Geospatial Data Standards Development
Using Data Model Patterns

Christine Hawkinson and Thomas A. Chatfield

US Bureau of Land Management

ABSTRACT

Recently, BLM data architects began to use patterns for developing data standards. A pattern is a business-view of the data which identifies the business rules and common data elements (logical data models for standard business situations). Patterns provide a more effective and consistent development process, and provide the basis for business rules.   Once the pattern is approved, it can be used by itself or in combination with other patterns in the design of the physical database and implementation of feature-level metadata. Patterns speed up the standardization process. And the data standards have enabled the new BLM National Geodatabase.

As an example, we will use one of the first patterns developed, Location. We will discuss how this pattern is being reused for other BLM Boundary layers and how it is advancing the development of geospatial data standards across our organization.

Introduction

The US Bureau of Land Management (BLM) is a federal agency that is responsible for the management of public lands concentrated in the western United States. It has 12 state offices and approximately 150 field offices. Much of the data in the BLM has a geospatial aspect to it.

The BLM is developing data standards for our national data sets using a data standardization process (see Appendix A) as discussed in a paper presented at the 2007 ESRI user conference by two members of the BLM Data Management group.

Background and Issues

At the BLM, the number of potential geospatial data sets is considerable. It often took up to18 months to develop each standard. Our GIS specialists collect and maintain data and cannot wait for the standards development process to be completed. Currently, there are different data implementations across the 12 BLM Administrative States and even within those states. Yet, it is important that we have consistent standards across the Bureau as we have so many users and types of users. It was imperative that the process is 'sped up,' so more data standards are completed and the necessary implementation information to GIS Specialists is provided.

As geospatial data standards were developed for different BLM administrative boundaries, there was a noticeable repetition of the same attributes across the standards. The BLM data architects began to discuss the concept of patterns. Patterns are a reusable business model of data which describes data common across the business, with a common vocabulary and taxonomy. A pattern is a logical data model consisting of 4-12 entities. For example, a data model for organization or person/address could be a pattern since many areas within a company have this information in their systems.
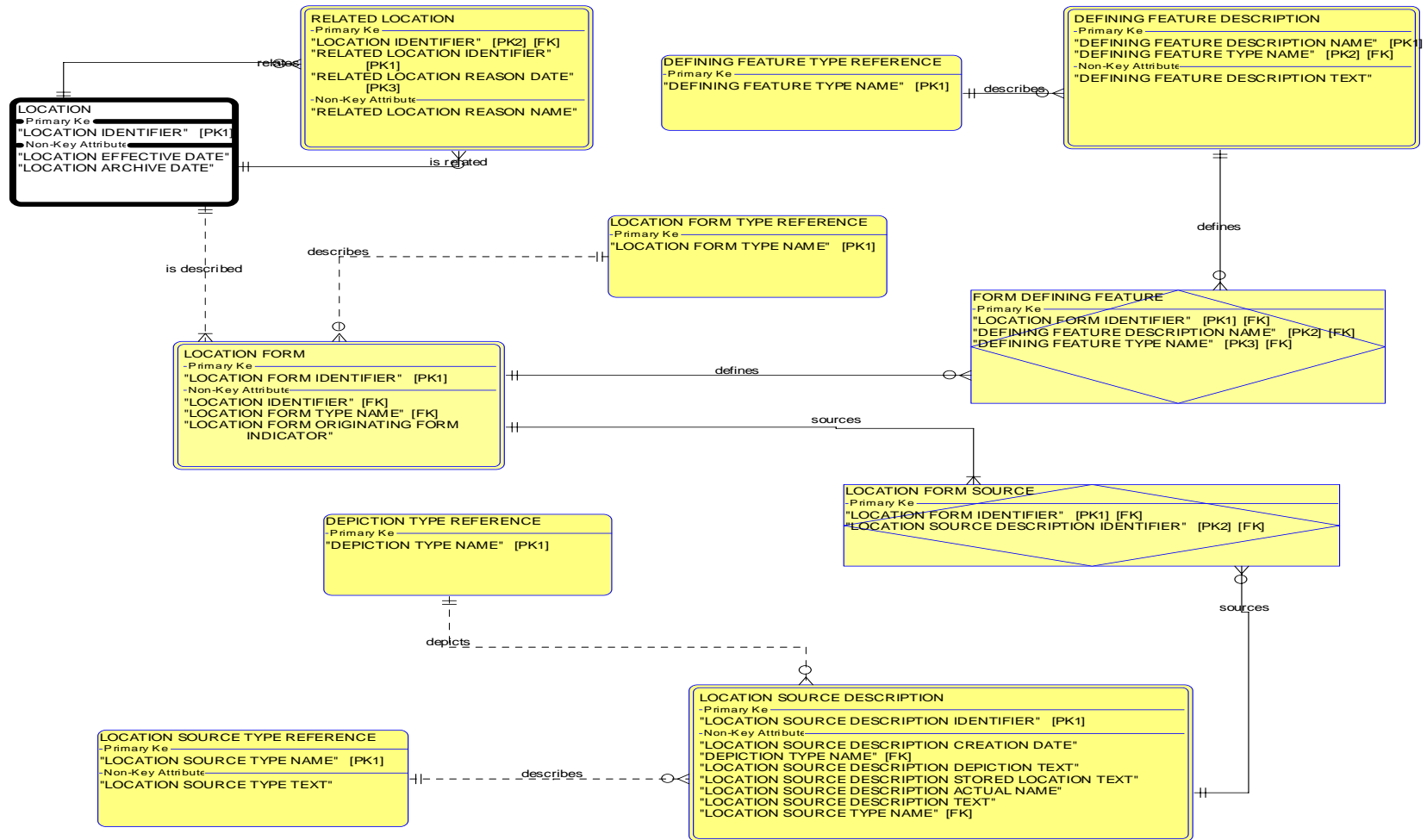
It was discovered that the shared feature level metadata attributes across our administrative boundaries were a pattern. This means that for each new geospatial data standard, the feature level metadata does not need to be remodeled.

BLM Location Pattern

As part of the data standards process, each data standard requires a logical data model. A logical model which includes feature level metadata (calling it the LOCATION Pattern) based on the review of data standards that we have been working on at the BLM was developed. The following diagram is a portion of this Pattern. When a data standard includes the entity "LOCATION," it indicates that the Location pattern entities and attributes may be required to implement the data standard.
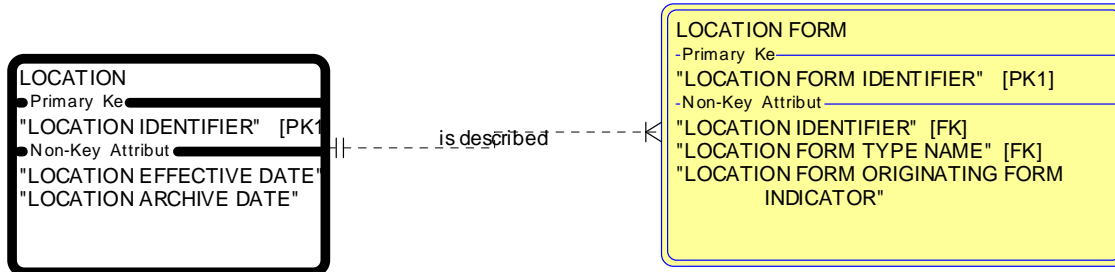
BLM LOCATION PATTERN

The following is a portion of the Location Pattern which includes the Defining Feature and the Coordinate Source Location metadata.
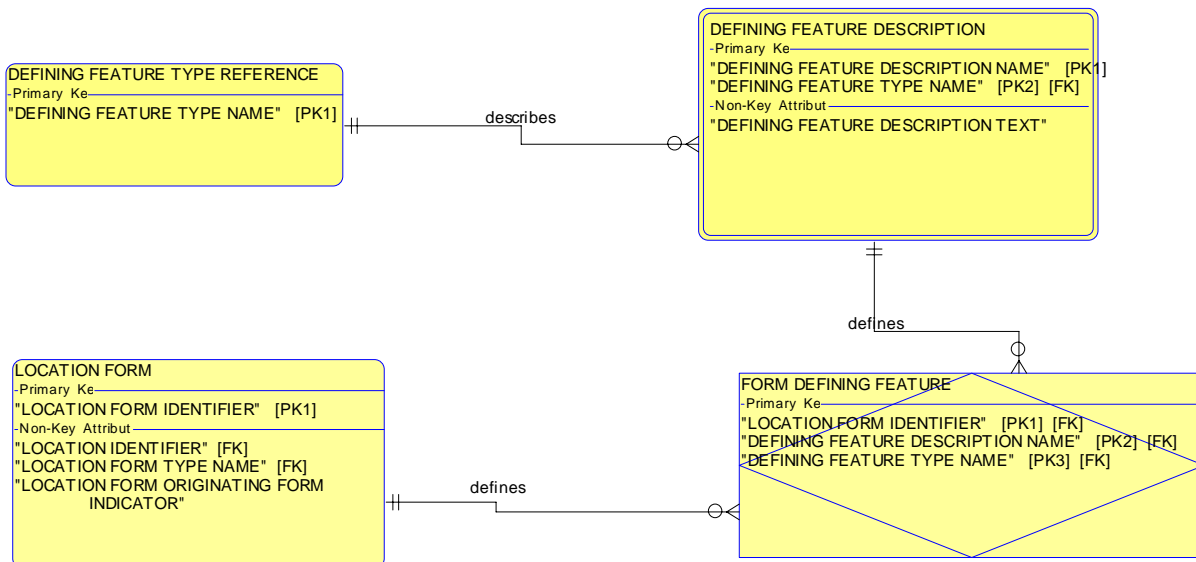
RELATED LOCATION
-Primary Ke-
"LOCATION IDENTIFIER"   [PK2] [FK]
"RELATED LOCATION IDENTIFIER"
          [PK1]
"RELATED LOCATION REASON DATE"
          [PK3]
-Non-Key Attribute-
"RELATED LOCATION REASON NAME"

LOCATION
-Primary Ke-
"LOCATION IDENTIFIER"   [PK1]
-Non-Key Attribute-
"LOCATION EFFECTIVE DATE"
"LOCATION ARCHIVE DATE"

relates

is related

DEFINING FEATURE TYPE REFERENCE
-Primary Ke-
"DEFINING FEATURE TYPE NAME"   [PK1]

describes

DEFINING FEATURE DESCRIPTION
-Primary Ke-
"DEFINING FEATURE DESCRIPTION NAME"   [PK1]
"DEFINING FEATURE TYPE NAME"   [PK2] [FK]
-Non-Key Attribute-
"DEFINING FEATURE DESCRIPTION TEXT"

defines

LOCATION FORM TYPE REFERENCE
-Primary Ke-
"LOCATION FORM TYPE NAME"   [PK1]

describes

is described

FORM DEFINING FEATURE
-Primary Ke-
"LOCATION FORM IDENTIFIER"   [PK1] [FK]
"DEFINING FEATURE DESCRIPTION NAME"   [PK2] [FK]
"DEFINING FEATURE TYPE NAME"   [PK3] [FK]

LOCATION FORM
-Primary Ke-
"LOCATION FORM IDENTIFIER"   [PK1]
-Non-Key Attribute-
"LOCATION IDENTIFIER"   [FK]
"LOCATION FORM TYPE NAME"   [FK]
"LOCATION FORM ORIGINATING FORM
          INDICATOR"

defines

sources

LOCATION FORM SOURCE
-Primary Ke-
"LOCATION FORM IDENTIFIER"   [PK1] [FK]
"LOCATION SOURCE DESCRIPTION IDENTIFIER"   [PK2] [FK]

DEPICTION TYPE REFERENCE
-Primary Ke-
"DEPICTION TYPE NAME"   [PK1]

depicts

sources

LOCATION SOURCE DESCRIPTION
-Primary Ke-
"LOCATION SOURCE DESCRIPTION IDENTIFIER"   [PK1]
-Non-Key Attribute-
"LOCATION SOURCE DESCRIPTION CREATION DATE"
"DEPICTION TYPE NAME"   [FK]
"LOCATION SOURCE DESCRIPTION DEPICTION TEXT"
"LOCATION SOURCE DESCRIPTION STORED LOCATION TEXT"
"LOCATION SOURCE DESCRIPTION ACTUAL NAME"
"LOCATION SOURCE DESCRIPTION TEXT"
"LOCATION SOURCE TYPE NAME"   [FK]

LOCATION SOURCE TYPE REFERENCE
-Primary Ke-
"LOCATION SOURCE TYPE NAME"   [PK1]
-Non-Key Attribute-
"LOCATION SOURCE TYPE TEXT"

describes

4

Location Pattern Diagram Explanation

The next section discusses the various parts of the Location Pattern.

```
LOCATION                                          LOCATION FORM
-Primary Ke-                                      -Primary Ke-
"LOCATION IDENTIFIER"   [PK1]                      "LOCATION FORM IDENTIFIER"   [PK1]
-Non-Key Attribut-                                -Non-Key Attribut-
"LOCATION EFFECTIVE DATE"      is described       "LOCATION IDENTIFIER"   [FK]
"LOCATION ARCHIVE DATE"                           "LOCATION FORM TYPE NAME"   [FK]
                                                  "LOCATION FORM ORIGINATING FORM
                                                          INDICATOR"
```

- A LOCATION is "A defined place that requires a way to locate it by some means. Note: Entities linked to Location have the potential for a geospatial aspect."

- A LOCATION is described by One To Many (1:M) LOCATION FORMS. We used the word FORM rather than Feature since we included Tabular as one of the forms.   For example, a LOCATION such as a river can be described as a line, polygon and/or tabular.

- A LOCATION FORM is "The form in which the location is described such as the description, shape, or appearance of the location." A Location Form Type Name is a point, line, polygon or tabular.

```
DEFINING FEATURE TYPE REFERENCE          DEFINING FEATURE DESCRIPTION
-Primary Ke-                             -Primary Ke-
"DEFINING FEATURE TYPE NAME"   [PK1]     "DEFINING FEATURE DESCRIPTION NAME"   [PK1]
                   describes             "DEFINING FEATURE TYPE NAME"   [PK2] [FK]
                                         -Non-Key Attribut-
                                         "DEFINING FEATURE DESCRIPTION TEXT"

                                                        defines

LOCATION FORM                            FORM DEFINING FEATURE
-Primary Ke-                             -Primary Ke-
"LOCATION FORM IDENTIFIER"   [PK1]       "LOCATION FORM IDENTIFIER"   [PK1] [FK]
-Non-Key Attribut-                       "DEFINING FEATURE DESCRIPTION NAME"   [PK2] [FK]
"LOCATION IDENTIFIER"   [FK]             "DEFINING FEATURE TYPE NAME"   [PK3] [FK]
"LOCATION FORM TYPE NAME"   [FK]   defines
"LOCATION FORM ORIGINATING FORM
        INDICATOR"
```

- The FORM DEFINING FEATURE is the characteristic (feature) constructed from a geographic feature that was used to create the location. Since a LOCATION FORM can be tabular, it has a 0:M relationship to FORM DEFINING FEATURE.

The other two related entities contain the values that contain the characteristic information for the feature is:

- DEFINING FEATURE TYPE REFERENCE contains the valid domain values for the Defining Feature Type. Defining Feature Type Name defines the specific features from which the arcs are derived to create the polygon boundaries; the information that describes the physical or mapping feature type that makes up the boundary. For example: Vegetation (change); Construction (pipe, fence); Coastal/Fluvial (stream, river, shorelines).

- DEFINING FEATURE DESCRIPTION is not a reference entity as there were too many values and the number of values change over time. The person entering the data can select a value already used or create a new one.
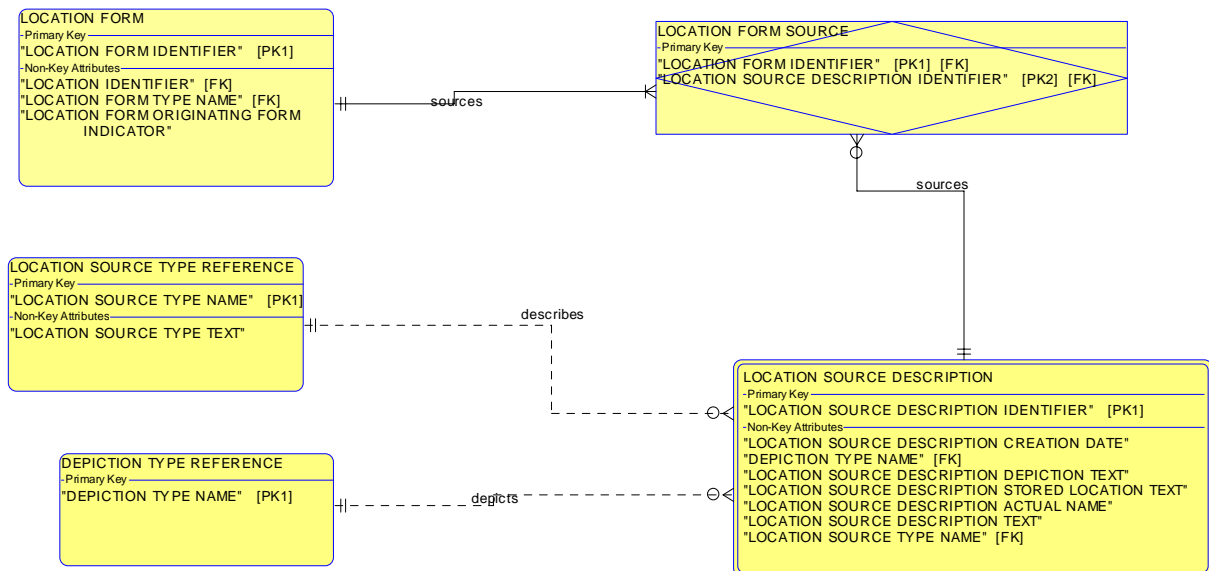
The LOCATION FORM SOURCE is the origin (actual source) for the location coordinate such as Map, Image or global positioning system. The other three related entities contain the values that can be selected for the coordinate source:

- DEPICTION TYPE REFERENCE contains the domain values of "scale" and "resolution."

- LOCATION SOURCE TYPE REFERENCE contains the domain values that can be selected for the LOCATION FORM SOURCE.

- LOCATION SOURCE DESCRIPTION is similar to DEFINING FEATURE DESCRIPTION in that there were too many values for the attributes in, so this entity could not support a discreet list of domain values.

This LOCATION / FEATURE LEVEL METADATA pattern was used to create table structures that will be standardized across our geospatial datasets. In a National Geodatabase, only one set of these tables are required, even if there are multiple feature classes (data sets), in the geodatabase.

6

Creating a Data Standard

The first two boundary data standards developed with patterns were Allotments and Areas of Critical Environmental Concern (ACEC). Allotments were a much simpler set of data than ACECs.

Allotments

Allotments are the basic geographic area used in administering BLM range land.

An allotment contains one to many Pastures. Overtime, a PASTURE can move to another ALLOTMENT. In the model, it shows that only PASTURE is related to LOCATION. In the logical model, it would be redundant to show an ALLOTMENT with a relationship to LOCATION since all ALLOTMENTS are made up PASTURES. In the Implementation of this standard, the PASTURES that make up an ALLOTMENT can be dissolved to determine the boundaries of the ALLOTMENT.
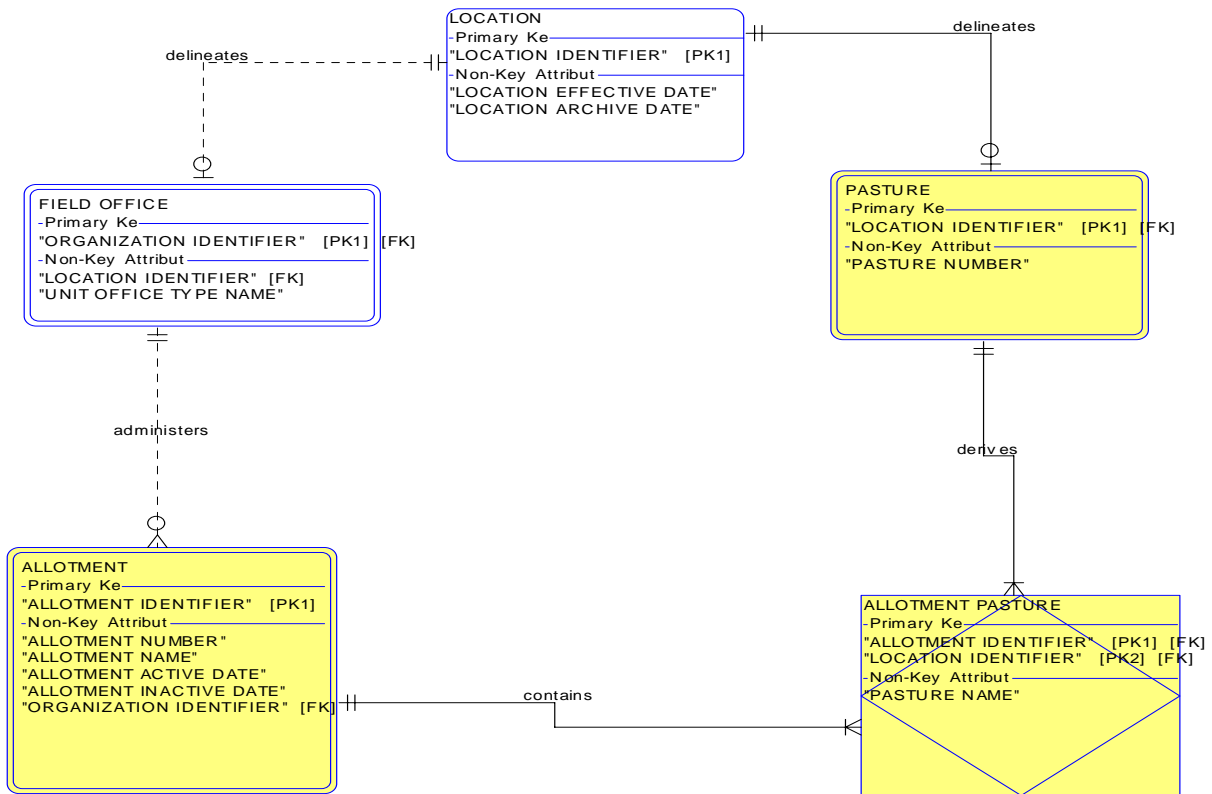
Allotment / Pasture Entities

The LOCATION Entity shows the relationship to the PASTURE and since LOCATION is also on the Location Pattern, there is a link back to the feature level metadata for this pattern.

Allotments are managed by a BLM field or district office, so the link to a BLM Administrative Office pattern needed to be incorporate as well. Most of the administrative boundaries are administered or managed by a BLM field or district office. With the BLM Administrative Office

pattern, standard tables that are reusable across all administrative boundary data sets can be developed, as we are doing with feature level metadata. The complete logical model for the ALLOTMENT data standard is in Appendix B. In that diagram one can see that the model also is linked to AUTHORIZATIONs.

When the logical model is fully defined, it is included in the Data Standard Report which is then sent out for comments to BLM state data administrators, data stewards and GIS specialists. The Data Standard Report will also include the Location pattern, if it is geospatial.

Implementation Guidelines Document

When consensus on the Data Standard Report is reached, the Implementation Guideline is written. The table structures, feature classes (where necessary) and domain values are included in an Implementation Guideline. This is sent out for review and comments to gain consensus on the design and implementation of the data standard across the BLM.

The tables and feature classes are designed based on the logical data model. Data Elements that are required for the data set are identified. Entities may be collapsed for performance or reporting requirements. Some of the domain tables are reused from other data standards (e.g., Defining Features and Coordinate Sources).

For the Allotment data standard, there are 4 tables, each representing a feature class: Pasture Arcs and Polygons, Allotment Polygons and Historical Allotment Polygons.
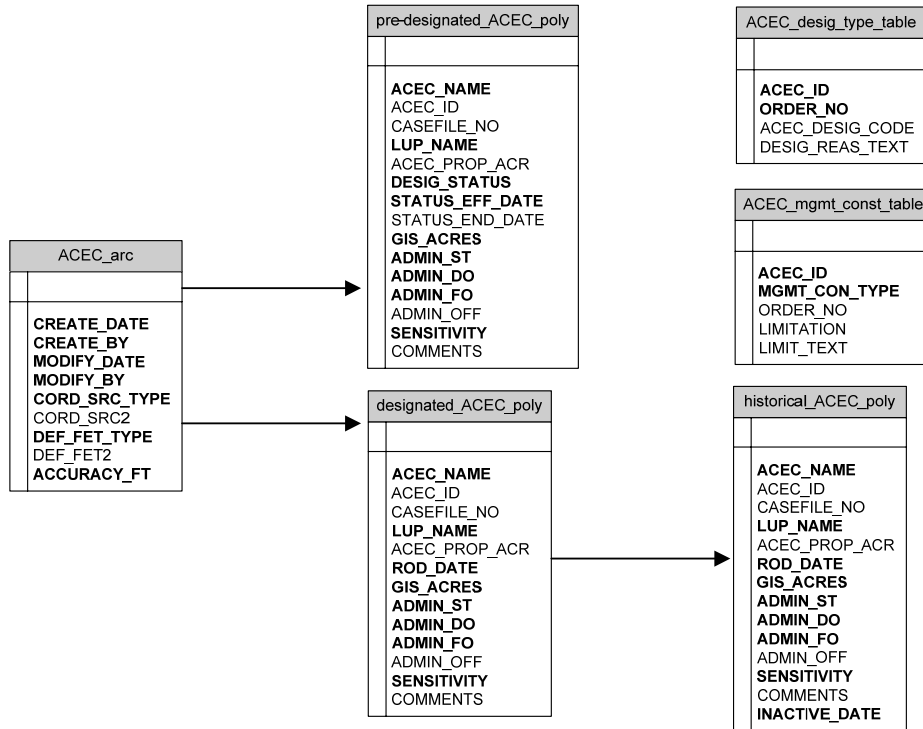
Allotment Tables:

past_arc:
CREATE_DATE
CREATE_BY
MODIFY_DATE
MODIFY_BY
CORD_SRC_TYPE
CORD_SRC2
DEF_FET_TYPE
DEF_FET2
ACCURACY_FT

past_poly:
ALLOT_NO
ALLOT_NAME
PAST_NO
PAST_NAME
GIS_ACRES
ADMIN_ST
ADMIN_DO
ADMIN_FO
ADMIN_OFF
ST_ALLOT_PAST

allot_poly:
ALLOT_NO
ALLOT_NAME
GIS_ACRES
ADMIN_ST
ADMIN_DO
ADMIN_FO
ADMIN_OFF
ST_ALLOT
ACTIVE_DT

hist_allot_poly:
ALLOT_NO
ALLOT_NAME
GIS_ACRES
ADMIN_ST
ADMIN_DO
ADMIN_FO
ADMIN_OFF
ST_ALLOT
ACTIVE_DT
INACTIVE_DT

In the above table structures COORD_SRC_ TYPE, DEF_FET_TYPE (Location Pattern), ADMIN_ST (state administrative office), ADMIN_DO (district office) and ADMIN_FO (field office) are data elements from other patterns, each with a standard set of domain values and format.

The Area of Critical Environmental Concern (ACEC) data standard was developed in much the same way as Allotments. The diagram of the full data standard is in Appendix B. By looking at the tables for ACEC below, several of the same data elements are included which appear in the Allotment tables. Other data elements that are in other patterns, such as Record of Decision (ROD) Date and Casefile_No, are included. Since they are from other patterns, it means they will have the same format and definition for all standards that use the same element.

ACEC tables:

pre-designated_ACEC_poly
ACEC_NAME ACEC_ID CASEFILE_NO **LUP_NAME** ACEC_PROP_ACR **DESIG_STATUS** **STATUS_EFF_DATE** STATUS_END_DATE **GIS_ACRES** **ADMIN_ST** **ADMIN_DO** **ADMIN_FO** ADMIN_OFF **SENSITIVITY** COMMENTS

ACEC_desig_type_table
ACEC_ID **ORDER_NO** ACEC_DESIG_CODE DESIG_REAS_TEXT

ACEC_mgmt_const_table
ACEC_ID **MGMT_CON_TYPE** ORDER_NO LIMITATION LIMIT_TEXT

ACEC_arc
CREATE_DATE **CREATE_BY** **MODIFY_DATE** **MODIFY_BY** **CORD_SRC_TYPE** CORD_SRC2 **DEF_FET_TYPE** DEF_FET2 **ACCURACY_FT**

designated_ACEC_poly
ACEC_NAME ACEC_ID CASEFILE_NO **LUP_NAME** ACEC_PROP_ACR **ROD_DATE** **GIS_ACRES** **ADMIN_ST** **ADMIN_DO** **ADMIN_FO** ADMIN_OFF **SENSITIVITY** COMMENTS

historical_ACEC_poly
ACEC_NAME ACEC_ID CASEFILE_NO **LUP_NAME** ACEC_PROP_ACR **ROD_DATE** **GIS_ACRES** **ADMIN_ST** **ADMIN_DO** **ADMIN_FO** ADMIN_OFF **SENSITIVITY** COMMENTS **INACTIVE_DATE**

National Geodatabases

The BLM is currently developing multiple National Geodatabases in a central location. Some of the data layers may be merged into the same geodatabase and some may remain as separate geodatabases, depending on the business requirements and feature class relationships. The National Geodatabases are being implemented to allow for access to the data by many different levels of the organization. This is occurring to help BLM reduce time and efforts put in to 'data calls' for information and the resulting time it takes to manipulate the data and merge it together.

Long term, the BLM hopes to have National Geodatabases with satellite geodatabases in State and Field Offices. Data replication will maintain the currency between the National Geodatabases and the satellite Geodatabases. Because many data elements across data sets are the same (feature level metadata, admin offices, designations, authorizations, land use plans), these domain tables only need to be stored once and maintained in the central National Geodatabase server, using data replication to maintain currency of their values across the BLM.

Summary

There were some issues during the review of the Data Standards Report and the Implementation Guide. The Data Administrators tend to understand logical data models and our GIS specialists understand the GIS Implementation Guideline, the physical data requirements. There were a lot of questions on the logical data model:

- "Why were there so many entities? We only need a couple of tables."

- "What do the relationship lines mean?"

It was determined that as long as the data administrators understood the data model and the business rules were modeled correctly, the GIS specialists did not need to understand the logical model as much as they did the Implementation Guidelines and how the tables were designed.

In some cases, there was information needed that was part of patterns that had not developed yet were needed. For Allotments, the Authorization Pattern was necessary, but there was no time to develop it. A relationship to the Authorization Entity was used; the Authorization Pattern will be developed that at a later time. The entities for Authorization will be updated as the pattern is developed, but the current relationship will not affect the grazing data standard. Land Use Planning (LUP) has a similar situation. Boundaries for LUP need to be developed, but at the same time, an ePlanning project is in progress that is developing data models for Planning.

Another issue was when patterns were first being created; it was discovered that additional time was spent on developing the initial patterns and reviewing them with the BLM communities.

Over the long-term, findings show that using approved patterns to develop new standards help reduce time spent in:

- Developing and gaining consensus on the Logical Data Model, since only the new entities in the data model need to be reviewed;

- Developing the Physical Implementation Guidelines, since the feature level metadata is already defined and can be dropped into the guidelines easily.

In the future, there will hopefully be several patterns that can be reused across the BLM. These include: Authorizations/Leases, Land Use Plan Boundaries, Treatments and Name/Address. These patterns can help identify who requires what data to complete the work in their program.

Because the same set of domain tables across data standards is reused, there is an improved quality and consistency of data. The connection between the logical data model and the implementation guidelines are better understood. The relationship between tabular entities and geospatial entities are better defined, and therefore, it is easier to determine which attributes are needed for the data set and which attributes are better suited in a tabular database with links to the geospatial data set.

Modifications to an existing standard will be simpler to maintain. When a domain table requires a new value, the data sets it will affect can be easily seen. Another benefit will accrue when a data standard or pattern is modified. By understanding which entity is changing and the pattern it belongs to, where that pattern is being used to determine which standards may require modification can be identified.

Appendix A: BLM Data Standard Process

Propose Data Standard

Business community requests a new data standard of National Data Steward responsible for that data area. The National Data Steward appoints a data standard adoption team (Team) with a designated leader which then develops a proposal for the new data standard. The draft proposal is evaluated by the business community, State Data Administrators, and others, and is finalized by Team.

Adopt Standard

The Team researches, drafts, and completes a draft data standard which includes a logical data model. Reviewers, including the business community, evaluate the draft data standard which provides the basis for preparing a final standards report. The final data standards report is attached to a formal Instruction Memorandum and is then goes through the approval process. Once the standard is adopted, the Implementation Guidelines document is developed which includes a listing of domain values, creation of an empty geodatabase, and the application of geodatabase topology rules. The Implementation Guidelines document ensures that the standards are applied as desired.

Implement Data Standard

The Team devises a practical implementation approach and develops a formal implementation plan, in close cooperation with the National Data Steward and stakeholders. The implementation plan receives widespread review by the established list of stakeholders. It is then submitted for management approval and funding. The plan is implemented by an Instruction Memorandum and executed under leadership of the National Data Steward.

Maintain Data Standard

Review existing standards for currency and completeness to determine if modifications are needed. There may be requests to make minor modifications to the standard or a major revision, which requires the full data standardization process is completed.

Appendix B: Grazing Allotment Data Standard

The completed Grazing Allotment Data standard is shown below. The shaded box that encompasses some of the entities shows the relationship to patterns. Not all entities from the BLM Administrative Data Pattern are included; only those that showed the relationships back to the allotment standard are included. An Authorization pattern has not been developed yet, but there was a need to show a relationship to the ALLOTMENT AUTHORIZATION, since all Allotments are related to some type of authorization by the BLM.

Allotment Pasture Boundary 3/10/2008 version 7 DRAFT

12
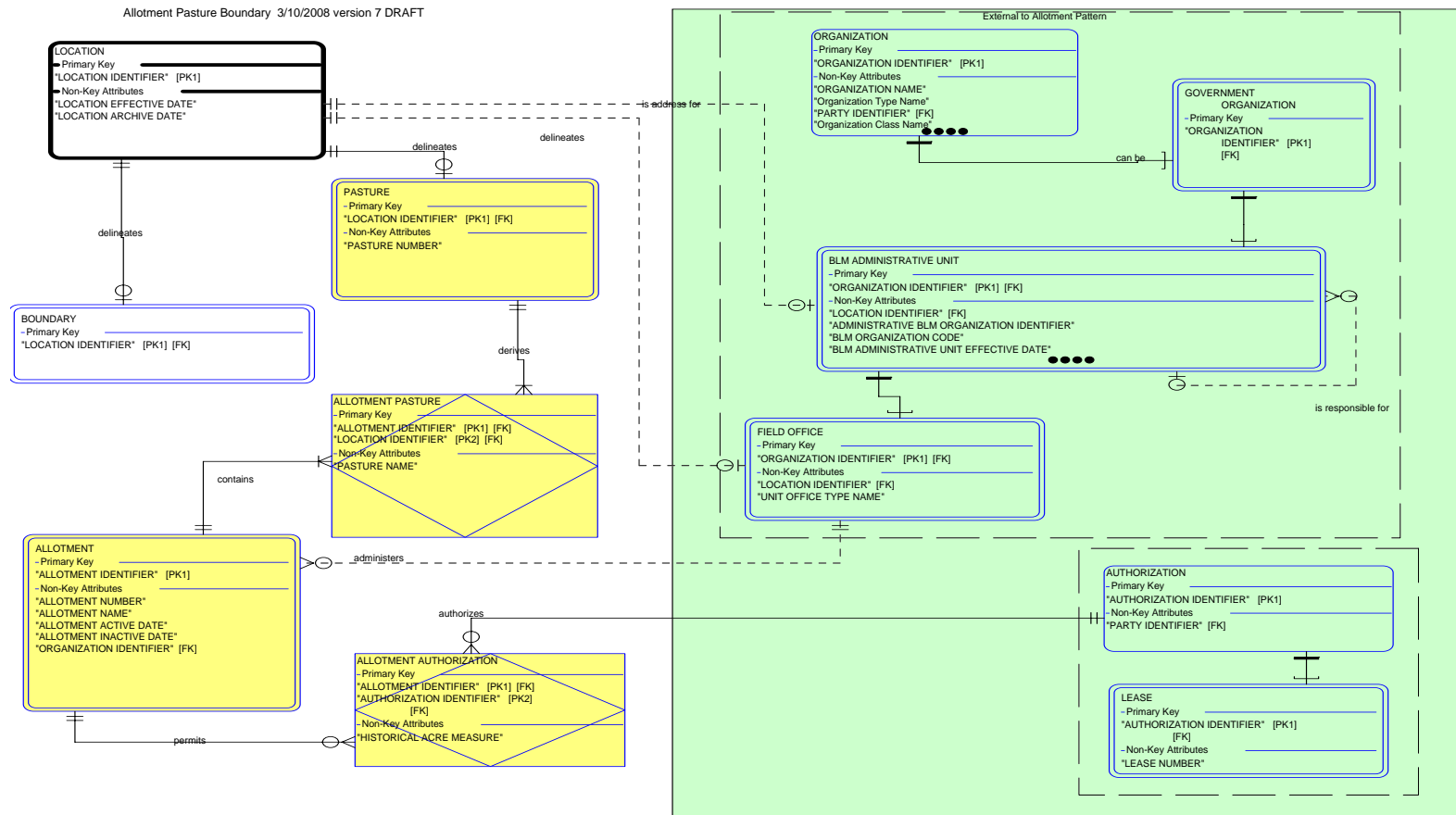
Appendix B: Areas of Critical Environmental Concern (ACEC) Data Standard

The completed ACEC Data Standard is shown below. The shaded box that encompasses some of the entities shows the relationship to patterns. ACECs require information from Land Use Plans and BLM Administrative Offices, therefore some of the entities from those patterns are included in the shaded boxes.

Areas of Critical Environmental Concern 3/5/08 version 6 DRAFT
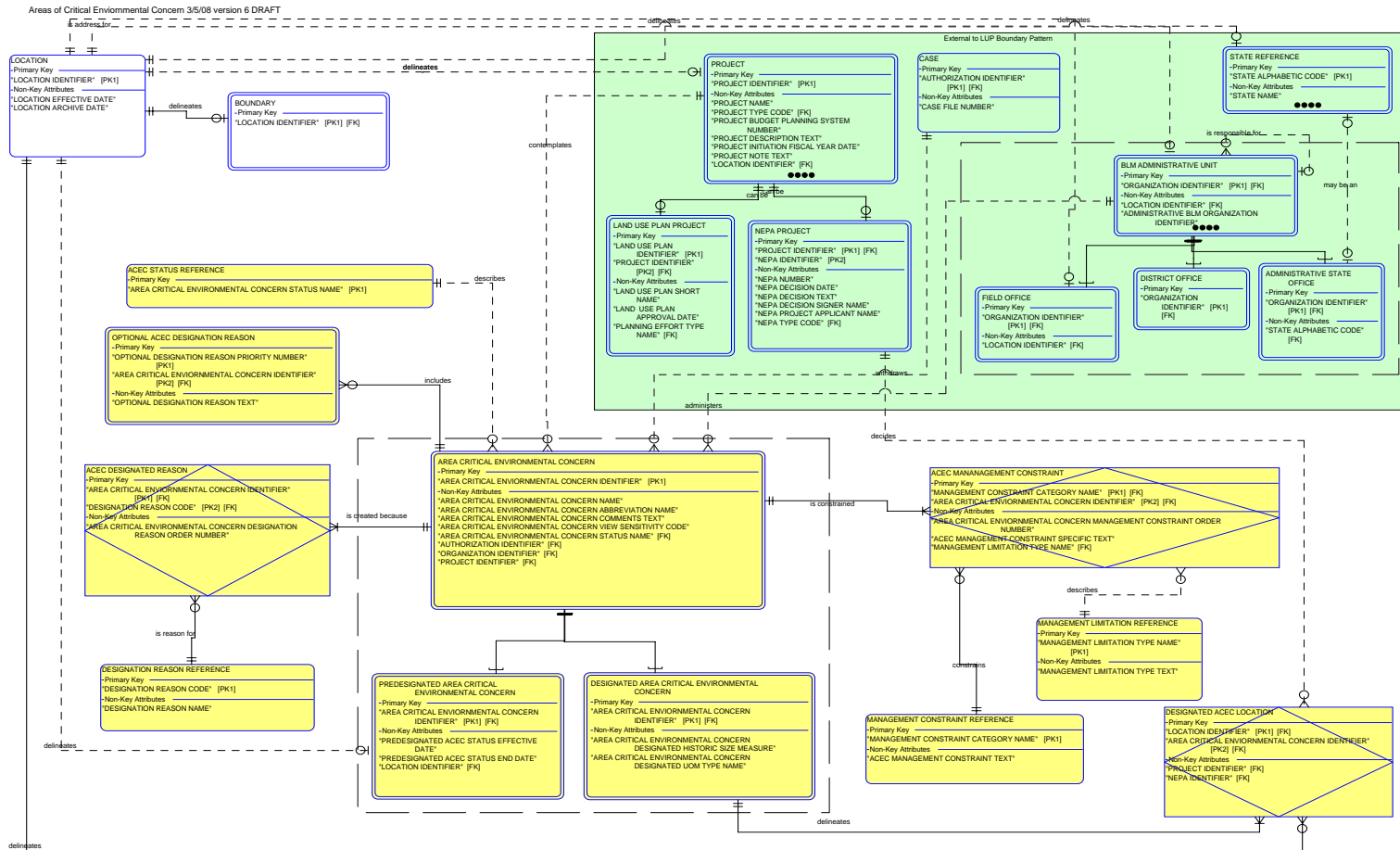
Appendix C: Glossary

Business Rules

Business rules describe what the business does and why, and provide a formal structure for understanding business operation. From an information system and data perspective, business rules dictate how data is managed and accessed. These types of rules specify constraints on the creation, update, use, archival, and disposal/deletion of data. Business rules on a logical data model include the relationship lines, definitions and domain lists.

Data Standard

A documented agreement on a set of rules or terms for a data entity (or set of entities) that applies to a collection and use of that data. The data standard applies to the full scope of the organization affected by the data and enables consistency and predictability.

Logical Data Model

A logical representation of the business data requirements of an organization independent of hardware or software constraints which provides a foundation for data control with specific and accurate data definition.

One reason LDMs are critical is that they make business rules explicit (rather than imbedded in code). Business Rules describe what the business does and why, and provide a formal structure for understanding business operation. From an information system and data perspective, business rules dictate how data is managed and accessed. Definitions, domain sets, relationships and optionality are considered business rules. Examples of business rules that can be shown in a data model:

- An ALLOTMENT is administered by 1 and only 1 BLM ADMINISTRATIVE OFFICE. A BLM ADMINISTRATIVE OFFICE administers zero to many ALLOTMENTS.

- An ALLOTMENT consists of 1 to many PASTURES. A PASTURE belongs to 1 to many ALLOTMENTS (over time).

Pattern

A reusable business model of data which describes data common across the business with a common vocabulary and taxonomy.

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

Acknowledgements

Author Contact Information

Christine Hawkinson

Data Architect , G&B Solutions

US Bureau of Land Management

Denver Federal Center

Bldg 40, PO Box 25047

Denver, CO 80225-0047

Ph: 303-236-4031

Fax: 303-236-1981

christine_hawkinson@blm.gov

Thomas A. Chatfield

Spatial Data Architect

US Bureau of Land Management

Denver Federal Center

Bldg 40, PO Box 25047

Denver, CO 80225-0047

Ph: 303-236-41936

Fax: 303-236-1981

tom_chatfield@blm.gov