Development of Data Standards for a Large Government Bureau

Thomas A. Chatfield and Barb Kett
US Bureau of Land Management

ABSTRACT

Enterprise GIS implementation within the Bureau of Land Management (BLM) has required the development of data standards for themes as they become 'National Data Layers'. As with many organizations, GIS technology within the BLM developed over time in a distributed environment. Creation of data standards during that time was limited by the lack of a centralized IT organization. Within the Enterprise Architecture, the BLM is establishing national standards for geospatially enabled data.

The BLM has created processes for data standards development for all types of data, whether spatially enabled or not. The process normally calls for consensus on a logical data model (LDM). Recognition of the difference of spatially enabled data occurs with the identification of geospatial features and the relationships between those features within the LDM. The standards process is completed by creating physical database models based on logical models and business rules, and implementation of Feature Level Metadata.

Introduction

The US Bureau of Land Management (BLM) is an organization of about 10,000 employees who are widely distributed mostly throughout the western United States. It has 12 state offices and approximately 150 field offices. While the BLM has a long history of geospatial activity, it has only been connected with a network bandwidth capable of handling the movement of geospatial data for a relatively short time.

This limitation, along with the independent can-do attitude of the field office level personnel, has led to deployment of GIS Technologies in a distributed environment. As result of that environment and the fact that each office develops and relies on its own Resource Management Plan, data standards were neither developed nor shared. Also, work tasks and other data were not shared above the field office level, due to the amount of effort required and perceived lack of driving need.

The interest in creating shared data sets has been sparked because the current view of natural systems and the means to gain efficiencies by sharing data and methods is greater than the area of a single field office. In some cases, these shared datasets are still at the local-level between field offices. In other cases, they are state-level or national-level datasets.

Background

Since 1999, there has been a growing trend toward viewing data and technology from an Enterprise Architecture (EA) perspective. The Federal Government and many individual agencies within the Federal Government have implemented EA programs. The BLM started its Bureau Enterprise Architecture (BEA) in 1999. While there is a definite push from upper management due to the growth of these efforts, it has been recognized that the BLM needs to be proactive in getting its own architecture in order before higher levels of government organizations dictate that structure to BLM.

For 20 years, the BLM has been involved in data standardization efforts, but there was renewed interest as EA became stronger. The early activities were geared primarily toward alpha-numeric data. Geospatial data was acknowledged, but little effort was made to work specifically on geospatial data standards. In response to BLM programs seeking national data layers, the BLM started working on geospatial data standards at the national-level in 2004.

There have been numerous attempts to create data standards at the state office level, but not at the national-level in the geospatial business area. In fact, some states were able to proceed on their own and develop state-wide data layers and data standards. However, not all states had the resources to do this. The mismatch of expertise and standards has hampered the national-level standardization process. The different standards (or lack thereof) meant that anytime a 'national view' of the data was required, a GIS specialist would have to 'mash' the data together, often losing valuable information in the process, due to lack of standards explaining what was being analyzed. This includes such issues as attributes and a mixing of values in individual attribute fields.

The geospatial standard development followed the same processes of existing BLM data standard development but added a section. There are four major sections of the process for the development of data standards within BLM: propose, adopt, implement, and maintain. After the standard has been implemented, it is posted into the BLM Corporate Metadata Repository (CMR) which is used to document data and the major systems within BLM, making information available for re-use.

The existing data standard development process is:

Propose Data Standard
1. Business community requests the National Data Steward to provide a new data standard.
2. The National Data Steward appoints a data standard adoption team leader and data standard adoption team.
3. The team develops a proposal for the new data standard.
4. The draft proposal is evaluated by the business community, State Data Administrators, and others, and is finalized by the data standard adoption team.
5. A request for approval is routed to management of the benefiting activities.

Adopt Standard
1. The data standard adoption team researches, drafts, and completes a draft data standard by going through the steps approved in the proposal.
2. The data standard adoption team produces a draft Data Standard Report.
3. Reviewers, including the business community, evaluate the draft data standard.
4. Evaluation results provide the basis for preparing a final standards report.
5. The final data standards report is attached to a formal Instruction Memorandum and is then sent through the approval process.

Implement Data Standard
1. The data standard adoption team devises a practical implementation approach, in close cooperation with the National Data Steward and all stakeholders.
2. The team develops a formal implementation plan.
3. The implementation plan receives widespread review by the established list of stakeholders.
4. The final implementation plan is submitted for BLM management approval and funding. The plan is implemented by an Instruction Memorandum.
5. The approved implementation plan is executed under leadership of the National Data Steward.

Maintain Data Standard
1. Evaluate adequacy and usefulness of the data standard by soliciting and analyzing feedback on how well the standard is working.
2. Evaluate requests for changes to determine whether they are valid and whether they represent a need to modify or maintain the data standard.
3. Manage data standards modification (requires changes to the structure or definitions) by beginning at the "Propose Data Standard" process.

4. Perform data standards maintenance (limited to minor changes that do not cause re-collection or edits to data).
5. Standards that are no longer needed or no longer suitable for the maintenance process will be formally retired, by Instruction Memorandum, and archived along with their maintenance history and records associated with their development.

The following additional geospatial process requirements have been identified to provide additional information required for geo-spatially enabled data.

1. Identify the spatial features within the entities.
2. Determine the form of the spatial entities (Point, line, polygon).
3. Develop a physical representation of the data.
4. Implement the Feature Level Metadata as appropriate.

One major aspect of authorizing data standards within the BLM is the use of Data Stewards. Data Stewards are designated from the business side of the organization rather than from the geo-technical side. Data is viewed as belonging to the business, not to a GIS person who is responsible for the maintenance of the data. While some Geospatial data themes fall within a single business area, many more cross business areas and require much more cooperation to develop and implement standards. Once the data stewards are identified, they agree to take the responsibility for the data theme and approve the issuance of the data standard for the organization. The need for data stewards who understand geospatial data along with business needs is forcing us to bridge some gaps that have not previously been bridged. While there may be a single person from the business side of the organization responsible for the data standard, BLM is also looking at trying to provide support to that person from the geo-technology side of the organization.

Issues
There were several issues that presented themselves soon after geo-spatial data folks started using the standardization process. These issues led to potential changes to the existing data standardization process, in order to adapt it to meet the needs of spatial data.

The placement of the data standards in the Corporate Metadata Repository (CMR) was one such process requiring modification. The CMR application and database were designed to handle only alpha-numeric data. The attribute information about a feature can be documented in the CMR, but the feature itself can not be documented.

Another process modification included the development of 'standard' documents and the alteration of one form. This has enabled creators and readers of multiple data standards to more easily prepare and review the data standards presented.

The domain values were originally included in the standards document. Through the process we have discovered that when those values are subject to frequent change, the standard document must also be updated. A decision was been made to allow the values to be maintained separately and referred to in the data standard.

Another issue was: What is a data standard and what should it contain? For BLM, the purpose of the data standard is to identify and clearly define the required data and relationships between that data. The requirements include definitions for all of the entities (features) and the associated attributes (fields).

Forcing users to create real definitions for entities and attributes that are not circular is a large task by itself. An example of a circular definition is: "GIS Acres: GISAcres are the acres from the GIS system." Requiring definitions to be more precise has helped clarify where attributes are duplicated and where they are truly different.

BLM standards include the following for attributes:
- Naming standards, including a representation term (and a unit of measure if appropriate)
- Data type, length and format
- Definitions must be clear and concise
- Is it required or optional, Null values allowed, default values
- Domain (where applicable)

Having clearly defined and standardized attributes allows the BLM to seamlessly merge the data together from various datasets when required (with the exception of the feature to feature topology, which is being addressed by the business process group).

What is not a data standard?
- A physical Database design
- Calling it a shapefile, coverage, or geo-database
- An XML Schema used to generate a database

The GIS Perspective about data and data management has historically been about physical implementation and a data standard is not the same as an implementation. GIS users immediately slip into talking about the features and attributes that are required for those features. The BLM is trying to approach the geospatial data standards development from a more disciplined data perspective. There is an "implementation standard" that is appropriate. However, the data standard and the implementation standard are not the same thing.

The existing BLM process for development of data standards and the parent DOI process require a Logical Data Model for each data standard. Because the process for the development of data standards came from the business world and started with a focus on data, the process included the development of a logical data model before any standard could be approved.

A Logical Data Model:
- Is a logical representation of the business data requirements of an organization
- Is Independent of hardware or software constraints
- Describes what data is required and how it should be organized, rather than what operations will be performed on the data
- Provides a foundation for data control with specific and accurate data definition

There seem to be a limited number of geospatial technology people who understand the real differences between a logical data model and a physical implementation. This appears to be driven by the very physical data nature of geospatial technologies. A logical data model is not the same as an inventory or a representation of the physical database. It is a unique view of the data and is useful from the business perspective to ensure the data needed is all captured with the data's relationships for business processes.

The development of the logical data model is a process where one has to ignore the fact that there are underlying features (points, lines, and polygons) that will represent these objects. The strength of the logical data model is the understanding of the relationships between the objects that the business has identified it needs. While many of these relationships get further defined in the physical implementation of the model, the understanding of the relationship is the important function.

The BLM has developed a term for a view of the logical model that we use to help bridge the gap between the logical and physical views of the data, the 'physio-logical' model. We use this view to help the geo-spatial staff to better see the links between the logical model and the more familiar features of the GIS.

Figure 1, below, is an example of a logical data model and Figure 2, also below, is an illustration of a geospatial data model produced as a result of the amended process. The logical data model shows generically defined 'boundaries' because there are so many types of boundary data within BLM. Figure 2 specifically shows the 'physio-logical' of the National Monument boundary.

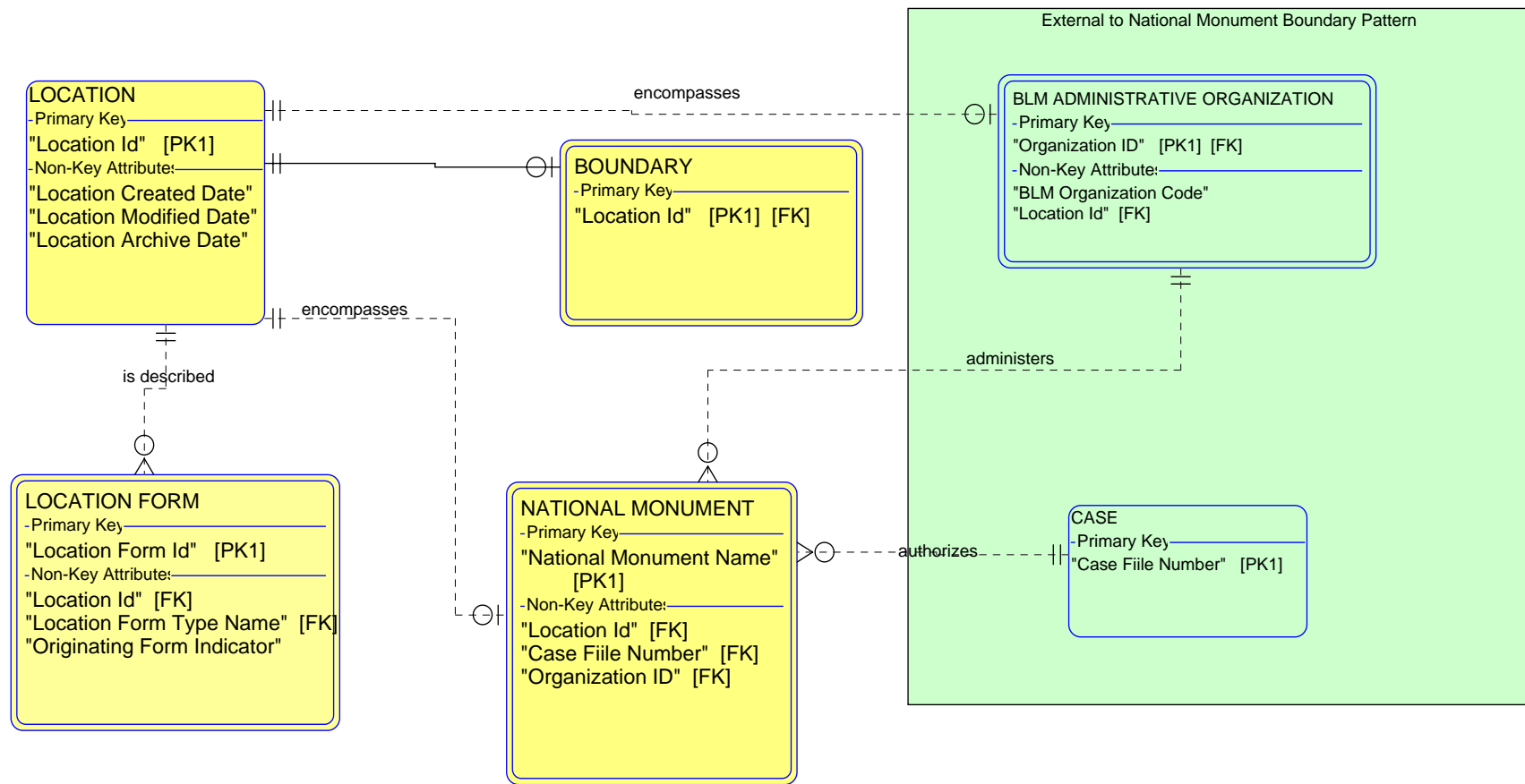National Monument Boundary 5/4/2007 version 1 DRAFT

External to National Monument Boundary Pattern

LOCATION
- Primary Key
"Location Id"   [PK1]
- Non-Key Attributes
"Location Created Date"
"Location Modified Date"
"Location Archive Date"

BOUNDARY
- Primary Key
"Location Id"   [PK1] [FK]

BLM ADMINISTRATIVE ORGANIZATION
- Primary Key
"Organization ID"   [PK1] [FK]
- Non-Key Attributes
"BLM Organization Code"
"Location Id" [FK]

encompasses

encompasses

is described

administers

LOCATION FORM
- Primary Key
"Location Form Id"   [PK1]
- Non-Key Attributes
"Location Id" [FK]
"Location Form Type Name" [FK]
"Originating Form Indicator"

NATIONAL MONUMENT
- Primary Key
"National Monument Name"
    [PK1]
- Non-Key Attributes
"Location Id" [FK]
"Case Fiile Number" [FK]
"Organization ID" [FK]

authorizes

CASE
- Primary Key
"Case Fiile Number"   [PK1]

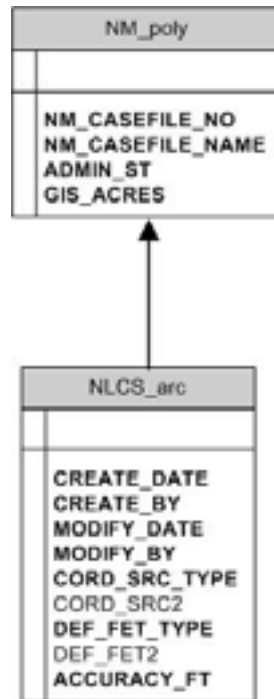Figure 1 - National Monument Boundary Logical Data Model

8

Figure 2 –
National Monument Boundary Physical Data Model

Part of the effort to develop geospatial data standards has been driven by the Data Quality Act (DQA). (Public Law 106-554 - Section 515 of the Treasury and General Government Appropriations Act for FY 2001.) On February 22, 2002, the OMB's Office of Information and Regulatory Affairs (OIRA) published the final version of its *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies* (DQA Guidelines). *See* 67 F.R. 8452. As provided in the DQA, the Guidelines mandate that each Federal agency:

1. By October 1, 2002, issue its own information quality guidelines ensuring and maximizing the quality, objectivity, utility, and integrity of information that it disseminates;
2. Establish administrative mechanisms to allow affected persons to seek and obtain correction of information maintained or disseminated by the agency that does not comply with OMB or agency guidelines;
3. Report periodically to OMB the number and nature of complaints received by the agency regarding the accuracy of its information and how such complaints were resolved.

Because of the DQA guidelines, a review of the BLM implementation of geospatial metadata was necessary. The BLM has used the Federal Geographic Data Committee (FGDC) metadata model for several years. Recently, it has found that the Content Standard of FGDC lacks the ability to track some information at a detailed enough level to respond to the DQA Guidelines.

The FGDC Metadata Content Standard has limitations about the information that is input into the system. If only one of 10,000 polygons in a dataset is updated and edited by one data maintenance person, should the entire dataset be listed as having been updated? Or should updates and edits be tracked at a feature level that will show which portions of the data have been updated and which portions have not? This metadata at the feature level would allow the Data Steward an opportunity to notify certain data maintenance staff to review their portion of the data. Twenty year old data may not have changed, but it should be reviewed for data quality at some point.

What is Feature Level Metadata (FLM) and how does it help the BLM? Accuracy is one potential component of a data dictionary. Attributes can be defined to keep track of "feature-level metadata" (information specific to features or attributes such as when they were entered, updated, by whom, by what method, and quality of method). For instance, road networks change over time as new roads are built and the data is updated. It is helpful to define attributes for tracking which roads were added, when they were added, how they were added, and other information. Keeping track of changing features in any dynamic dataset is helpful in maintaining and assessing data quality.

Feature-level metadata is also useful when datasets are comprised of more than one source of information, especially if the sources are of varying data quality. If any concerns about the dataset's quality are brought up in the future, the portions of the dataset whose sources are in question can be easily identified and isolated.

Since few of today's more complex datasets are developed in a single effort at a single point in time, it is helpful to be able to track which portions of the data come from which sources and at what time. Attributes for the source and the date of creation or date of last change/edit are useful in this area.

It is has been a slow process and difficult to obtain momentum. But with numerous efforts coming from both inside and outside the agency, such as the Land Use Planning effort, the migration of several states to a more centralized IT structure, and the implementation of Citrix capabilities, more people are sharing data and applications which is helping to drive the need for data standards.

The BLM Enterprise GIS (eGIS) project has assisted in getting several state offices migrated to a more state centric geospatial infrastructure. This is a first step toward some form of centralized database. It is also a contributing factor toward the push for development of data standards. This work is tiered down from the Federal Enterprise Architecture (FEA) Geospatial Line of Business and the DOI Geospatial Blueprint. In both of those efforts there is a great deal of emphasis on data standards and data quality. Data is one of the 5 cornerstones of the FEA.

There are also several data layers that have been identified as required to geo-enable some national applications. One of these data layers is the Wild Horse and Burro Herd Areas and Herd Management Areas. Since the data layer is being elevated to a 'National Data Layer' on a central national level server, a data standard is an essential piece of the puzzle to allow its usage by other applications.

The standards process is not just handling data. It is also about standardizing work processes and products for not only internal use but external use. Standards work at BLM is also going beyond just creating a document. Creating a document and handing it over to be implemented is a poor way to communicate the true goals. As with many activities within large organizations, if the standards document is created and not followed up, it will wither and die from a lack of maintenance and use. There are now other standards for map disclaimers, map templates and collars, symbol sets, data disclaimers, metadata keywords standards.

Where will BLM geo-spatial data standardization go from here?
We plan to continue to develop data standards for BLM datasets and to participate in the development of data standards for the OMB Circular A-16 authoritative data sources. BLM is the responsible agency for Cadastral data and Federal Land Ownership data based on the A-16 Circular. There are also a number of data layers to be completed that are of high value to BLM, but have limited value to other users since they are related BLM work processes and are not in the realm of 'reference data.'

      Examples:
      Grazing Allotments

Wild Horse and Burro Herd Areas and Herd Management Area
National Landscape Conservation System Boundaries

BLM has been involved in developing data standards for other agencies with the same data layers. Some examples of these datasets include: Inter-Agency Trails Data Standards (ITDS) and Aerial Photo Index.

Beyond the issues currently identified, there are some additional issues that, while important, are not necessarily a part of developing data standards. These issues have more to do with the transfer of information about the data standard and how it is to be implemented and maintained. The development of the data standard is not enough. There is also a need for:

1. Implementation instructions
    i. Definitions in a logical data model will not be as specific as the field definitions listed as choices in the Arc Catalog (ESRI) listing.
    ii. Information about database topology
    iii. Instructions for the creation of the geo-database
2. XML Schema for the generation of the geo-database
3. Review and update process

Summary
The BLM has been working toward the development of geospatial data standards and is continuing to define the areas that need to be addressed. The standards development process will continue to evolve as more data standards are developed. It is expected that the development of standards will be easier as the user audience becomes more knowledgeable. This has been demonstrated by the increased interest and questions about standards.

Acknowledgements

The authors would like to thank Christine Hawkinson for her contributions in modeling and helping to define some of the processes used in standards development. We would also like to thank Heather Karlson for her editing.

Author Contact Information

Thomas A. Chatfield
Spatial Data Architect
US Bureau of Land Management
Denver Federal Center
Bldg 40, PO Box 25047
Denver, CO 80225-0047
Ph: 303-236-1936
Fax: 303-236-1981
tom_chatfield@blm.gov

Barb Kett
Data Architect
US Bureau of Land Management
Denver Federal Center
Bldg 40, PO Box 25047
Denver, CO 80225-0047
Ph: 303-236-4025
Fax: 303-236-1981
barb_kett@blm.gov