

Lynda Wayne  
FGDC Metadata Education Coordinator  
GeoMaxim  
60 Cherokee Road  
Asheville, NC 28801  
P 828.250.3801  
F 828.251.6242  
[Lwayne@GeoMaxim.com](mailto:Lwayne@GeoMaxim.com)

## **INSTITUTIONALIZE METADATA *BEFORE IT INSTITUTIONALIZES YOU***

**Abstract:** Geospatial metadata production commonly occurs at the end of the data development process. The approach is both cumbersome and questionable as individuals attempt to re-create data development methods and recall specific values. Metadata captured during data development are more accurate and provide greater functionality. The following operational measures can be utilized to better integrate metadata production into the geospatial data development process: build administrative and staff support, create organizational metadata templates, map metadata fields to workflow and distribute production, develop/integrate metadata collection tools, and establish metadata procedures and policies.

### **INTRODUCTION**

As the concept of geospatial data documentation, or metadata, is introduced to organizations, efforts generally focus on the documentation of existing geospatial data resources. As a result, metadata are captured after the data development process is complete. Unfortunately, most organizations continue this approach as they document new and evolving geospatial data resources. Metadata production outside of the data development process is both cumbersome and the results can lack integrity. Those producing the metadata must attempt to re-create stages of data development and recall specific values. The resulting metadata is often inaccurate and/or incomplete.

## **METADATA AS A DATA COMPONENT**

Metadata is a key component of the geospatial data set. It carries critical information as to the dataset purpose, location, content, and lineage. Geospatial software and analysts increasingly rely on metadata to ingest, display, and manage data. Perhaps most significantly, metadata is the consumer information needed by a rapidly growing geospatial data market to locate available geospatial data resources and assess their fitness for a particular use. Metadata instills data accountability and limits data liability. The timely capture of metadata is fundamental to the quality of the data set as a whole.

## **OBSTACLES TO METADATA PRODUCTION**

Most organizations have standing geospatial data development methodologies. While some methodologies are more formal than others, the integration of new processes and technologies is always disruptive. Interviews with metadata workshop participants and a survey of county GIS data developers conducted by Mecklenburg County, North Carolina reveal the following major obstacles to metadata production:

- metadata standards are too extensive and difficult to implement
- metadata production requires time and other resources
- there are few immediate and tangible benefits and fewer incentives to produce metadata.

## **MAKING METADATA PART OF THE PROCESS**

The objective of this effort is to address the obstacles outlined above by streamlining metadata creation into the workflow process and providing guidance in the development of policies and procedures that will encourage and enforce metadata production. The following recommended operational procedures were derived from national Federal Geographic Data Committee (FGDC) metadata coordination efforts and related implementation efforts in the State of Minnesota; Mecklenburg County, North Carolina; and the City of Salisbury, North Carolina.

### **Build Administrative Support**

Technical staff members are typically the first to learn of metadata as a means of documenting and managing geospatial data resources. It is the administrators, however, that are tasked with allocating the resources and instituting the policies critical to successful implementation. Therefore, those inspired to integrate metadata production must champion an effort to educate administrators. Education can occur during individual discussions with administrators or as short presentations and briefings during meetings. Demonstrations of data clearinghouse technology and metadata success stories can also prove effective. It is important that metadata advocates highlight the organizational benefits of metadata production (Table 1) as well as the commitment of resources (staff time, software, and training) and policies required of the organization.

**TABLE 1. ORGANIZATIONAL BENEFITS OF METADATA**

<b>Data Archive</b>
Data are the most expensive components of a GIS. Metadata is a means of preserving the value of data investments. This is of particular significance to local and regional governments experiencing rapid staff changes.
<b>Data Assessment</b>
GIS data development has shifted from data producers to data consumers. From a consumer perspective, metadata is the truth in labeling required to assess available data products. From the producer's perspective, metadata is a means of declaring data limitations and serves as a form of liability insurance.
<b>Data Management</b>
Metadata enables organizations to retrieve in-house data resources by specific criteria for global edits and annual updates
<b>Data Discovery</b>
Metadata is the primary means of locating available geospatial data resources via the Internet. Metadata is a primary public information resource as it is a non-technical means of presenting technical information.
<b>Data Transfer</b>
Metadata is increasingly used by software systems as a means of properly ingesting data and by analysts as a means of properly displaying data.
<b>Data Distribution</b>
By building metadata in compliance with national standards, you can participate in the Global Spatial Data Clearinghouse. Participation promotes your agency and frees staff from answering data inquiries.

### **Build Technical Support**

Since the primary responsibility of metadata production falls to technical staff, metadata advocates must also make the case that metadata production is beneficial to the individual (Table 2.). Technical staff should also be included in the decision-making process as they bring critical insight to metadata program design and feasibility analysis. The more staff are included in the decision-making process, the more likely they are to commit to the goals of the program.

**TABLE 2. INDIVIDUAL BENEFITS OF METADATA**

<b>Reduced Data Management Workload</b>
Metadata contain information that can be used to quickly locate and retrieve data resources by specific criteria including: keywords, time period, contacts, data type, entities and attributes, etc.
<b>Fewer Inquiries as to Data Availability and Content</b>
Most of the information required by data consumers to locate, evaluate, access, and ingest available data resources is contained within the metadata. As a result, data developers/managers are faced with fewer and more specific data inquiries.
<b>Document Personal Contributions</b>
The initial metadata record produced for a given data set establishes the core content that will persist, with updates, for the life of the data set. This provides data developers an opportunity to document their efforts and contributions and serves as a tangible performance indicator that may be incorporated into organizational and individual evaluations.

## **Create Organizational Metadata Templates**

Much of the angst regarding metadata is associated with related standards including the Content Standard for Digital Geospatial Metadata (CSDGM) and the forthcoming International Standards Organization (ISO) Metadata Standard for Geographic Information. The standards are extensive and somewhat overwhelming because they are written to address a wide range of geospatial data types (imagery, GIS files, GPS data, geocoded databases) developed by a wide range of organization types. Individual organizations can address the problem by building custom metadata templates.

Templates are built by extracting those metadata fields pertinent to the organization and the specific data types and geographies of the organization. In addition, libraries can be built to provide information about contacts, sources, and methodologies common to the organization. A suggested method for building such templates includes:

1. Adopt all mandatory fields specified by the national metadata standard
2. Adopt all 'mandatory if applicable fields' pertinent to the data type or organization
3. Identify 'optional fields' of interest to the organization
4. Create a pilot record from the draft template
5. Have the pilot reviewed and revised by administrators, analysts, technicians, and contributing scientists
6. Identify those fields that tend to remain consistent. This may include:
  - access and use constraint statements
  - data distribution methods and contacts
  - contact information
  - north, south, west, east bounding coordinates
  - coordinate system and datum
  - place keywords
  - native data set environment
  - source citations

## **Map Metadata Fields to Existing Data Development Workflow**

Once a template, or template set, is created for the organization, specific fields of the template can be mapped to key stages of data development. By doing so, metadata production is transformed from a looming complicated task to a series of manageable steps that can be distributed throughout the data development process. In addition, the responsibility for metadata production is distributed among those responsible for various stages of the process. A general mapping of fields is provided in Table 3.

**TABLE 3. METADATA FIELDS MAPPED TO WORKFLOW**

<b>Data Development Stage</b>	<b>Metadata Information</b>
<b>Data Planning</b>	Identification Information title, originator, abstract, purpose, keywords, time period Data Organization point, raster, vector Spatial Referencing coordinate system and datum Entity and Attributes (planned)
<b>Data Processing</b>	Data Quality completeness, positional accuracy, geoprocessing steps
<b>Data Analysis</b>	Data Quality attribute accuracy, analysis steps Entity and Attributes (results) Metadata Reference

### **Develop/Integrate Metadata Collection Tools**

An information collection tool is required to effectively implement metadata production. Tools can be built using text editor, word processing, database, and/or programming software. Commercial software products are available as stand-alone metadata producers or internal components of GIS software. The tools should be selected/developed to meet the specific needs of the organization. Tool options include:

#### Form Documents

Hardcopy or digital forms can be developed to capture metadata throughout the lifecycle of a data set. Hardcopy forms are often the easiest method of collecting information from analyst and data contributors alike, but to realize the full utility of the metadata, the information must be input into digital format at some point. Digital forms can be built to collect information or as an interface into a metadata database table. The primary advantage of forms is the ability to segregate the metadata information asked of the various data contributors and to limit exposure to the metadata standard.

#### Database / Spreadsheet

Data tables can be designed to store metadata information. Column headings relate to specific metadata fields and records are built for each data set. In addition, a series of related tables can be built to reflect the workflow outlined above. The primary advantage of the metadata database is the ability to integrate production rules such as population of mandatory fields, use of 'pick lists' for sources and contacts, and automated QA/QC routines.

#### Shareware Metadata Products

Government agencies developed in-house metadata production software when faced with federal requirements for metadata and a lack of commercial products to support the effort. Most of these products are freely distributed to the public. In addition to the low/no costs, these products offer the benefit of organizational or discipline specific features. For example, *Corpsmet* was developed for the US Army Corps of

Engineers and includes references to Tri-services data standards observed by the military community. Another product, *Metamaker*, was developed by the US Geological Survey's Biological Resources Division and is built to the CSDGM 'biological profile' used by many biological and natural resource organizations. Potential users should be aware that most shareware products provide no user support and few are actively updated.

#### Commercial Metadata Products

Commercial metadata products fall into two categories: stand-alone and GIS-internal. Stand-alone products generally allow the user to 'harvest' some metadata information directly from the geospatial data set and provide a user-interface for additional data entry. These programs are typically robust production tools that facilitate the building of templates and libraries and enable interaction with a range of data types. As such, they are particularly useful to organizations that produce and manage data using multiple data development software. GIS-internal products also provide a data entry interface but, due to their proprietary nature, are able to harvest more information directly from the data set.

When purchasing commercial metadata production software, consumers should consider if the software:

- is easily understood and implemented
- is built to existing and applicable metadata and Internet standards
- automatically captures and updates much of the metadata
- supports the construct of digital forms that can be used to customize data entry
- supports global edits and updates
- provides optional and/or custom formats for viewing the metadata
- 'bundles' the metadata with the data
- supports required metadata import and export formats.

#### **Develop Metadata Procedures and Policies**

To effectively implement metadata production into the data development process, procedures and policies are needed to guide participation, streamline operations, and encourage compliance. Procedures and policies can be developed to address each of the following:

##### Assign Responsibilities

If metadata collection has been mapped to the workflow and tools have been developed to facilitate metadata capture, it is now possible to assign metadata collection and management responsibilities.

**Managers** can be responsible for:

- documenting metadata information mapped to the data planning stage
- coordinating the overall collection of metadata
- enforcing metadata policies

**Technicians** can be responsible for:

- documenting metadata information mapped to the data processing stage
- building source and contact citations

**Scientists and Field Staff** can be responsible for:

- reviewing and revising metadata information pertinent to data collection methods and findings

**Analysts** can be responsible for:

- documenting metadata information mapped to the data analysis stage
- assisting technicians in metadata documentation

**Information Technology / System Managers** can be responsible for:

- developing and maintaining the metadata collection 'tool'
- managing and updating metadata records
- distributing metadata within and external to the organization

**Data Stewards** can be established as responsible parties for the maintenance and distribution of specific data and metadata records.

#### Assign Priorities

Organizations embarking on metadata production are often overwhelmed by the prospect of documenting large numbers of both archival and actively produced data sets. A plan of action can be established that prioritizes the order in which data sets are documented based upon the following:

- value of data set as a core, or framework, data product of the organization (note: these data are prime candidates for data stewardship programs).
- utility of the data set within the organization
- number of external requests for the data set
- historical significance of the data set to the organization

#### Establish Administrative Guidelines

Guidelines are needed to aid users in the comprehension and implementation of program objectives. Options for guideline development include:

- define organizational compliance through the use of standards and templates
- establish standardized language for data and metadata distribution liability and access/use constraint statements
- develop boilerplate contract language to include metadata as a required deliverable when contracting for external data development
- include metadata skills in RS/GIS position descriptions and performance measures
- require organizational units to publish their metadata holdings
- publish a metadata standard operating procedure (SOP) manual to document the policies and procedures outlined above.

### Implement and Advocate

To fully integrate metadata into the geospatial data development process it is important to advocate participation by:

- providing staff metadata training
- publishing your efforts in professional and research journals
- providing incentives such as prizes and awards
- enforcing participation using project ‘punch lists’ that indicate work is not complete until the metadata is complete
- presenting metadata as a management priority during meetings, memo’s, and presentations.

### **CONCLUSIONS**

Metadata provides many benefits to organizational and individual geospatial data producers. However, quality metadata production requires additional resources and changes to existing data development procedures and policies. The recommendations outlined above were developed in direct response to metadata production obstacles cited by geospatial data developers and are summarized in Table 4.

**TABLE 4. SUMMARY OF METADATA PRODUCTION OBSTACLES AND RECOMMENDATIONS**

<b>Metadata Production Obstacle</b>	<b>Metadata Production Recommendation</b>
<b>Standards are overwhelming</b>	<ul style="list-style-type: none"><li>▪ create organizational templates</li><li>▪ develop/integrate metadata collection tool</li></ul>
<b>Requires excess time and resources</b>	<ul style="list-style-type: none"><li>▪ build administrative support</li><li>▪ provide training</li><li>▪ map metadata fields to workflow and distribute metadata production</li></ul>
<b>Few tangible benefits and incentives</b>	<ul style="list-style-type: none"><li>▪ highlight organizational and individual benefits</li><li>▪ include metadata in job descriptions and performance measures</li><li>▪ establish policies that require metadata production</li><li>▪ provide prizes and awards</li><li>▪ publish your efforts</li></ul>

### **ACKNOWLEDGEMENT:**

This research was sponsored by the Federal Geographic Data Committee and supported by the experiences and insight of various members of the ‘national metadata cadre’.

### **REFERENCES**

Cialek, Chris, Nancy Rader, Nicole Beck. May 2001. *Minnesota Metadata Mind Meld – Meeting Summary*, unpublished.

Clifton, Kathryn, August 2000. *City of Salisbury - Metadata Worksheets*, unpublished.



Federal Geographic Data Committee, March 1997. *Geospatial Metadata Factsheet*.

<http://www.fgdc.gov/publications/documents/metadata/metafact.pdf>

Federal Geographic Data Committee, accessed July 2001. *The Value of Metadata, Manager's Brochure*.

<http://www.fgdc.gov/publications/documents/metadata/metabroc.html>

Federal Geographic Data Committee, accessed July 2001. *Instructions for Implementing the FGDC Metadata Standard and Testing Conformance*.

<http://www.fgdc.gov/metadata/toollist/instru~1.html>

Federal Geographic Data Committee, accessed July 2001. *Top Ten Most Common Metadata Errors*,

<http://www.fgdc.gov/metadata/top10metadataerrors.pdf>

Mecklenburg County GIS, July 2001. *Metadata Survey Results*. unpublished.

Schweitzer, Peter. July 2000. *Editorial: Top Ten Metadata Mistakes*.

<http://geology.usgs.gov/tools/metadata/tools/doc/opinion/mistakes.html>

US Forest Service, 2001. *Forest Service Handbook (FSH) 6609.15, Chapter 30: Geographic Information Systems (GIS) Core Data Structure Standards*. Washington, D.C.

Wayne, Lynda. February 2001. *Making Metadata Work For You*. 2001 North Carolina GIS Conference, Winston-Salem, North Carolina.