

GIS Core Database Revision:  
Minnesota DNR  
Extended Data Dictionary  
Design and Deployment Overview  
(DRAFT)

April 17, 2002

Robert Maki  
Hal Watson  
Sandra Paddock  
Minnesota DNR  
Management Information  
Services Bureau

# Minnesota DNR Extended Data Dictionary Design and Deployment Overview

## 1.0 Overview

The Minnesota Department of Natural Resources (DNR) Extended Data Dictionary (EDD) is an Oracle database that serves as metadata registry describing a wide variety of spatial and non-spatial data resources. The EDD serves a variety of purposes within the DNR GIS enterprise environment, including: 1) a repository of corporate knowledge of specific data resources, promoting the long-term value of department data investments, 2) an explicit description of information resource content, to facilitate informed data application, 3) centralized coordination of physical data format and access method information to instruct corporate business applications, and 4) distribution and access control of data resources to data centers throughout the DNR wide area network (WAN).

This document consists of the following sections:

<i>Section Number</i>	<i>Section Name</i>	<i>Section Description</i>
1	Overview	EDD Purpose
2	Background	Circumstances leading up to EDD development and deployment
3	Design Overview	Overview of general design considerations, strategies, and rationale
4	Logical Model Sections	Detailed discussion of various portions of the EDD data model. Collections of entities that function together to meet a business need
5	Future directions	Anticipated changes and enhancements to the database

## 2.0 Background

The Minnesota DNR has maintained descriptive narratives and code definitions for their GIS data resources since the mid-1990s. These data were managed in desktop database form with customized programs written to export the data into a variety of documentation types to support on-line metadata publication and indexing for participation in national and state spatial clearinghouse search engines (FGDC-SDI, Minnesota Geogateway). Through time, this documentation environment became inefficient. The existing database design was ill-suited to handle descriptions of data products that were closely related to other data products (resulting from generalization, aggregation, thematic subset, or format conversion). These so-called “derived products” tended to be incompletely described and difficult to maintain within the existing database schema. Maintenance also posed challenges since the narrative documentation had multiple contributors, some of whom existed in regional offices with moderate to poor network connectivity (in terms of bandwidth). Other challenges were presenting themselves at the same time. Base level GIS software was changing, with new data types becoming available. Internet-based GIS applications were calling for new types of data sources. Distribution of data resources to both networked and non-networked sites was cumbersome and inefficient. In general, the existing GIS infrastructure was due for an overhaul.<sup>1</sup>

During the EDD data modeling effort a variety of existing concepts were considered (e.g. Marble, 1991) as were other metadata design registry efforts. A fragment of the ISO11179 (Metadata Registry) data model was adopted to handle data descriptions below the table level as well as provide a framework for integration with a conceptual data element registry component. Descriptive narrative elements were reviewed to ensure compliance with FGDC Content standards, Minnesota Geographic Metadata Guidelines, and Dublin Core standards as exported document types.

Originally conceived as a repository for storing metadata narratives, the development team soon realized the database’s potential for serving in a more active role. Provisions were made for the database to inform applications that access, manipulate, and transfer data. These included elements for data location, access/authentication, format, administrative affiliation, feature class content, and time parameters for scheduled maintenance events.

EDD deployment progressed amidst a variety of related system design and development activities, including the development of a system-level application that conducts intelligent replication-update of fileserver-based data through the wide area network that relies on the EDD.<sup>2</sup>

The EDD became operational in September of 2001, being deployed on a Sun Microsystems 250 server, running Solaris 7.X, in Oracle 8.1.X.

---

<sup>1</sup> Minnesota DNR, Management Information Services Bureau, “Revised Core Database General Architecture Design”, March 2000; Minnesota DNR, Management Information Services Bureau, “GIS Core Database Architecture Revision”, 1999.

<sup>2</sup> The data distribution-replication system is described in: Minnesota DNR, Management Information Services Bureau, “Update Data Resource Site Application Administrator’s Guide (Version 2.0),” 2002. EDD-Administrator interactions related to data distribution are handled through a Visual Basic application described in Minnesota DNR, Management Information Services Bureau, “Data Resource Site Administration Tool User’s Guide,” 2001.

### 3.0 Design Overview

The Extended Data Dictionary was guided by several design principles, including:

1. Should follow a normalized approach, minimizing information redundancy.
2. Should hold a complete description of data at multiple levels of abstraction from a broad “dataset” perspective (translated as “layer” in GIS lingo) to specific code value meanings.
3. Must include provisions for storing non-spatial data.
4. Appropriate and applicable metadata documents (Dublin Core, FGDC CSDGM, MIMG) need to be generated from the database.
5. Must be accessible from a variety of client operating system and development environments.
6. Must allow for selective access through DNR administrative group authentication.
7. Must support active systems for spatial product creation and data distribution.
8. Must minimize dependencies on vendor specific technologies (e.g. base level GIS management or application software).

### 4.0 Logical Model Sections

The Extended Data Dictionary includes logical entities that can be roughly viewed as having membership in four overlapping groups: 1) Descriptive narrative support, 2) Element content definition, 3) Data distribution support, and 4) Product generation. These are described in separate subsections below. The complete EDD data model with column listings is provided in Appendix A and would likely be a useful reference when reviewing the contents of this section.

#### 4.1 Descriptive Narrative Support

The Descriptive Narrative Support section is the most complex section of the database. A simplified data model of entities involved in the Descriptive Narrative Support group function is provided in Figure 1. This is the portion that stores standard metadata element information. The complexity of the section follows the definition of two separate types of GIS data layer types: 1) Core, and 2) Derived Product (DP). Core data are optimized for maintenance while DP’s are optimized for user application. The department strategy is to collapse the “Core” business data of the department to a succinct set of geography and associated attributes, minimizing redundant cartographic features, and normalizing attribute tables. DP’s are generated from Core data and feature planned cartographic redundancy and preferred user data formats. They are also rich in long name descriptions that would normally be excluded from Core attribute tables. DP’s are created on an as-needed basis, and may exhibit significantly overlapping content. In concept, DP’s are stored and accessed within a “data mart” paradigm.

From a modeling perspective, some DP properties are inherited directly from Core sources (e.g. positional accuracy), while others are unique to the derived product (e.g. Long Name, Entity Overview), while still others exhibit properties of both (e.g. Lineage). One of the principle

challenges in designing the data model was to allocate data properties to their proper home while minimizing redundancy in content. Descriptive DP metadata reports could then be constructed that draw upon elements from a combination of sources. The resulting data structures were largely effective in meeting these objectives, although some deficiencies remain.

The primary entities in this model fragment are: CORE\_LAYER, CORE\_PARENT\_CART, METAPRODUCT, DERIVED\_PRODUCT, and TABLE\_REG. These hold the principle identifying characterizations; essentially, Sections One and Two of the FGDC content standard. CORE\_LAYER and DERIVED\_PRODUCT have some common elements. Each delegates many narrative elements to the METAPRODUCT table, the records of which could be shared between CORE\_LAYER and DERIVED\_PRODUCT if a core product and a DP are essentially the same at a gross level of description. A clear distinction between CORE\_LAYER and DERIVED\_PRODUCT relates to feature class handling. DP's are by definition single feature class layers, while Core layers can support multiple feature classes. CORE\_PARENT\_CART is the intersection table between CORE\_LAYER and FEATURE\_CLASS, and contains one portion of the layer lineage information, on the premise that individual cartographies have separate processing histories. This "base-level cartography" information provides one part of the DP lineage report, where DP feature classes inherit properties from their Core sources. The TABLE\_REG entity is essentially a table registry that stores the identity of the tables that are explicitly associated with core layers and DP's. Attribute accuracy and lineages are stored here, in this case, contributing to both core layer and DP metadata reports. An entity associated primarily with the product generation subsystem, CREATION\_PROCESS, holds a description of the processes that were used to create a DP from the core sources and tables, which is the final component of DP lineage report narratives.

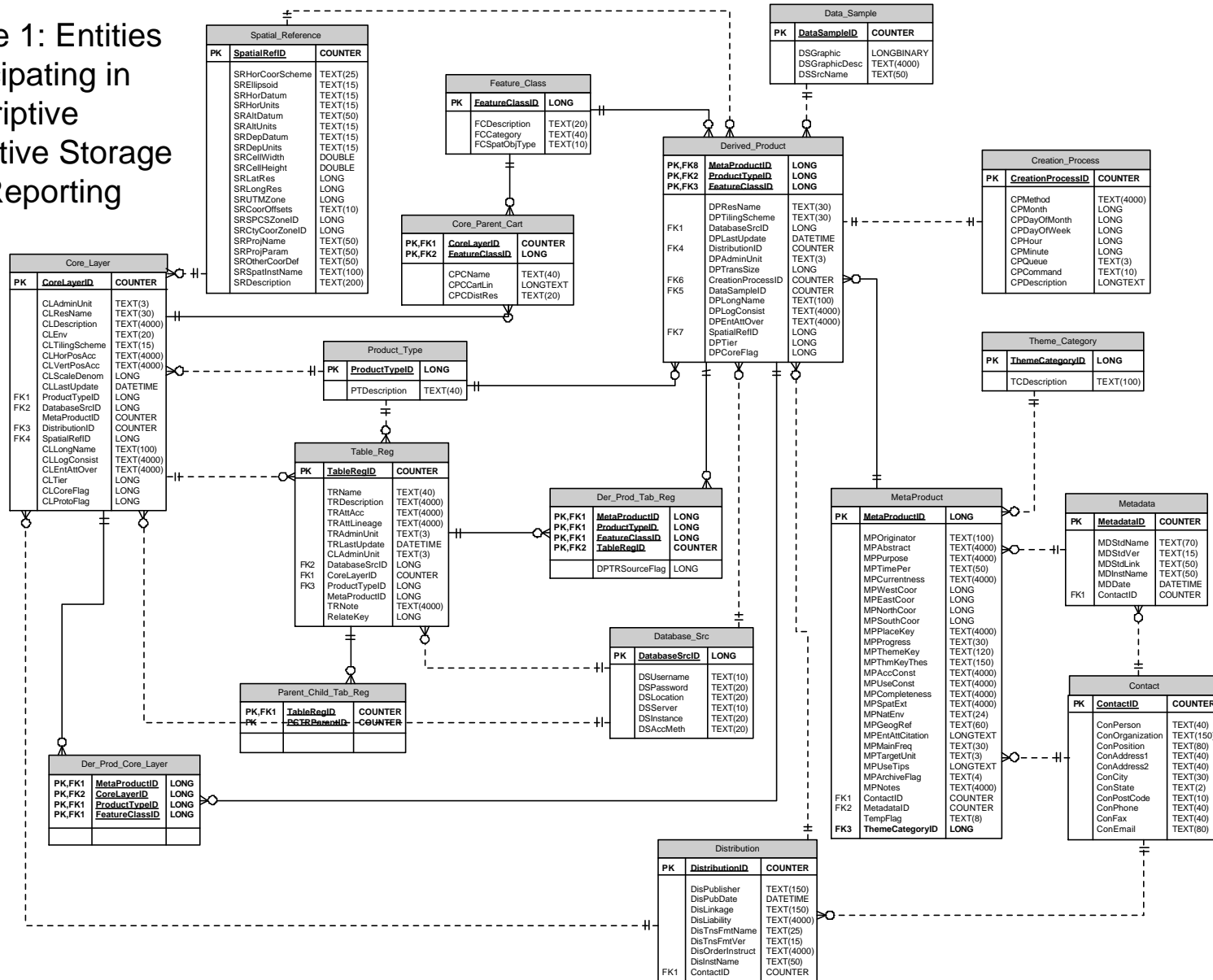
A constellation of tables exists around METAPRODUCT and each of the layer registry tables. These include spatial reference information, structures for handling contact information, product type (physical format), distribution notes, and metadata version references. An intersection table (DER\_PROD\_CORE\_LAYER) forms the relationship between CORE\_LAYER and DERIVED\_PRODUCT. Another intersection table (DER\_PROD\_TABLE\_REG) forms the relationship between DERIVED\_PRODUCT and TABLE\_REG. This entity includes a column that defines the relationship type: either a DP source, or an attribute table associated explicitly with a DP. A DP may be derived from multiple Core and Tabular sources. All data types (Core, DP, and Table) have relationships to DATABASE\_SRC, a table that holds the information necessary to access the data.

This portion of the database is expressed most commonly in Java Server Pages (JSP) reports to internal GIS users. It is maintained by a server-side Java application that manages the full range of content and relationships described here.<sup>3</sup>

---

<sup>3</sup> Minnesota DNR, Management Information Services Bureau, "Extended Data Dictionary Application Functional Requirements", May 2001.

Figure 1: Entities Participating in Descriptive Narrative Storage and Reporting



## 4.2 Element Content Definition

This section of the database is a natural extension of the Descriptive Narrative Support section. In essence, it describes the tabular attribute portion of DNR's spatial data resources. It is described separately because: 1) narrative metadata standards do not typically include provisions for description at this level, and 2) it is equally applicable to all data stored in tables regardless of whether the tables are associated with GIS or not. This is a schema based almost entirely on the ISO11179 data model, with some simplifications. A simplified schema for this section is presented in Figure 2.

The Element Content Definition section describes a hierarchical set of relationships beginning with the object class TABLE\_REG, and moving to DATA\_ELEMENT (column), VALUE\_DOMAIN (set of permissible values and associated value meanings that can be viewed collectively), and finally a PERMISSIBLE\_VALUE table and a VALUE\_MEANING table that intersect within an ENUMERATED\_DOMAIN. This highly normalized structure allows column definitions to be shared between tables. The PERMISSIBLE\_VALUE table holds a bare minimum number of values that can be combined with individual "meaning" records to create value domains (comprised of enumerated domain records). Similarly "meanings" can be reduced to a succinct set and projected onto different enumerated domains, allowing more than one value domain to share a single meaning. Value meanings are also a bridge to a conceptual domain portion of the model that has yet to be implemented. Structures to store non-enumerated domain information are also present in the model, although these had not been implemented at the time of this writing.

The element content definition section is currently used for two purposes: 1) documentation support to GIS users and developers in the form of JSP-generated reports, and 2) a source for translating succinct code descriptions into long name descriptions. In the future, this part of the database will support applications designed to assist users in identifying data resources that fit their business requirements. Creation of an enterprise-wide attribute validation service is also envisioned.

## 4.3 Data Distribution Support

For a variety of reasons, the DNR makes extensive use of fileserver-based spatial data sources. The process of replicating and distributing data resources to remote sites can be labor-intensive and presents significant opportunities to reduce costs associated with a key internal business function. Derived Products are the department user's primary spatial data resource. Data distribution activities are centered on a network of file servers that host an encapsulated set of DP data, site-describing metafiles, and access instructions that are collectively referred to as a "Data Resource Site" (DRS)<sup>4</sup>. Conceptually, DRS data content is an intersection of a DRS registry and a DP registry. Adding the option of associating "Areas of Interest" (AOI's) with each layer assigned to each site results in the EDD approach to tracking data distribution. The intersection of DRS, DP and AOI in the database reflects a state that is termed the "desired condition" of the

---

<sup>4</sup> Minnesota DNR, Management Information Services Bureau, "GIS Data Resource Site Specifications, version 2.0", August, 2001.

site (i.e., the data content that the administrator wants the site to have). The DRS Site Update application reads these desired conditions, analyzes the target site for the existing conditions, develops a workplan to update the site, and executes the workplan. It should be noted that centralized data server sources are supported in the DRS environment and that the EDD provides full support for documenting these sources and associated them with specific DRS's.

Correctly performing updates within the context of the target file system organization, source data type, and feature class content requires that the DRS Site Update Application draw upon a variety of EDD elements. A simplified model fragment of these is provided in Figure 3.



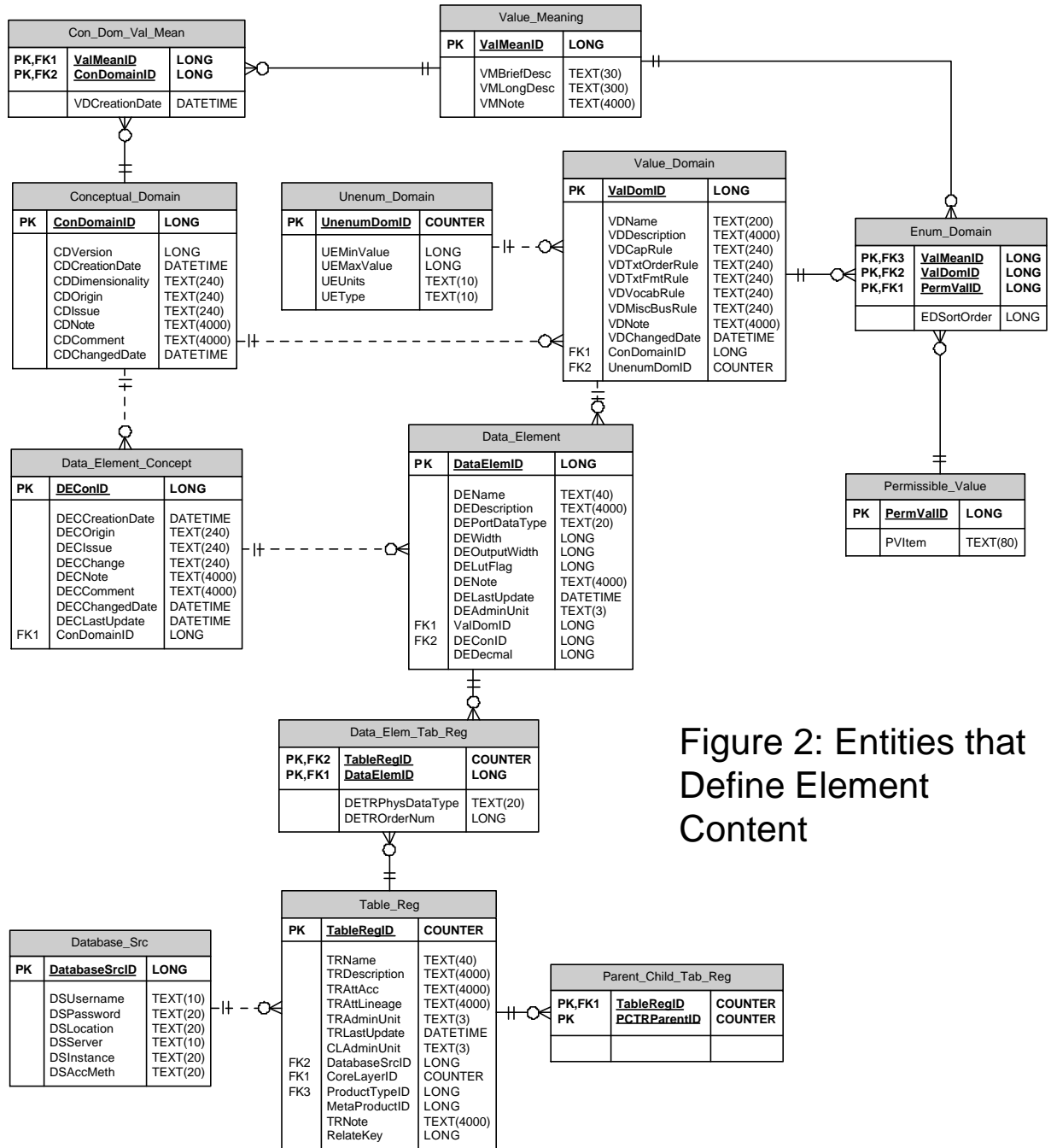
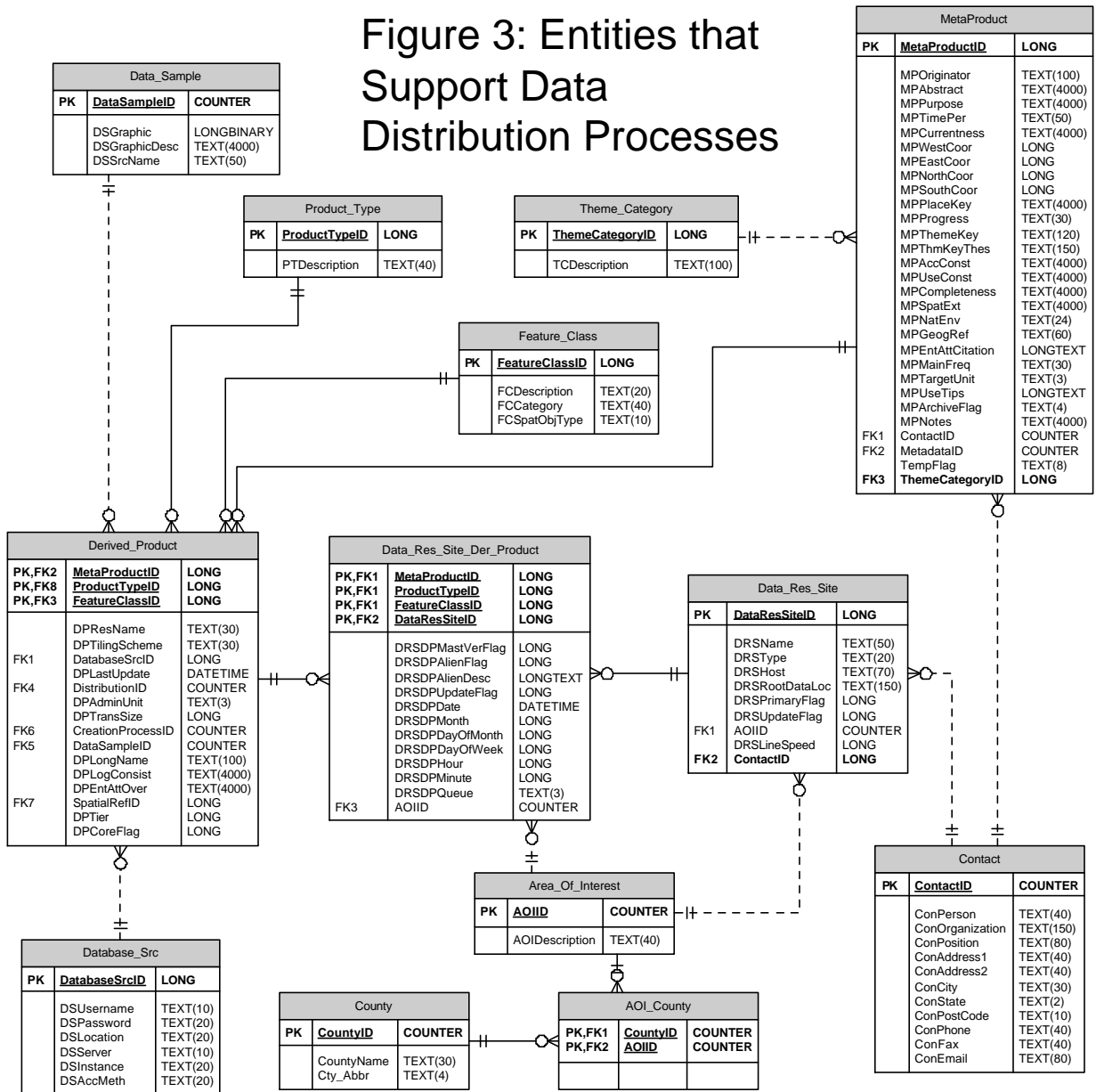


Figure 2: Entities that Define Element Content

# Figure 3: Entities that Support Data Distribution Processes



In addition to DERIVED\_PRODUCT, DATA\_RES\_SITE, and AREA\_OF\_INTEREST, the application draws on METAPRODUCT, THEMATIC\_CATEGORY, FEATURE\_CLASS, and DATABASE\_SRC.

Area of interest handling within the database requires some additional discussion. Early in the design process, technical staff decided to express AOI using “county” building blocks for the initial roll out of the environment (this is reflected in the relationship between AREA\_OF\_INTEREST and COUNTY). Avoiding ad hoc AOI definition simplified the database design and related application development tasks. Since most DNR staff operate roughly within a county context, this seemed adequate. The approach taken here also allowed for the establishment of stored, persistent AOI’s that could be used repeatedly, as needed.

#### 4.4 Product Generation

The EDD is designed specifically to guide the creation of Derived Products from Core Layers and tabular data.<sup>5</sup> In a general sense, this is an example of a business support application that makes active use of EDD elements. Product creation programs, which exist outside of the EDD are guided by EDD-generated “product parameters” (feature class, source data type, data location, data source instance, etc.) of both DP and Core sources to (in part) inform applications that manipulate data. One of the central concepts of DNR product generation is to create general programs that are reactive to product parameters, thereby reducing to the amount of custom code required to create the products. Product generation is supported by a CREATION\_PROCESS entity that stores time parameters for scheduling product creation events. Operating system level processes monitor the EDD and trigger creation processes based on the stored time parameters. A model fragment of the entities involved in product creation is provided in Figure 4.

---

<sup>5</sup> A general discussion of the product generation environment can be found in: “Minnesota DNR, Management Information Services Bureau, “Product Generation Toolbox”, March, 2001.

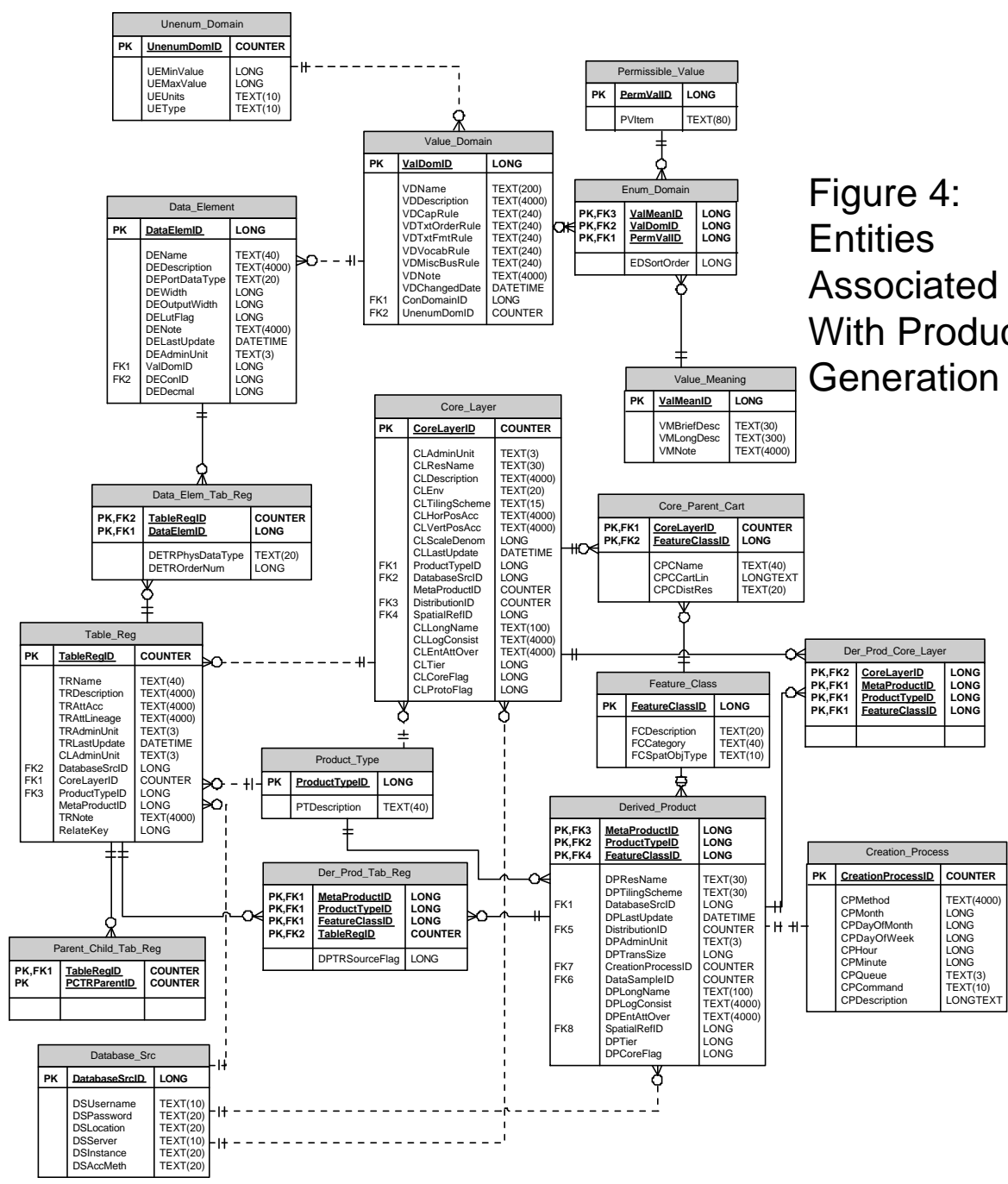


Figure 4:  
Entities  
Associated  
With Product  
Generation

## 5.0 Future Directions

The EDD is a production database that should exist in one form or another for at least the next six to eight years, surviving two more general changes in base level GIS software. It is designed to be neutral with regards to data characteristics and could continue to function even if the department changed their entire GIS application architecture. MIS staff anticipate a steady migration of so-called “Core” data resources into new data formats and new application environments. In this process, the EDD will serve as a stabilizing influence as we systematically make these transitions and document the resulting changes in the physical descriptions of the data.

The EDD will also be the basis for managing the content of DNR Data Deli, version 2, which will be a complete rewrite of DNR’s popular data distribution website. In this context, the “Deli” data sources will be a particular Data Resource Site type that will be managed identically to other DRS’s.

## Bibliography

Dublin Core Metadata Initiative. (1999). Dublin core metadata element set, version 1.1: Reference description. URL: <http://dublincore.org/documents/dces/>

Federal Geographic Data Commission, Content standard for digital geospatial metadata, <http://www.fgdc.gov/metadata/contstan.html>, 1998.

Federal Geographic Data Commission, National Spatial Infrastructure, <http://www.fgdc.gov/nsdi/nsdi.html>, 1997.

ISO/IEC. Specification and standardization of data elements, <http://www.diffuse.org/meta.html#ISO11179>, 2001.

Governor's Council of Geographic Information. Minnesota geographic metadata guidelines, <http://www.gis.state.mn.us/stds/metadata.htm>, 1998.

Marble, Duane. The extended data dictionary: a critical element in building viable spatial databases, Proceedings of the 11<sup>th</sup> Annual ESRI User Conference, 1991.

Minnesota Land Management Information Center, Minnesota Planning, <http://geogateway.state.mn.us/documents/index.html?>, 2001.

## Appendix A:

# The Extended Data Dictionary Data Model





Minnesota Department of Natural Resources  
**Extended Data Dictionary Project**  
 Physical Data Structure  
 Data Structure Diagram  
 Note: This model is still under development

Draft: 01.25.2002  
 Tool: Visio 2000  
 Filename: dd\_012502.vsd  
 Custodian: Hal Watson  
 Contacts: Robert Maki, Hal Watson

