

Feb. 28, 2005

**NSDI Cooperative Agreements Program
Don't Duck Metadata
Project Report: Final**

**Agreement Number: 03HQAG0174 (award number)
Final Report**

Organization:

Name/unit/office: University of Maryland Center for Environmental Science
Street Address: UMCES Appalachian Lab
Line 1: 301 Braddock Rd
Line 2:
City/State: Frostburg, MD
Zip: 21532

Website address: <http://www.al.umces.edu>

Collaborating Organizations: UMCES Horn Point Laboratory

Project Leader

Name: Dan Fiscus
Telephone: 301-689-7121
Fax: 301-689-7200
E-mail: fiscus@al.umces.edu

1. Number of Metadata files created as a result of this project: 12 metadata files that represent or refer to 135 GIS grids and coverages – some of the individual metadata records describe multiple data files, from 1 up to 25.

2. Clearinghouse Service. Is the metadata resulting from this project being served at a Clearinghouse site where it can be discovered and accessed?

Yes.

What is the Clearinghouse address:

FGDC clearing house node search page. <http://130.11.52.184/>

Comments (optional):

Our metadata files were posted via CIMS Online Metadata Entry Tool (COMET) of the Chesapeake Bay Program and our data is also available online there at <http://www.chesapeakebay.net/comet/>

3. For projects who received training assistance:

We held our own training.

Number of individuals that received training: 10

Is metadata documentation and creation a part of your organizations workflow? Describe

It is becoming a part of our workflow, but progress has been slower than originally hoped. At the training in January 2005, two major GIS faculty members agreed to begin to make metadata creation part of standard operating procedures, given ongoing support and guidance from myself. Unfortunately, we hit several obstacles, and I was not able to provide enough support and guidance to keep the momentum going, and the process slowed down. More information on this below.

6. Project Narrative (no more than 120 words): Summarize the project activities. Include its accomplishments, strengths and weaknesses, and next steps.

Our goals were to 1) build metadata capacity; 2) create metadata for 150 datasets; 3) make metadata available via NSDI; and 4) contribute to metadata, FGDC and NSDI communities. I learned to post metadata on NSDI via the Chesapeake Bay Program's metadata system, COMET. In a workshop, 10 staff learned basics of COMET, ArcCatalog and our data catalog, CODEX. GIS faculty committed to making metadata a standard practice. We did not reach consensus on a single metadata tool or a way to integrate multiple metadata systems. We have slowed in our metadata capacity development process. I posted to the NSDI basic metadata covering the majority of GIS coverages that we have created or to which we added significant value.

What areas need work?

We ran in to several problems that I did not foresee. One problem was the lack of a clear plan for integrating metadata across the lab, university, regional (Chesapeake Bay) and national/global (NSDI) scales. At the outset I did not realize we would need to navigate and participate in multiple metadata systems as connected to multiple agencies, purposes and audiences. Now it makes perfect sense that the ideal situation is for metadata to be entered only once yet be available on 1) local/lab/university (CODEX), 2) regional/Chesapeake (COMET) and 3) national/global (NSDI) platforms simultaneously. Also, we confronted the continued reality that metadata is not as important a priority here

as research, publication, and teaching. As one faculty member put it, we are not paid to document data, but to collect it, analyze it and publish the scientific results. Thus when metadata creation and capacity development are third, fourth or lower priority, and funding and incentive are not explicitly provided (beyond that for this grant, for the time required of other staff and faculty in ongoing ways) those tasks seem to get delayed or fall off the to-do list completely. Lastly, I believe that I provided insufficient follow-through, guidance and initiative as project leader. I under-estimated the complexity and then the amount of time and effort required to initiate such major cultural and workflow changes at Appalachian Lab and UMCES. This combination of events has led to a slow down in our metadata capacity development process. Thus we are now in need of reassessment of the project and perhaps new approaches to get past these obstacles. I begin that process here.

In assessing and brainstorming on the realities of developing a metadata capacity at our university research lab, several factors could help us make progress in the next phase. In addition to the buy-in and commitment we had from several key GIS faculty, it would help to have support and even a mandate from the Lab directors and the UMCES president. Additional support and stated needs for metadata from other administrators and faculty, such as the Vice President for Science Communication and the faculty involved in environmental education could help as well. In short, the consensus and internal mandate for metadata needs to be broad and well developed to achieve a “critical mass” of internal peer-pressure, effort and momentum sufficient to help get past the inertia of past patterns and workflow that has not included metadata as a value or as a product. Another aspect of this same issue relates to having additional staff in responsible positions to lead and push metadata capacity development. For example, it could have helped us to have a second and third staff person as back-up staff in charge of metadata in addition to myself. That way, when I hit my multiple personal health crises (diagnosis with diabetes in June, mother broke hip in September, baby born 4 weeks prematurely in October) and PhD rough spots (conditional pass on comprehensive exams in June followed by assignment of special research project to be completed in six months) the metadata development project would not have been so severely hampered. These are not excuses – I still take full responsibility for this project – but they did serve to raise the issue of staff “depth”, cross-training and teamwork for success in a long term, complex and culture-change organizational process. As such, the funding provided may not have been enough to expect completion of a metadata initiation process – additional funding to cover some salary for additional staff may be required.

Several technical and software issues caused us delays as well. Some of these may already be resolved in software updates that we have not yet gotten and installed. Our main GIS program, ESRI’s ArcGIS, did not have a metadata export function that was truly FGDC-compliant. This meant that either we had to enter metadata twice – once in ArcGIS and again in the COMET site, our truly FGDC-compliant portal to the NSDI – or we had to use an additional software tool to convert Arc metadata export files to be FGDC-compliant. One of my tasks was to get this conversion tool (developed and made available for free download by the National Park Service) and send it out to all

participants in our metadata training workshop, along with instructions on use within our metadata procedures. I did not complete this task, and I think that hindered our efforts and momentum. Within COMET, the online metadata entry tool of the Chesapeake Bay Program, one key function gave us errors and never worked. This was the option to use an existing metadata file as a template to create a new metadata file. I was counting on this tool to streamline and speed up the process for us, as many of our files share information such as projection, permission for use, etc. I worked with COMET technical support staff and emailed several rounds of information on the specific details of the problems and error messages I got. The problem still has not been resolved. A technical issue I had hoped to solve myself was to make import and export functions to get metadata files into and out of our in-house, web-based, open-source data catalog, CODEX. My perl and sql and other web programming skills were not strong enough to enable me to complete these functions in a timely way, and they still have not been implemented. Overall, despite the FGDC standard being well-documented, it is not universally available in all software, not even ArcGIS, and communication between metadata platforms at various organizational levels (lab, university, regional and national/global) posed greater technical obstacles than I had originally taken into account in project planning.

In terms of culture and values at our university research lab, metadata creation is still not a central priority despite our limited efforts in this project. I have three ideas toward improving the odds of metadata being adopted as a central value. These ideas are in addition to the faculty and administration buy-in and commitment mentioned above. First, it would help to have some way in which metadata creation leads to a publication in a science journal or other forum that is recognized as a bona fide publication on the resume of faculty or staff. That way, metadata creation as integrated with publication would lead directly to products and results that are integral to job performance, annual reviews and individual professional advancement. There are some journals or web repositories for datasets that seem to fill this need (Ecological Society of America and/or its journal, *Ecological Applications*, may publish datasets?), but I did not make the necessary connections to learn about them and integrate them into our organizational and technical processes. Second, assuming the software and system integration issues mentioned above could be resolved (i.e., metadata could be entered only once yet be available on local/lab/university/CODEX, regional/Chesapeake/COMET and national/global/NSDI platforms simultaneously), it would help to have a critical mass of metadata files and datasets available to make the value of metadata more clear to UMCES faculty, administration, staff and students. My original goal to make metadata available for 150 of our best GIS datasets was oriented to achieve this critical mass, but I did not get as far as I had hoped on metadata creation, and did not resolve the multi-platform issues. Last, and as related to the “critical mass” and “inertia of getting started” problems, one short-cut I took may be of general utility – to create a single metadata file that references a larger group of related datasets or datasets created under a single project. This short-cut serves the dual purposes of making some basic information discoverable while also shortening the time of metadata creation during the initiation process. As datasets are requested, shared, used, re-analyzed and questions asked of the original developers, the need for additional and more detailed metadata and perhaps the splitting into individual metadata

files would likely proceed naturally, or “by popular demand”. This approach does not alter or reduce the FGDC standard fields as some “metadata lite” approaches do. Instead it uses the full standard but lumps multiple datasets into a single metadata file. I did this and was able to post some basic information on 135 GIS grids and coverages in just 12 metadata files. Granted, not all of these files are as fully discoverable as they would be if each had its own metadata file.

Combining these kinds of modified approaches and strategies, I believe metadata development is still possible, and would still be greatly beneficial for us, at UMCES.

7. Feedback on Don't Duck Metadata Program

The goal of DDM program is to provide organizations with assistance for metadata creation and clearinghouse service through (a.) training, and (b.) metadata creation experience so that metadata documentation becomes part of an organizations normal workflow.

**What are the program strengths and weaknesses?
Where does the program make a difference?**

Helping groups get over the initially hard hurdles is great help! I guess our hurdles were bigger than I expected.

**Was the assistance you received sufficient or effective?
What would you recommend doing differently?
Are there factors that are missing or need to be considered that were missed?**

This section is not so much for DDM or CAP as for FGDC metadata standards and NDSI.

I see an area of potentially great importance in an increased role for visual modes of metadata or interacting with metadata. The UMCES Integration and Application Network (<http://ian.umces.edu>), which funded some of our initial work in metadata, has developed expertise in science communication and conceptual model diagrams. I had originally proposed and hoped to explore ways that this work could aid national efforts in metadata and data sharing by serving to augment the visual "channel" as an effective mode for discovering and also assessing the value of metadata and data for one's work. While the FGDC standard is comprehensive and well designed, I believe that it could be improved by a stronger and more integral role for visual modes of communication. This issue is especially emphasized in our field of environmental science and ecology, where systems we study are complex, highly interconnected and dynamic, and thus visualization can be a crucial aid to understanding, communication and management. A possibly useful analogy is to consider how much computer interfaces have improved, user-friendliness has increased, and user technical skill requirements and learning curves have been reduced by the integration of visual modes in the form of graphical user interface, icons and related

techniques. Especially for that set of the population that are “visual learners”, an enhanced visual mode to metadata creation and discovery could lead to increased use and value. Being a strongly visual person myself, I have to admit this as a personal bias. I don’t know how prevalent this preference is in the general population or within the environmental science research and management communities.

A somewhat related angle that I see as a potential avenue to benefit metadata standards and infrastructure involves the use of stories. An editorial in the Berkeley Wellness Letter mentions a report in medical journal (I can get these references if needed) that proposes that people remember stories better than numbers, tables or statistics. The spread of “urban legends” and anecdotal “truths” such as the need for car seats for children flying in airplanes based on a single child’s death are given as examples of how people not only remember but also retell stories. Stories are so powerful that they are often spread while ignoring facts and statistics that would alter the message in the story (airplanes are safer than car travel, for example, and if the car seat anecdotal scare story served to get more people driving instead of flying it likely resulted in more injury rather than less). The author of the original article suggests that humans may be “wired” to prefer stories or to think in stories. The author of the wellness newsletter editorial suggests that perhaps we scientists should get better at telling stories, ones that are based on facts and scientific findings. This combo would be the best of both worlds – compelling and “infectious” stories that are spread, and ones that also carry, convey and teach scientific understanding. The same idea could help with metadata, especially metadata for such complex topical matter as environmental science. If we could form stories from our research results, this might be a concise, readily accessible, and easily communicable form of information or knowledge to aid in sharing of both our science results and the data that went into those results. In relation to metadata, one of the potentially most interesting and desirable aspects of a dataset is its integral role in the telling of a story about how the world works. It may be that (again, for perhaps only a subset of learners or population that prefer stories, as in the example of visual learners above) a compelling story would serve to increase the interest in and sharing of datasets that figure in the telling and retelling of that scientific, environmental dynamics story.

A last idea for future metadata research and development is the combination of the visual, symbolic mode of communication with the story or narrative mode. In this combination I imagine “story boards” like those used to develop movies or animations: a story is told in a few, symbolic pictures. Another analogy is the use of symbols in sequence to give instructions for assembly or use of various products. My home water filter has four or so images in sequence that tell me how to take out my old filter from the faucet connection, insert the new filter and then run the water for five minutes to flush it out. These four images convey important semi-technical information in very quick and compact form. It seems such a format could be adapted to aid discovery and retrieval of metadata – each symbol panel might be an active web link to a dataset or science research project. Seen and “read” as a whole sequence, the series of symbols would tell a story of environmental science or management significance. The hypertext nature of the links would allow getting more details on the datasets that figure centrally in the telling of this story. The

story line as a whole would thus be a kind of “hypercontext” – the digital and linkable science dynamics context in which the individual datasets “make sense”. Likewise, it seems to me that context is always important for wise use of data, and that this is one of the primary functional purposes of metadata. FGDC and NSDI have developed “hypercontext” around tabular and alphanumeric data structures. It could be that visual and story-oriented data structures could complement this existing infrastructure in useful ways. I hope to help with work in this area in the future. One set of related products we developed are animations that show the dynamics of gypsy moth caterpillar defoliation of forests followed by increased nitrogen export from their watersheds. These are simple and just a start, and online here:
<http://al.umces.edu/~fiscus/gypsy>

Are there program management concerns that need to be addressed? Timeframe?

I found USGS and FGDC program management to be excellent with quick and helpful communication. Thank you!

--

Dan Fiscus
Ecologist/Research Assistant/PhD student
University of Maryland
Center for Environmental Science
Appalachian Lab
301 Braddock Rd
Frostburg, MD 21532
301-689-7121 (phone)
<http://al.umces.edu/~fiscus/research>