded through a trial and error process and even if the upload was successful to the database, our stylesheets could not display the metadata properly. Additionally, the metadata editing interface on the Discovery Portal could not load these records so future updates can't be performed. At this stage, I wouldn't recommend attempting to work on metadata outside of the Discovery Portal. It is likely that we will have to remove the XML import/export functionality but we hope to further develop this component and make it available again in the future.
 - ii. **Q** - If we need to create more metadata for other data layers, do we need to start from scratch or can we copy data from existing metadata files? **A** - Copies of metadata records can be made by using the [Save As] button at the top-right of the editing interface. You will be asked to enter a new Product Collection Name and upon clicking [Save] a new entry with a new ID will be created under your account. This should save you from having to retype everything from scratch if an entry that could be used as a template already exists under your account. If the template entry exists under another account you could either be granted permissions to make a copy of it or a GDP Admin could provide you with a copy.

3) Recommendations

Recommendations herein are restricted to registration of metadata with Canada's GeoConnections Discovery Portal. The Executive Committee Chair made the decision for this pilot project that each partner would register his/her own metadata, with review and advice by a designated, knowledgeable individual (Lou Van Guelpen), rather than have the individual register all metadata for the five partners with review by each. This decision likely was sound, since each partner has all required information in-house, understands his/her dataset much better than an outsider, and thus should create a more clear and accurate metadata description of the dataset. The alternative approach would require each partner to determine and provide many of the metadata entries to

the designated individual, with substantial revision by the partner to reflect the desired metadata product. This seems a cumbersome and wasteful approach. Nevertheless, the results of the approach taken in this pilot project clearly indicate the need for guidance to GoMODP partners when registering metadata. Therefore, I make the following recommendations to the GoMODP for registration of partner metadata:

- a) The GeoConnections Discovery Portal seems a reasonable option for metadata registration, especially with the development of metadata sharing among portals. The GCMD is another viable option, and shares metadata with GeoConnections.
- b) Each partner should register his/her own metadata for each dataset in question.
- c) Metadata training for GoMODP partners is being arranged by the Metadata Training Committee. This step should provide a level of expertise within partner organizations. Nevertheless, partners likely will require guidance from a designated resource during metadata registration. Often the resource will require access to the dataset or a detailed data description to give responsible advice.
- d) The GoMODP Executive must decide how to assist partners lacking the resources to undertake metadata registration, whether to fund individual partners or a designated resource to assist in or perform the registration.

4) Partner experiences with “data authority”

The Technical Committee requested a report on the experiences of pilot partners regarding “data authority” – whether a partner is authorized to serve a dataset.

Questions asked of pilot partners were:

- a) Are you the primary creator/collector of the dataset? Who else has the data (are they being served elsewhere)? Is there shared ownership?
- b) Who will correct errors?
- c) Should the data user contact you (or the originator if appropriate) prior to use?

The responses of pilot partners were as follows:

NOAA-Northeast Fisheries Science Center

- a) No response.

Maine Department of Marine Resources

- a) Maine DMR is the sole, primary collector of the dataset. The data are not being served elsewhere.
- b) The project scientist, Sally Sherman, is responsible for any data discrepancies or corrections.
- c) Any user should contact the project scientist, Sally Sherman. Data are online in report form only.

Stellwagen Bank National Marine Sanctuary

- a) Stellwagen Bank NMS is authorized to serve their dataset, as the primary creator/collector. The data are not being served elsewhere at the moment,

though they may be posted on the NMSP(?) server in the future. There is no shared ownership.

- b) If necessary, errors will be corrected by Stellwagen Bank NMS.
- c) Data users require permission from the originator prior to downloading and use, and should cite the source.

Tufts University SEANET

- a) SEANET is the primary creator and collector for the dataset. The data also will be served on the NBII Wildlife Disease Information Node (WDIN) at the National Wildlife Health Center, USGS in Madison. There is no shared ownership (for now), only data distribution/sharing.
- b) SEANET will correct errors. An online data entry system, developed by NBII WDIN, will be utilized soon. SEANET still will be responsible for QA/QC.
- c) SEANET, as originator, would like to be contacted prior to data use. The Program Coordinator wants to add such a disclaimer, "particularly to try to prevent people from publishing data without contacting/giving SEANET credit, etc."

Wells National Estuary Research Reserve

- a) The contact, Jeremy Miller, responded that he was the creator of the dataset, but not Principal Investigator. He believed Wells Reserve to be the only owner of the dataset.
- b) Not answered directly, but questions are to be directed to Miller.
- c) Not answered, but data are not yet online.