Extending Geospatial Data Standards in the BLM with Implementation Guidelines

Thomas A. Chatfield and Lorena H. Creely

U.S. Bureau of Land Management

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ABSTRACT

While developing National data standards, BLM has ascertained that these data standards are more consistently applied when supplemented by Implementation Guidelines. BLM has found that Implementation Guidelines documentation is needed to ensure that the standards are applied as desired. While this advances beyond the logical data modeling required for standards development to the area of physical database design, the removal of the existing barriers between the two have become necessary to ensure that data structures are standardized (especially in the geospatial realm).

The Implementation Guidelines documentation includes a listing of domain values, creation of an empty geodatabase, and the application of geodatabase topology rules. As an example of this, the data standards process within BLM and, specifically, one of the BLM Administrative boundary data standards and geodatabase will be discussed.

Introduction

Development of national-level geospatial data standards has been ongoing at the U.S. Bureau of Land Management (BLM) for several years. The BLM is an organization of about 10,000 employees who are widely distributed, primarily throughout the western United States. The Bureau has 12 state offices and approximately 150 field offices. While the BLM has a long history of geospatial activity, the process of developing and implementing national data standards is a recent undertaking. BLM personnel have noticed that even when standards are developed within the Bureau, variations in implementation of the standard occur due to incongruous operating systems and database management systems, as well as local interpretation issues. This includes such things as inconsistency in text case in field names, variations in field definitions, data format, and domain values. To reduce the 'lack of standards' within the data standards, BLM has extended the development process to include implementation guidelines documentation.

Background and Issues

As a large organization with multiple existing data stores, development of data standards in the Bureau is difficult at best. Even after data standards are created, implementation of those standards can become a logistical nightmare.

Wide distribution of BLM staff is one reason why data standardization is a challenge. Spatial data is often acquired and/or developed on a state-by-state basis, and in many instances, data collected for a program or project is disparate even between neighboring field offices. Lack of coordination between offices creates redundant data, or similar data that is difficult and costly to reconcile, translate, and share. Without the structure of a national data standard, each state and field office does their own thing.

Various people with diverse roles in the organization have different views, perceptions, and levels of understanding about BLM data and what a data standard should look like. Department of Interior (DOI) and BLM policy requirements only extend through the logical model. Data Administrators and Data Stewards understand the data standard at this level of Logical/Business Rules and Requirements. GIS personnel will understand it at the Physical/Implementation level. Time and again, we have received comments such as, "*I am not a database designer and so I am unfamiliar with the notations that are used on the logical model to relate one table to another. I need a legend to tell me what these symbols mean*." Or, "*This proposal fails to provide any information on the nuts and bolts of implementation. How will this proposal help the field offices complete our work*?"

The challenge becomes, how do we reconcile the disparate perspectives and concerns so that everyone in the organization feels like their needs are being understood and addressed?

The Bureau has identified several stumbling blocks attributable to differences in the knowledge and experience between those with a data administration background and those with a GIS background. Personnel in data administration roles are almost always from an alpha-numeric data background, and thus may not understand the unique requirements of geospatial data storage formats as a GIS is implemented. They don't always comprehend how the value of tabular data can be extended with a spatial component, how features are identified from data entities, what forms these features may take and how they relate to each other. Yet they need to have some level of knowledge of what the actual geospatial information will look like. Conversely, GIS personnel are rarely trained in logical data modeling, but they need to have a basic understanding of the business requirements as a driving force behind the need for standardization in the first place.

The process that the BLM has developed, under guidelines/requirements from the DOI, helps to ensure that key personnel with different areas of expertise have opportunities, at different points in the process, to have input into what should go into a national data standard.

There are three required pieces of documentation for establishment of a BLM national data standard. The documents are:

1) Data Standard Proposal: The BLM business community requests the appropriate National Data Steward to provide a new or revised data standard. The steward identifies a team who develops a proposal for the new standard. This proposal sets out all the information needed by management to decide whether the effort should go forward. The draft proposal is evaluated by the business community, State Data Administrators, and others.

2) Data Standard Report: Approval of the proposal triggers the steps to research, coordinate, draft, obtain consensus, and complete a data standard, which is documented in a Data Standard Report. This report publishes the new standard at a Logical/Business needs level. The subsequent reviews may yield recommendations to expand or redirect the scope of the standard.

3) Data Standard Implementation Plan: The implementation plan provides the Bureau with the information about who the responsible parties are and what the expected timelines and impacts are for the actual implementation of the data standard.

Even though the DOI guidelines actually only require the above three documents, we have found that these alone are not enough to ensure that the standard is accurately and consistently implemented. We take it a step further with the inclusion of:

4) Data Standard Implementation Guidelines: After the final data standard has been approved, the data standard adoption team coordinates the creation of a practical physical implementation approach, in close cooperation with the National Data Steward and all stakeholders. The Implementation Guidelines explicitly spell out the suggested physical structure of the database, including tables, attributes, a listing of domain values, creation of an empty geodatabase, and the application of geodatabase topology rules.

Without these Data Standard Implementation Guidelines, we often have a mismatch in interpretation of the logical model. Some states/offices will interpret the standard one way, while others will do something contradictory to suit their specific needs, or just because “that’s the way it’s always been done.” Some of the various database management systems used by the Bureau store and interpret data in conflicting or incompatible ways. Implementation Guidelines bridge the gap between the logical model and the physical database design, paving the way for better data quality, integrity, and usability (Figure 1).

In addition, we have found that BLM GIS personnel are oriented to physical database design, and have difficulty implementing the data standard from the logical model-based Data Standard Report. Having the Implementation Guidelines document with its physical data orientation allows those people to provide in-depth comments and assist in improving the logical data model and standard based on what is really occurring in the field.

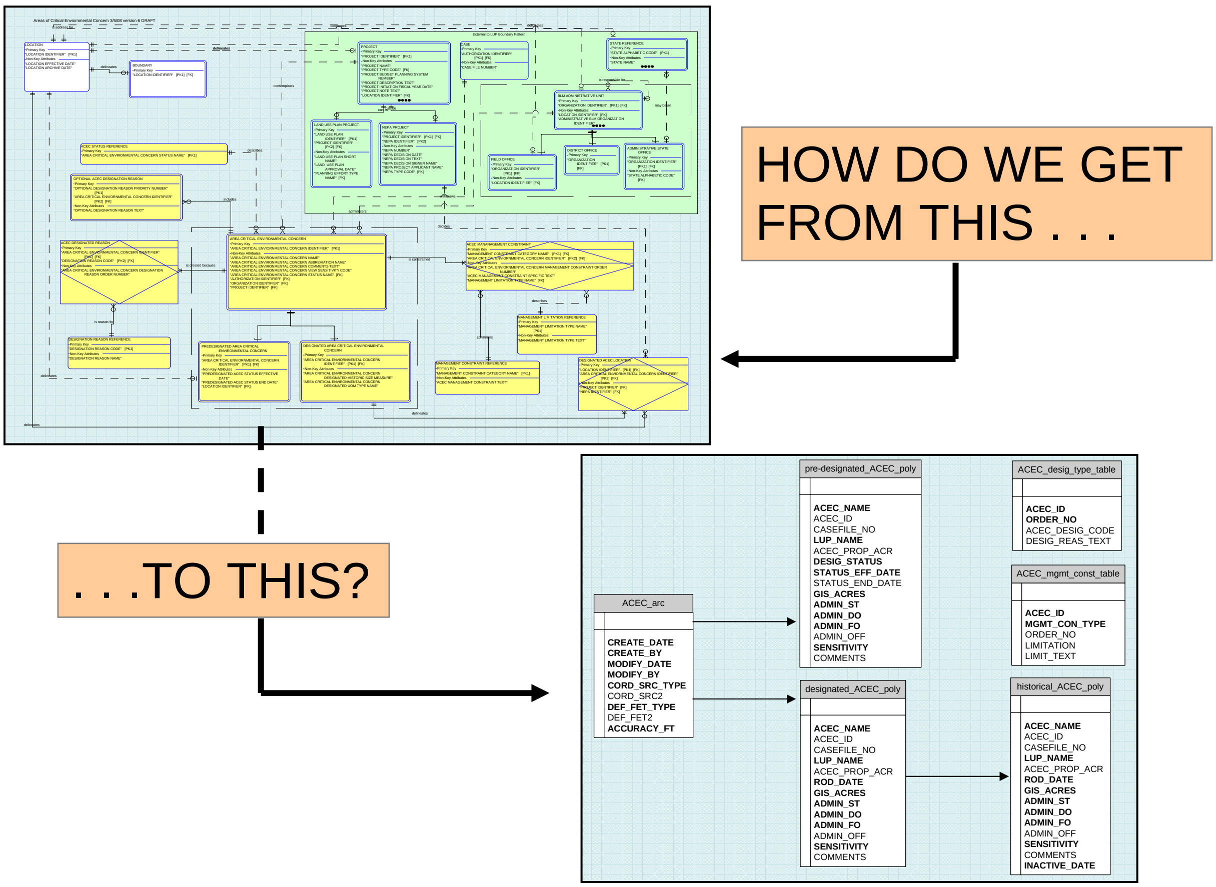


Figure 1: Logical Data Model to Physical Database Design

The ultimate goal of the BLM geospatial data standards effort is to create national data standards, wherein each state/office maintains core data in the same format with the same attributes, etc. These data layers can then be merged and maintained in a centralized location within the Bureau. BLM is evaluating methods for the maintenance of those national data sets by spatial segregation of data editing privileges. The deployment of centrally hosted, common editing environments (via Microsoft Terminal Services or Citrix) and geodatabase replication are two technologies that are being examined to support this business requirement.

The following examples illustrate the process of taking a data standard from the logical model to the physical database design, and the resultant Implementation Guidelines. As there are many different methods and means to achieve the same results within geospatial data structures, the natural variations in the development of geospatial datasets must be addressed.   The following examples show some of the methods that BLM has developed to ease the transition and structure the data standards implementation.

Examples

Following are examples taken from the development process for a data standard for Areas of Critical Environmental Concern (ACEC) boundaries (Figure 2). ACEC's are designated areas where special management attention is needed to protect important historic, cultural and scenic values, fish or wildlife resources or other natural systems or processes, or to protect human life and safety from natural hazards. The "ACEC" designation indicates that the BLM recognizes that an area has significant values and has established special management measures to protect those values.

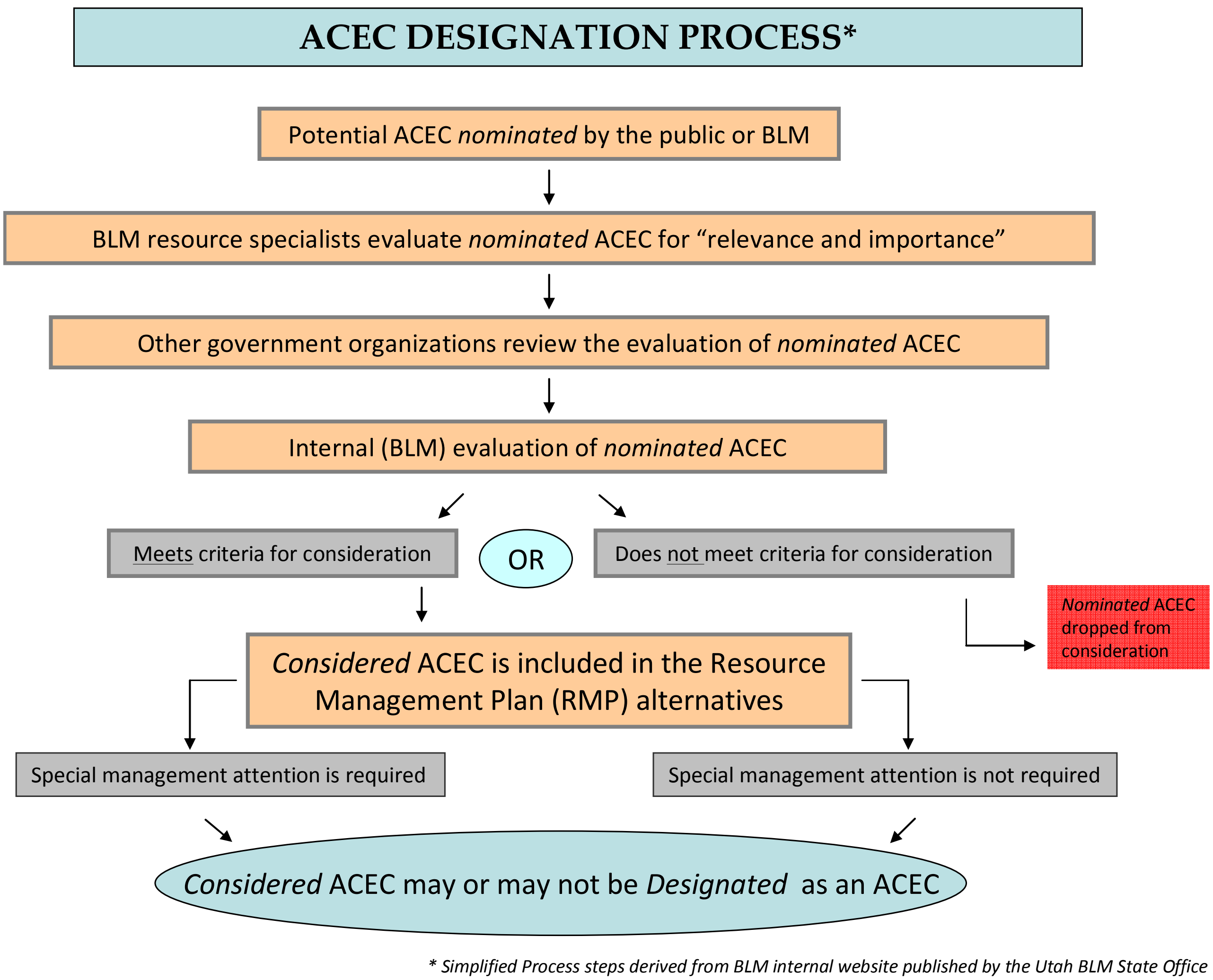


Figure 2: ACEC Designation Process

[Note: The following examples represent small pieces of the ACEC logical model and the corresponding transition to physical implementation. Logical model and attribute table examples have been simplified for purposes of illustration.]

Example 1: Collapsing Entities and Attributes

1a) While making the transition from a logical model to a physical implementation, one invariably has to decide which entities and attributes from the logical should migrate to the physical. Sometimes it makes sense to collapse attributes from complex parts into something simpler.

In the ACEC example, the logical model has an entity “PROJECT,” along with two other entities “LAND USE PROJECT” and “NEPA PROJECT” (Figure 3). These second two entities have a one-to-one relationship with the first entity (PROJECT). Each ACEC is linked to one and only one Land Use Plan (LUP), and one and only one NEPA (National Environmental Policy Act) project. Because of these simple one-to-one relationships, the attributes “LAND USE PLAN NAME” and “NEPA DECISION DATE” from the PROJECT, LAND USE PROJECT, and NEPA PROJECT can be collapsed into the ACEC Polygon table.

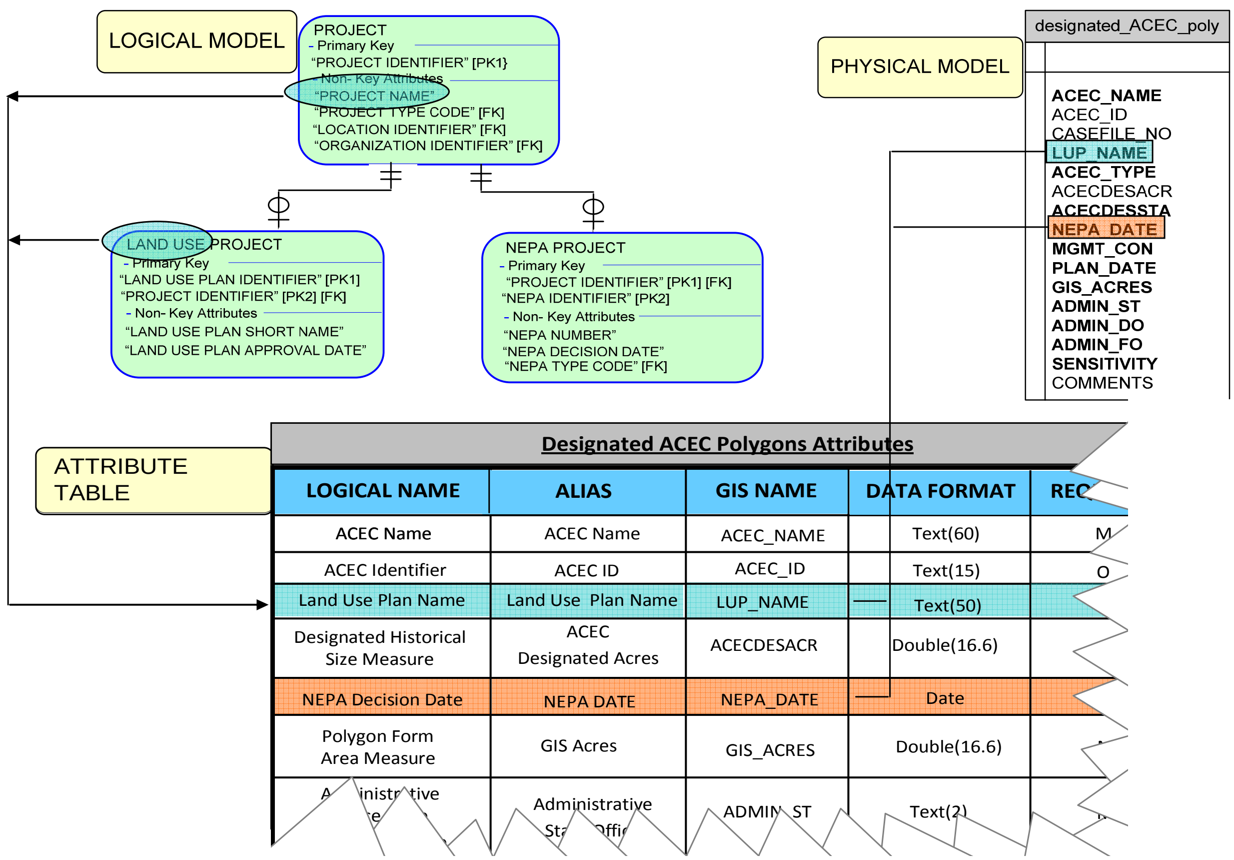
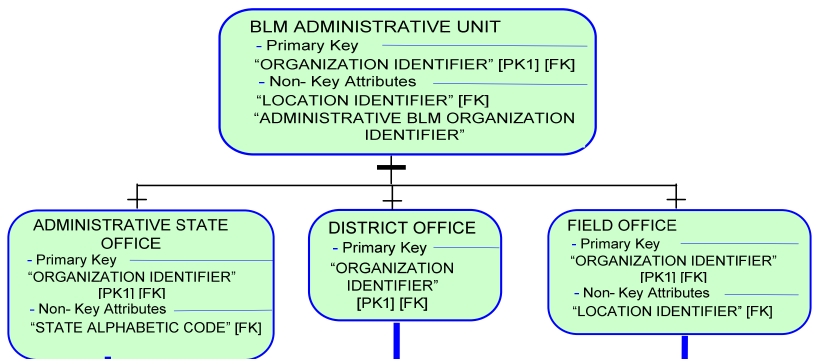


Figure 3: Collapsing Entities and Attributes

1b) Conversely, while it might sometimes initially make sense to collapse attributes from several entities together, it may be discovered later that this strategy will not work in the real-world physical implementation. An example of this concept is shown in the case of the BLM Administrative Unit entities of “FIELD OFFICE,” “DISTRICT OFFICE,” and “STATE OFFICE” (Figure 4). Because each ACEC is administered by one-and-only-one administrative office (either at the state, district, or field level), it originally appeared to make sense to collapse the codes for these entities into one 7-character concatenated attribute in the ACEC Polygon table. When we initially proposed this structure, the feedback we got from those who were actually using this data was that, for various reasons, they needed for each BLM office level to be its own attribute in the table. Thus, the table does include all three separately. Because the value for any given field or district office code is not necessarily unique in-and-of itself, we needed to include the concatenation as well to create unique values.



Designated ACEC Polygons Attributes

LOGICAL NAME	ALIAS	GIS NAME	DATA FORMAT	REQUIREMENT	ALLOW NULLS?	DEFAULT VALUE
ACEC Name	ACEC Name	ACEC_NAME	Text(60)	M	No	UNK
Administrative Office.State Alphabetic Code	Administrative State Code	ADMIN_ST	Text(2)	M	No	
District Office.BLM Organization Code	Administrative District Office Code	ADMIN_DO	Text(2)	M	N	00
Field Office.BLM Organization Code	Administrative Field Office Code	ADMIN_FO	Text(3)	O	No	000
N/A	Administrative Office Code	ADMIN_OFF	Text(7)	O	Yes	XX00000
ACEC View	View Sensitivity Code	SENSITIVITY	Text(10)	M	No	UNK

Figure 4: Keeping Entities Separate

Example 2: Logical Data Model vs. Physical Database Structure

2a) As illustrated in Figure 2, there are three stages in the process when an area identified as a potential area of critical environmental concern officially changes status (“Nominated,” “Considered,” and “Designated”). The area is first Nominated as a possible ACEC. From this point it can either be Considered or not. If it is not Considered, it goes no further in the process. If it is Considered, it may or may not be Designated as an ACEC.

In the original logical model, a single entity (ACEC) with an attribute of “ACEC Designation Status” appeared to cover the needs for the structure of the data. Attached to the Designation Status attribute was a Domain with the codes D)esignated, C)onsidered, and N)ominated (Figure 5).

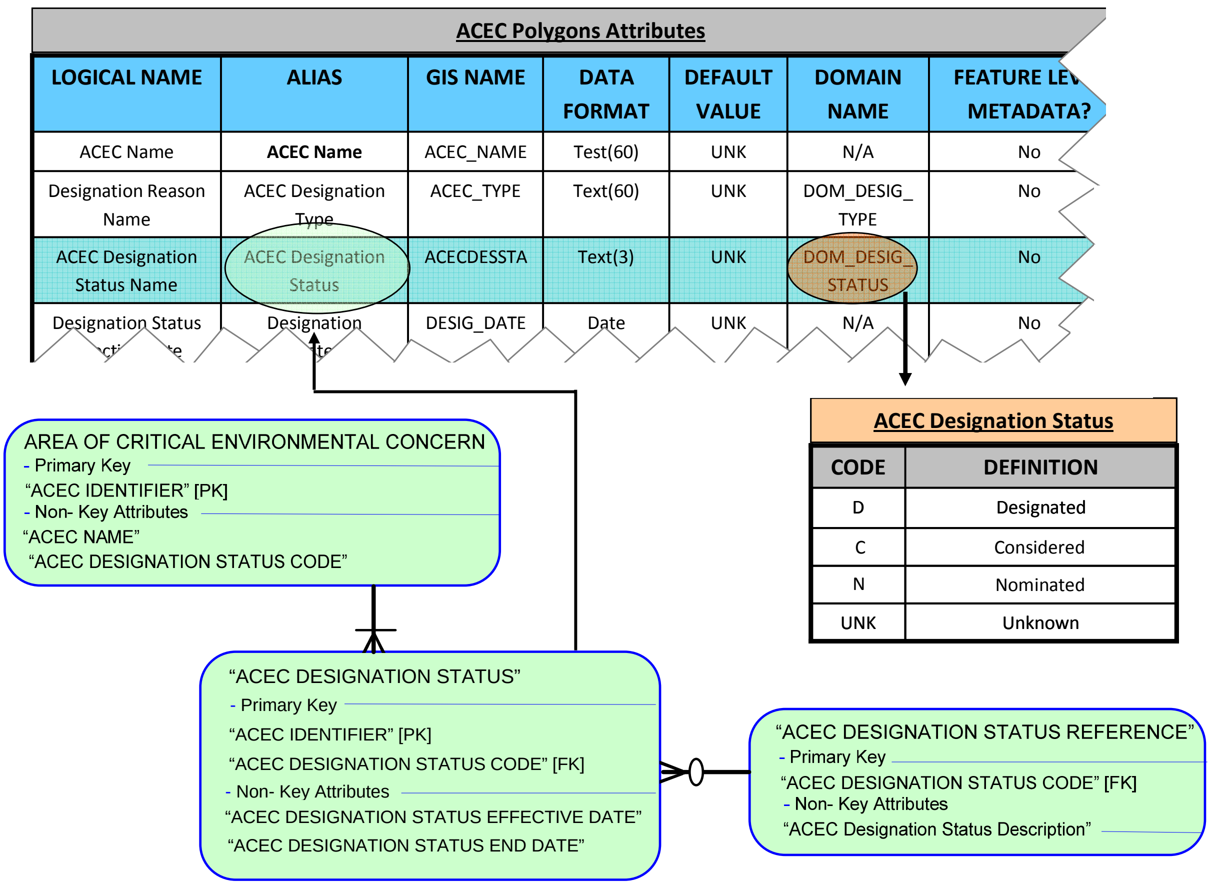


Figure 5: Logical Data Model vs. Physical Database Structure

2b) When the initial draft of the logical model was published for review from the field, we came to realize, through their comments, that our initial model would not suffice. The logical model was revised, making the ACEC entity a super type with two sub-types, Pre-Designated ACEC and Designated ACEC. There were many attributes in common between the two, but others were specific to only one or the other. One very important distinction was that only the Pre-Designated subtype had attributes for Status Effective and Status End Dates, referring to the status being N)ominated or C)onsidered which is due to BLM Policy (Figure 6).

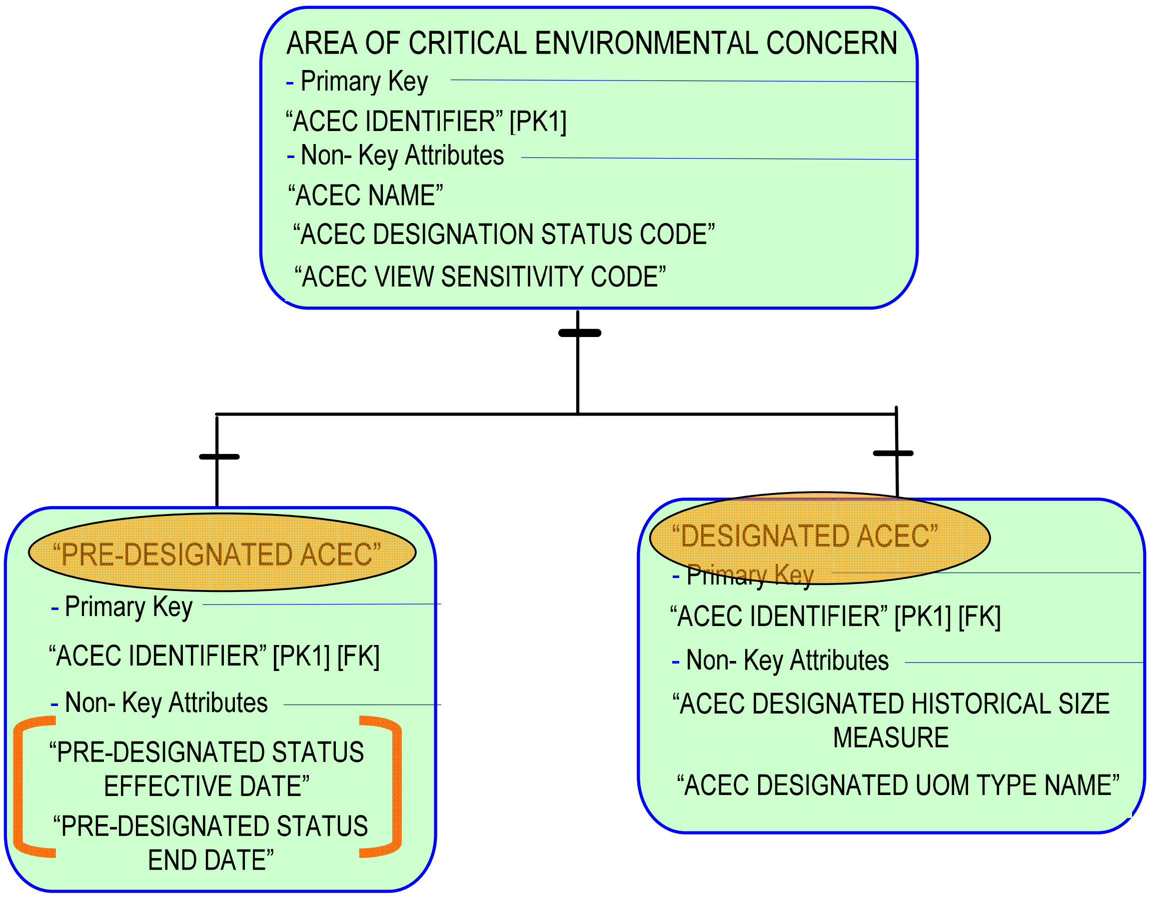


Figure 6: Logical Data Model vs. Physical Database Structure

During the development of the physical database design, we ran into an issue that required us to break the Pre-Designated and Designated ACEC subtypes out to be separate and distinct entities. The boundaries of a given Designated ACEC might be different than the boundaries were in the Pre-Designated state. The Designated ACEC may not be edited or changed without a change in the Land Use Plan that created it, while the Pre-Designated features may be edited during the planning process. The two different states of Pre-Designated ACEC's (Nominated or Considered - Not Nominated) could then be differentiated based on the single "ACEC Designation Status Name" attribute, with the accompanying Designation Status domain table (Figure 7).

Pre-Designated ACEC Polygons Attributes

LOGICAL NAME	GIS NAME	DOMAIN NAME	ALLOW NULLS?
ACEC Name	ACEC_NAME		No
ACEC Identifier	ACEC_ID		Yes
Polygon Form Area Measure	GIS_ACRES		No
ACEC Designation Status Name	DESIG_STATUS	DOM_DESIG_STATUS	No
District Office.BLM	ADMIN_DO		No

ACEC Designation Status

CODE	DEFINITION
C	Considered
N	Nominated
UNK	Unknown

Designated ACEC Polygons Attributes

LOGICAL NAME	GIS NAME	DOMAIN NAME	ALLOW NULLS?
ACEC Name	ACEC_NAME		No
ACEC Identifier	ACEC_ID		Yes
Polygon Form Area Measure	GIS_ACRES		No
District Office.BLM Organization Code	ADMIN_DO		No

Figure 7: Logical Data Model vs. Physical Database Structure

Even though these revisions to the model made the design of the physical implementation more complex, it ended up being a much better "real-world" representation of what happens during the process of ACEC designation.

Example 3: Simplification by Addition

3a) Physical implementation of a logical data model can sometimes become simpler and cleaner with the addition of attribute information related to, but not directly attached to, the features. Each ACEC must have at least one "reason" for designation. Initially, we came up with the domain list of 29 Designation Reasons from one of the State offices, as that was the pick-list they had been using for years.

During discussions with business subject matter experts (SME's), we learned that an ACEC can actually have more than one reason for designation. We initially allowed for implementation of this by instituting a large text field (60 characters) for the ACEC Designation Type (reason) attribute. The intent was that offices could include numerous designation types separated by commas in this field, with the primary type being listed first (Figure 8).

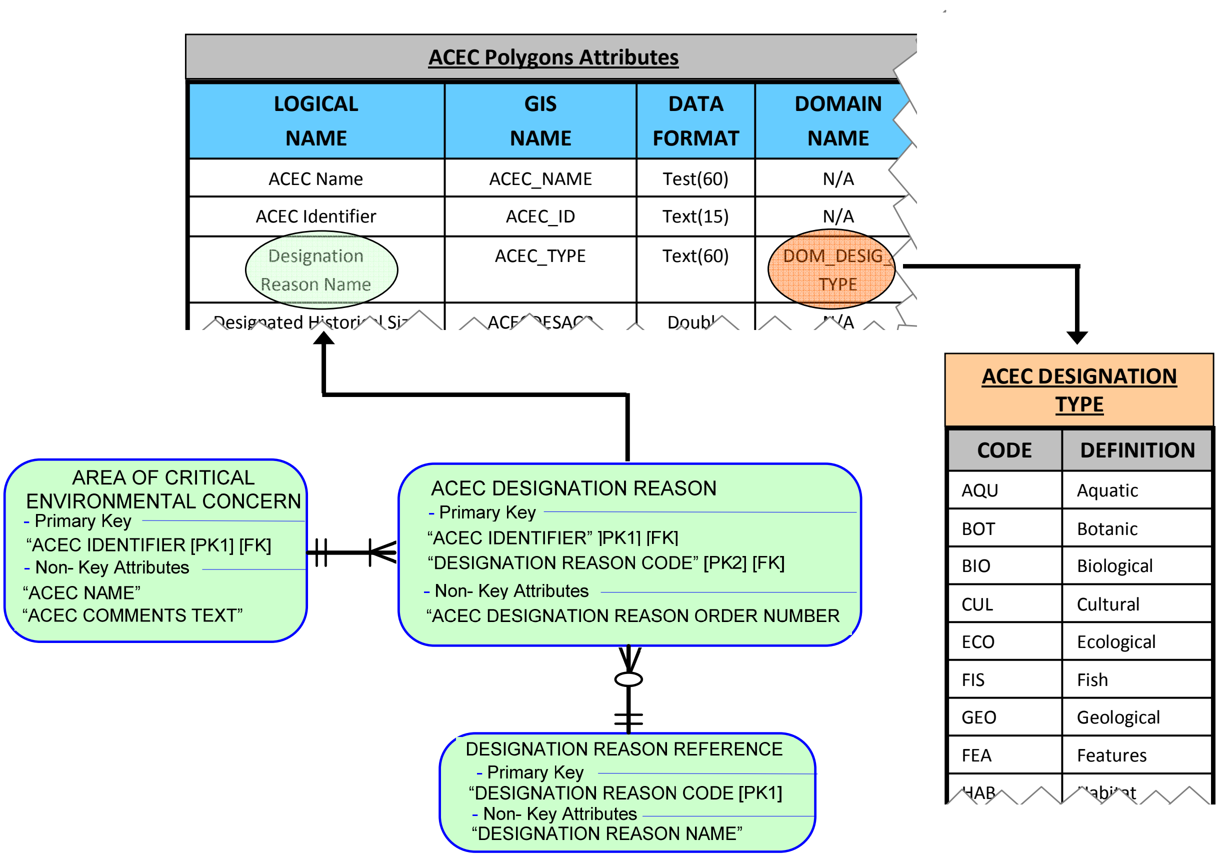


Figure 8: Simplification by Addition

3b) The comments we got from the field review was that this was a very cumbersome implementation and did not allow for any flexibility. The resolution of the one-to-many relationship was not clean and did not answer the business needs, even though at least one group was actually already using that method. We decided it made more sense to create a separate, related table, "ACEC Designation Type Table," linked to the ACEC polygon table by a unique ACEC Identifier. This provided a cleaner resolution of the one-to-many relationship, whereby each ACEC and "reason" could have its own line in the table, facilitated by an accompanying Designation Code domain for the "reason" codes. We also added an ORDER_NO attribute for ranking, and an optional comments text field. The Logical Data Model was updated to show the true business requirements (Figure 9).

Designated ACEC Polygons

LOGICAL NAME	ALIAS	GIS NAME	DATA FORMAT
ACEC Name	ACEC Name	ACEC_NAME	Text(60)
ACEC Identifier	ACEC ID	ACEC_ID	Text(15)
ACEC Casefile Number	ACEC Casefile Number	CASEFILE_NO	Text(15)
Land Use Plan Name	Land Use Plan Name	LUP_NAME	Text(50)

RELATED TABLE

ACEC Designation Type Table

LOGICAL NAME	ALIAS	GIS NAME	DATA FORMAT	DOMAIN NAME	SOURCE
ACEC Identifier	ACEC Identifier	ACEC_ID	Text(15)		Operator Input
ACEC Designation Reason Order Number	ACEC Designation Reason Order Number	ORDER_NO	Integer		Operator Input
ACEC Designation Reason Name	ACEC Designation Reason Code	ACEC_DESIG_CODE	Text(4)	DOM_DESIG_CODE	Operator Input
Optional Designation Reason Text	Designation Reason Text	DESIG_REAS_TEXT	Text(40)		Operator Input

Figure 9: Simplification by Addition

Example 4: Creating Accurate Codes

The other issue we ran into (again, from feedback from the various offices) was that the original list of 29 Designation Types were in fact not the “official” list we should be using. The official list was 8 ACEC designation “reasons,” taken directly from the Code of Federal Regulations (CFR). The CFR is our legal authority for creating ACECs. Once we had that list, it was a simple matter to create a crosswalk table for the values to fit each value on the original list of 29 into one of the 8 categories (Table 1).

<table>
<tr><th colspan="4"><u>Crosswalk Table for ACEC Designation Type Codes</u></th></tr>
<tr><th>DESIG_TYPE (Old Version)</th><th>Description</th><th>DESIGNATION TYPE CODE</th><th>Description</th></tr>
<tr><td>CUL</td><td>Cultural</td><td rowspan="2">CUL</td><td rowspan="2">Cultural</td></tr>
<tr><td>PAL</td><td>Paleontology</td></tr>
<tr><td>FIS</td><td>Fish</td><td>FRSC</td><td>Fish Resource</td></tr>
<tr><td>HIS</td><td>Historic</td><td>HIS</td><td>Historic</td></tr>
<tr><td>HAZ</td><td>Hazard</td><td rowspan="2">NHAZ</td><td rowspan="2">Natural Hazard</td></tr>
<tr><td>SAF</td><td>Safety</td></tr>
<tr><td>FEA</td><td>Features</td><td rowspan="6">NPRO</td><td rowspan="6">Natural Process</td></tr>
<tr><td>HAB</td><td>Habitat</td></tr>
<tr><td>REF</td><td>Reference Site</td></tr>
<tr><td>SSP</td><td>Special Status Plant</td></tr>
<tr><td>SYS</td><td>System Process</td></tr>
<tr><td>WQY</td><td>Water Quality</td></tr>
<tr><td>AQU</td><td>Aquatic</td><td rowspan="9">NSYS</td><td rowspan="9">Natural System</td></tr>
<tr><td>BOT</td><td>Botanic</td></tr>
<tr><td>BIO</td><td>Biological</td></tr>
<tr><td>ECO</td><td>Ecological</td></tr>
<tr><td>GEO</td><td>Geological</td></tr>
<tr><td>RIP</td><td>Riparian</td></tr>
<tr><td>VEG</td><td>Vegetation</td></tr>
<tr><td>WSD</td><td>Watershed</td></tr>
<tr><td>WTD</td><td>Wetland</td></tr>
<tr><td>SCE</td><td>Scenic</td><td>SCE</td><td>Scenic</td></tr>
<tr><td>RAN</td><td>Rare, Endemic, relicta</td><td rowspan="4">WRSC</td><td rowspan="4">Wildlife Resource</td></tr>
<tr><td>RAP</td><td>Rare, Endemic, relictp</td></tr>
<tr><td>SSA</td><td>Special Status Animal</td></tr>
<tr><td>WLD</td><td>Wildlife</td></tr>
<tr><td>REC</td><td>Recreational</td><td>N/A</td><td>Not Applicable</td></tr>
</table>

Table 1: Crosswalk Table for Designation Type Codes

The review of the Implementation Guide provided an effective feedback loop to the logical modeling process too. Since many BLM employees have little or no training in creating or reading logical data models, when they see how the standard is to be implemented, they begin to understand and provide feedback.

A complete attribute table (the Designated ACEC polygons feature class) appears on the following page (Figure 10 – see page 17).

Designated ACEC Polygons Attributes										
LOGICAL NAME	ALIAS	GIS NAME	DATA FORMAT	REQUIREMENT*	ALLOW NULLS?	DEFAULT VALUE	DOMAIN NAME	FEATURE LEVEL METADATA?	DERIVED?	SOURCE
ACEC Name	ACEC Name	ACEC_NAME	Text(60)	M	No	UNK		No	No	Operator Input
ACEC Identifier	ACEC ID	ACEC_ID	Text(15)	O	Yes	UNK		No	No	Operator Input
ACEC Casefile Number	ACEC Casefile Number	CASEFILE_NO	Text(15)	O	Yes	UNK		No	No	Operator Input
Land Use Plan Name	Land Use Plan Name	LUP_NAME	Text(50)	M	No	UNK		No	No	Operator Input
Designated Historical Size Measure	ACEC Proposed Acres	ACEC_PROP_ ACR	Double(16.6)	O	Yes	UNK		No	No	Operator Input
NEPA Decision Date	NEPA Date	NEPA_DATE	Date	M	No	9/9/9999		No	No	Operator Input
Polygon Form Area Measure	GIS Acres	GIS_ACRES	Double(16.6)	M	No	0		No	Yes	Operator Input
Administrative Office.State Alphabetic Code	Administrative State Code	ADMIN_ST	Text(2)	M	No	UNK	DOM_ ADMIN_ ST	No	No	Operator Input
District Office.BLM Organization Code	Administrative District Office Code	ADMIN_DO	Text(2)	M	No	00		No	No	Operator Input
Field Office.BLM Organization Code	Administrative Field Office Code	ADMIN_FO	Text(3)	O	No	000		No	No	Operator Input
N/A	Administrative Office Code	ADMIN_OFF	Text(7)	O	Yes	XX00000		No	Yes	Operator Input
ACEC View Sensitivity Code	View Sensitivity Code	SENSITIVITY	Text(3)	M	No	UNK	DOM_ SEN_ CODE	No	No	Operator Input
ACEC Comments Text	Comments	COMMENTS	Text(250)	O	Yes	UNK		No	No	Operator Input
* M=Mandatory O=Optional C=Conditional										

Figure 10: Attribute Table for Designated ACEC Polygons

The final physical diagram appears below (Figure11). The arc features are linked to the polygon features as they are the features that are used to create the polygons. The Designated Polygons are related to the Historical Polygons by the fact that when a Land Use plan is updated, the old ACEC polygon may be updated or changed and the business has decided that it would like to keep a record of any polygons that are no longer active ACECs. The Designation Type table and the Management Constraint table are both related to the ACEC polygon features through the ACEC ID.

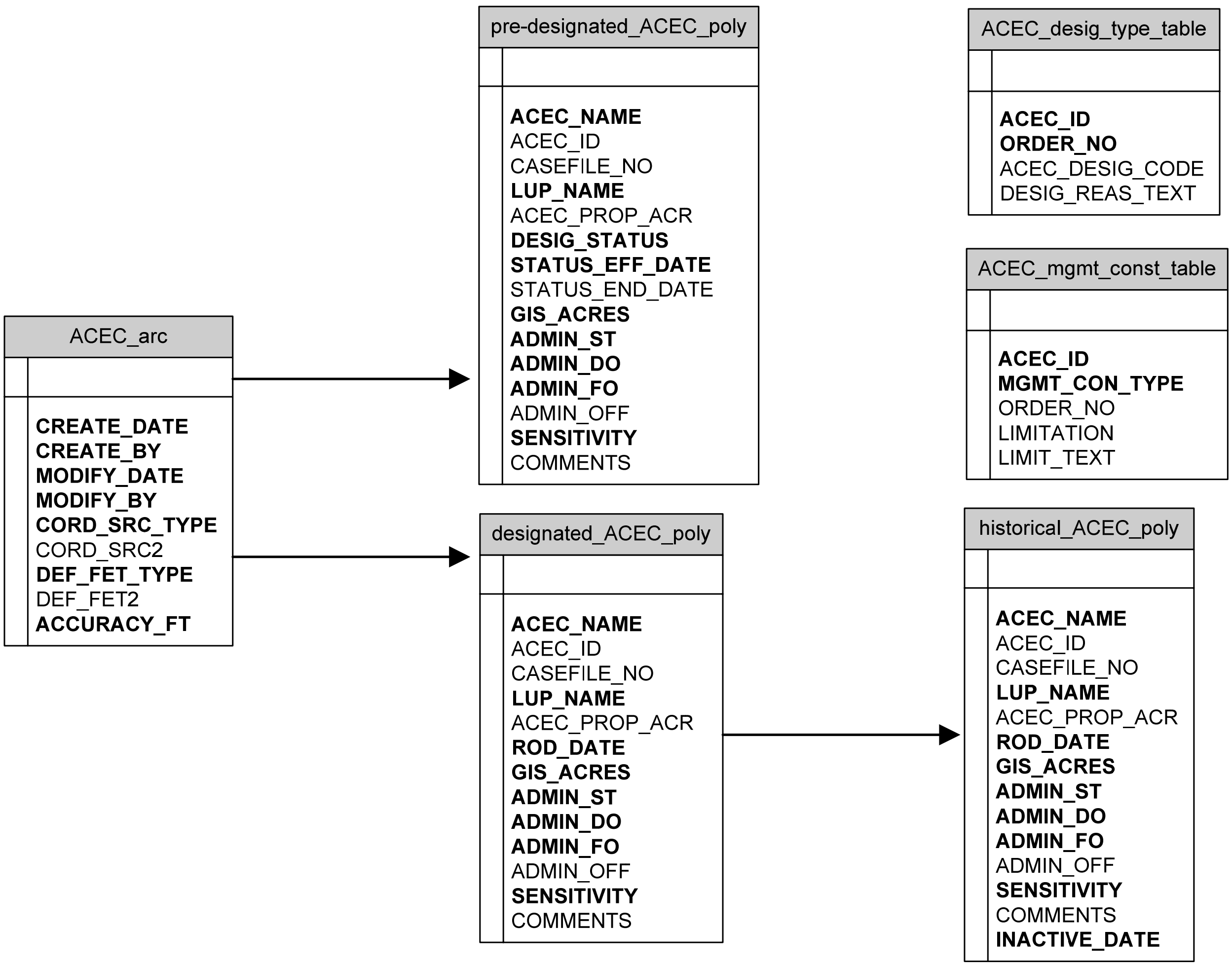


Figure 11: Physical Diagram for Areas of Critical Environmental Concern

Summary

The development and application of documented Implementation Guidelines to extend data standards beyond the logical model has allowed BLM staff to bridge knowledge gaps between different groups involved in the standards development process. It has also increased the number of feedback comments that deal with the actual content of the standard rather than mechanics of format.

The largest improvement has been seen in the increased conformity of data structure across the organization. By homogenizing the ‘interpretation’ of the standards report document, there are fewer issues with the data sets that are developed throughout the various offices of the Bureau.

References

Development of Data Standards for a Large Government Bureau: Thomas A. Chatfield and Barb Kett, U.S. Bureau of Land Management. Paper presented at the 2007 ESRI International User Conference

Guidance for Managing BLM Data Standards: U.S. Bureau of Land Management, March 2003

Data Standardization Procedures, United States Department of the Interior, Office of the Chief Information Officer, April 2006

Manual 1613 - Areas of Critical Environmental Concern: U.S. Bureau of Land Management, September 1988

Code of Federal Regulations (CFR) 1610.7-2

Acknowledgements

The authors would like to thank Barb Kett, Christine Hawkinson, and Heather Karlson for their editing assistance.

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

Lorena Creely

Spatial Data Analyst, Northrop-Grumman

US Bureau of Land Management

Denver Federal Center

Bldg 40, PO Box 25047

Denver, CO 80225-0047

Ph: 303-236-2695

Fax: 303-236-1981

lorena_creely@blm.gov