Metadata in Action

Expanding the Utility of Geospatial Metadata

Lynda Wayne

The production of geospatial metadata is now established as a GIS best practice. However, few organizations utilize metadata to support functions beyond data distribution and archive. Robust, quality metadata can serve a critical role in geospatial data and project management. To realize this potential, metadata must be produced with this end in mind and contain key content for specific elements. Metadata creation and maintenance must also be fully integrated into the geospatial data lifecycle. Organizations willing to incorporate these measures into their geospatial data development practices will realize greater return on their geospatial data investments.

KEYWORDS

Metadata, Geospatial data, Data management, Project management, GIS

INTRODUCTION

As organizations slowly come to recognize the value of geospatial metadata, new horizons emerge with regard to their utility. Metadata are no longer limited to preserving the lineage of data investments or facilitating data discovery. Complete, robust data documentation is increasingly used to manage in-house data resource, assess the utility of available data resources, instill data accountability, establish data liability, monitor project status, evaluate return on geospatial data investments, and provide a common language for both data contributors and consumers.

Unfortunately, no strong business case has been established to illustrate the utility of metadata as a data and project management resource. Geospatial data developers often perceive metadata creation as a tedious process whose benefits are realized only by those that later utilize or inherit their data. Managers of geospatial projects and data collections have even less opportunity for exposure to the benefits of metadata and can therefore be resistant to requests for resources to cover the costs of metadata-related staff time, training, and software tools.

Many in the geospatial community, however, have come to consider metadata as both a component of the data and a Geospatial Information Science (GIS) best practice. While this realization is sometimes made when the data producer is faced with consumer requests for information about data processing methods, attribute definitions, and data quality assessments; others have simply learned, through exploration and experience, that this 'data about data' has strong potential to support geospatial data and project management.

It should be noted that this research is primarily based upon the experiences of remote sensing and GIS professionals utilizing the US Federal Geographic Data Committee's (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) metadata standard. The research findings, however, are broadly applicable to the utilization of metadata created using other geospatial and non-geospatial metadata standards.

DATA MANAGEMENT APPLICATIONS

The application of metadata to geospatial data management can be organized into four key components: data maintenance and update, data discovery and reuse, data accountability, and data liability.

Data Maintenance and Update

As the number of data sets within a collection grows and storage space is reduced, it can be difficult

to determine those data that should be maintained and those that should be dispensed. Candidates for data update, archive, or removal can be identified using temporal and data processing metadata elements.

Metadata temporal elements can be used to identify data that were developed prior to:

- some period of currency such as 'twenty years or older',
- administrative events such as redistricting and annexations that resulted in new political boundaries, or
- geophysical events such as earthquakes and storms that created new land forms and patterns.

Metadata data processing elements can be used to identify data that were developed using:

- outdated source data or
- analytical methodologies that are no longer considered current or adequate.

Metadata can also serve as an easy means of maintaining data currency during organizational and technological changes. With basic scripting, global edits can be made to the metadata to perform efficient updates to:

- contact names, phone numbers, and physical, mailing, and email addresses;
- distribution policies, availability, and pricing;
- URLs for project websites, data downloads, source data, data dictionaries, or distribution sites;
- keywords that incorporate ISO 19115 Topic Categories and other standardized vocabularies, and
- newly available derivations of the data set.

Data Discovery and Reuse

Metadata is the fuel for data discovery. Through the use of data portals and other forms of online data catalogs, users can post documentation about their data holdings and those in need, can query the Internet for available data using parameters of geography, time, and theme. Once potential data resources are discovered, users can access the metadata to learn more about the data content, availability, and limitation and better assess the fitness of use of the data to meet their specific need.

However, efforts to promote metadata as a means of data discovery have generally failed with organizations disinterested in widespread data distribution. Few local government data managers are inspired at the thought of people in other countries being able to access their street centerline files. In addition, many organizations simply prefer not to distribute their data. The real incentive to these and other organizations is the use of metadata to locate their own internal data resources.

Geospatial analysts are constantly faced with the need for data and no means of determining if the data exist in-house, if a similar in-house data resource can be adapted, or if the data must be newly created. In many cases, the last option is the default as more effort may be required to locate and evaluate existing data holdings than to generate new ones. Metadata provides a standardized format to capture critical information about data holdings and organize that information into an *in-house* data catalog. The data catalog, if effectively designed and well-maintained, can serve to preserve data lineage as new data products are derived and data contributors retire or change jobs.

To fully realize the benefits of internal data discovery, metadata should be created using rich and precise content for elements such as: keywords, time period references, geographic footprints, accuracy assessments, processing descriptions, and attribute descriptions. The metadata must also

be maintained with current URLs and file system locations. If these practices are implemented, in-house data can be easily retrieved and assessed with regard to:

- themes and attributes,
- geographic extent,
- temporal character,
- data processing methods,
- data contributors, and
- data quality.

Data Accountability

The practice of writing metadata is an exercise in data comprehension and accountability. The metadata creator must be willing to associate themselves with the metadata content. The individual documenting the data set must fully understand error analyses and data quality assessment, the character and limitations of the source data, and the definitions and domains of all attributes. Otherwise the individual is faced with either stating a lack of knowledge or seeking out the needed information. For data developed by a team of contributors, the metadata can be crafted such that the contribution of each team member is recorded and the individual accountability that is often lost in group projects can be maintained.

Data developers can also instil data accountability by recording the data processing steps and variables. This provides both a repeatable process and a defensible process. Repeatable process is simple scientific method and key to efficient revisions, updates, and application of the process to companion data and geographies. Repeatable processes are especially valuable in processing digital imagery and capturing the many choices made as to algorithms, class parameters, and acceptable deviations and anomalies. Defensible processes are vital to public participation projects where decisions about land use, transportation, and environmental quality are subject to scrutiny by scientists, the media, and the general public. The documentation of process methods becomes progressively more important as GIS is increasingly used to promote community decision-making and the public becomes aware of geospatial data resources, scalability, and mapping technologies.

Another measure of data accountability is the return on the investment (ROI) that the data provides. Metadata can serve to support data ROI assessment by providing an indication as to the value of the data as a foundation for the development of additional data resources. A simple query can be made to identify the number of times the data is cited as a source within the metadata of other in-house data. Regional, national, and international data portals can be searched to identify external data products generated using the target data as a source. In addition, specific metadata information about the role or significance of the target data to the derived data can be used to refine the assessment.

To facilitate discovery of the data when used as a source, organizations should consider the use of:

- standardized data titling schemes and
- a statement, within the use constraints, indicating the preferred source title reference.

Data Liability

A well-written metadata record is also an opportunity to state what the data are not. Metadata can prove most useful in advising others in the appropriate and misappropriate application of the data. An explicit purpose statement can clearly outline special project conditions and requirements that may affect the applicability of the data to other projects. Use constraint statements can be crafted to express scale, geographic, or temporal limitations to the data. Liability statements should be written by legal staff to ensure that the legal requirements for use of the data are fully outlined. In general, it is far better to publish your data set limitations within your metadata than to later

attempt to generate them in response to an inquiry or lawsuit.

PROJECT MANAGEMENT APPLICATIONS

The application of metadata to geospatial project management can be organized into four key components: project planning, project monitoring, common language for multi-participant projects, and contract deliverables.

Project Planning

A metadata record can serve as a standardized means of outlining a proposed project and establishing key data parameters. The metadata abstract can provide the overall context for the data. A purpose statement can be crafted to specify the role that the data is to serve within the project. Bounding coordinates are generated to establish the geographic extent of the project. The time period of content is used to express the temporal extent of the project. Preferred data resources can be indicated in the source documentation. Finally, a data dictionary can be can drafted to outline attributes labels, definitions, formats, and domains.

By establishing a core metadata record early in the project planning stage, several key benefits are realized:

- the manager's expectations are clearly communicated to the data developer,
- metadata creation is integrated into the data development process, and
- a medium is established for recording data processing and changes to the project parameters.

Project Monitoring

Metadata can be used to monitor project development if the core metadata record is maintained and expanded throughout the lifecycle of the data. Managers can periodically check the data processing information to assess the status of data development and to perform quality checks with regard to process methodologies, output and error analyses, field verifications, and the use of prescribed text, vocabularies, and attributes. Data developers can use the metadata record to document interim data products for later incorporation or to meet progress reporting requirements.

Common Language for Multi-participant Projects

The benefits of establishing a metadata record to communicate project planning information and monitor development can also be extended to multi-participant projects. This is most easily accomplished by the creation of a project metadata template that: 1) establishes the metadata elements considered vital to the project, and 2) provides specific content for those metadata elements that should be standardized across all project-related data. Standardized project metadata content may include:

- project descriptions,
- keywords and standardized vocabularies/thesauri,
- project contacts,
- common attributes, and
- distribution information.

The metadata template also provides a standardized reporting format for participants to document and share their data-specific information. If the metadata are regularly reviewed by all participants, there is improved opportunity for coordinating source data, complimentary analytical methods, and attributes of value to the broader project team. Used in this manner, the metadata record serves as a key means of communication among members of the data development team.

Contract Deliverables

Metadata should also be specified as a deliverable when contracting with others for the development of data. The metadata specification should include clear language as to the metadata standard that should be used and provide some indication as to the quality of the metadata expected. Metadata quality can be described through the use of a:

- recognized metadata classification scheme that defines 'levels' of metadata
- project metadata template, as described above, that indicates required metadata elements and the use of standardized language and vocabularies, or
- metadata specification manual that outlines the organization's standard operating procedures and requirements for creating metadata.

In each case, a sample metadata record should be provided to illustrate the expected content.

CONCLUSIONS

The utility of metadata as a data and project management resource can prove significant if organizations are willing to invest in the time, tools, and training necessary to create robust metadata and instil policies, guidelines, and practices that support metadata creation and maintenance. Metadata created with this end in mind can serve a multitude of functions. To the data archivist metadata are a means of recording lineage; to the accountant, metadata are a means of preserving investments; to the data developer, metadata are a means of documenting process; to the project manager, metadata are a means of tracking progress; to the project participants, metadata are a means of communicating; and to the data consumer metadata are a means of locating, assessing, and accessing data resources.

This is only part of the picture. The activities outlined within this paper represent only the human utility of metadata. Saarela (2005) suggests that software applications can also utilize metadata to identify, sort, and deliver custom information across enterprises, business processes, and end-user devices. Data developers interested in accessing their own information via mobile phones and personal data assistants (PDAs) or controlling the level of content presented to others, may finally find the incentive necessary to fully integrate metadata creation into the complete lifecycle of the data.

ACKNOWLEDGEMENTS

The author would like to recognize the contributions of the international geospatial metadata community via metadata list server discussions, email contributions, and personal dialogs. Most notable among these many contributors are:

Sharon Shin, *US Federal Geographic Data Committee Secretariat*

Doug Nebert, *US Federal Geographic Data Committee Secretariat*

Peter Schweitzer, *US Geological Survey*

Nina Savar, *University of Illinois at Chicago*

Learon Dalby, *Arkansas Geographic Information Office*

Ken Miller, *Maryland Department of Natural Resources*

Chris Cialek, *Minnesota Land Information Management Center*

Nancy Rader, *Minnesota Land Information Management Center*

Jill Trubey, *Florida Fish and Wildlife Research Institute*

Jill Leale, *US Minerals Management Service*

Dave Connell, *Australian Antarctic Data Centre*

Eric Landis, *Natural Resources Information Management*

Bruce Westcott, *Intergraph/SMMS Consultant*

REFERENCES

1. Federal Geographic Data Committee, Creating and Publishing Metadata in Support of Geospatial One-Stop and the NSDI, Federal Geographic Data Committee (FGDC) Publication, Reston VA USA, 2004.
   http://www.geo-one-stop.gov/metadata/CreatePublishMeta.pdf

2. Landis, Eric, Enabling FGDC Metadata Development in Biodiversity and Ecosystem Sciences, unpublished report to the Federal Geographic Data Committee, Stillwater MN USA, March 2005.

3. Saarela, Janne, Evolution From Content Management to Context Management, in Report of the Workshop on Content Management and Metadata sponsored by the European Standards Organisation, Information Society Standarisation System, Multimedia Information- Dublin Core (CEN ISSS MMI-DC) Working Group, Brussels, 2005.
   ftp://ftp.cenorm.be/public/ws-mmi-dc/mmidc140.htm

4. Wayne, Lynda, Institutionalize Metadata Before It Institutionalizes You, in Proc. Urban and Regional Information Systems (URISA) 2001 Annual Conference, URISA Publication Series.
   http://www.urisa.org/store/urisa_conference_proceedings.htm

5. Westcott, Bruce, and Trubey, Gillete, Case Study: Implementing an Enterprise Metadata Strategy, in Proc. 2005 GeoTec Event, Vancouver, BC, Canada, February, 2005.

AUTHOR INFORMATION

Lynda Wayne
Lwayne@GeoMaxim.com
GeoMaxim / US Federal Geographic Data Committee

An earlier version of this paper was published in the Proceedings of the 2005 GISPlanet Conference Proceedings.