

Digital Geologic Database Model

Stephen M. Richard

Arizona Geological Survey

Draft v. 1.0

July 14, 1998

Contents

.....	2
INTRODUCTION	3
DATABASE OBJECTIVES	3
<u>Database requirements</u>	<u>3</u>
<u>Information management functions</u>	<u>3</u>
<u>Querying/analysis functions</u>	<u>4</u>
<u>Derivative map functions</u>	<u>4</u>
<u>Map Layout/printing functions</u>	<u>4</u>
PHILOSOPHY OF GEOLOGIC DATA MODEL	4
DATA MODEL SUMMARY	5
DATA MODEL DETAILS	8
<u>Entity Domain</u>	<u>8</u>
Spatial Entities	8
Observation or Measurement Entities	9
<u>Classification Domain</u>	<u>9</u>
Database-defined classifications:	9
User defined classifications:	10
<u>Relationship Domain</u>	<u>10</u>
<u>Visualization Domain</u>	<u>11</u>
<u>Symbolization Domain</u>	<u>12</u>
INTERCHANGE FORMAT	13
IMPLEMENTATION ISSUES	13
<u>Relationship tables</u>	<u>14</u>
<u>Hierarchical trees</u>	<u>14</u>
<u>Map Explanation</u>	<u>14</u>
<u>Other</u>	<u>15</u>
IMPLEMENTATION EXAMPLES	15
<u>Stratigraphic section</u>	<u>15</u>
<u>Description of Mappable Units</u>	<u>15</u>
<u>Structural measurements</u>	<u>16</u>
<u>Geochronology</u>	<u>16</u>
<u>Age assignments</u>	<u>16</u>
<u>Cartographic Object (symbol) definitions</u>	<u>16</u>
REFERENCES	16

INTRODUCTION

Modern data storage and communication technology has created an opportunity to rethink the manner in which geologic information is archived and presented. The National Geologic Map Database must provide a mechanism to allow rapid access to up-to-date geologic data in a form tailored to the needs of the user requesting the data. Several groups of users need to be accounted for in the planning of the database: curiosity-driven users from the general public, land managers/planners needing information pertinent to regulatory, planning, and development functions, mineral exploration geologists, and researchers in search of detailed technical information. The data model underlying this database must be flexible enough to encompass a wide range of earth science information, storing it in such a fashion that advances in geologic science do not obsolete the database.

One of the major goals of a geologic database is to simplify the process of generating derivative maps tailored to specific needs—e.g. tectonic analysis, hazard assessment, mineral exploration. Such maps may need to be based on information from many sources. Presently, derivative maps are usually compiled by manually extracting (i.e. tracing on paper) information from a number of printed maps, probably generated by different people with different motivations at different scales, and graphically combining the information (i.e. redrafting) into a single, derivative map. The goal of the geologic database is to provide basic information from a variety of sources in a consistent, flexible, and searchable structure.

The model outlined here provides a conceptual framework that could accommodate newly acquired field data or data ‘mined’ from previously published sources into a common framework. The starting points for this model are Model 4.2 of Johnson, Raines and Brodaric [1998], and the geologic classifications proposed in the EARS data model [Matti et al., 1997a, b, c].

DATABASE OBJECTIVES

The goal of this geologic database is to archive data in a standard structure to allow querying, retrieval, and display of geologic information tailored to the needs of environmental, exploration, and research geologists.

Development of a detailed list of functional objectives is an essential first step for any database design. The requirements listed below are the basis for designing a data model.

Database requirements

a set of minimal or core information needs to be specified to determine if a dataset is complete

model needs to be extensible to allow for new sorts of data or interpretation

a standard vocabulary of terms needs to be defined to make data from different sources compatible, and insure consistent and unambiguous usage..

Information management functions

Compilation/Archive to produce and maintain a statewide geologic information system with data derived from regional and detailed sources. Continuous updating must be possible, as well as source tracking.

Geologic database should provide an archive for as much basic geologic information, independent of interpretation, as possible.

Geologic database should provide an archive for cataloging and describing classifications/interpretations of data.

Database supplies a set of standard classifications to allow consistent attribution of features

Database must track the origin of data and interpretation/classification objects.

Track updating of map data for irregular areas as new data become available

An interchange format must be specified to allow use of data by other producers and users of geologic data (USGS, ALRIS, academic and business community, public).

Must be able to relate geographic data objects to data in a standard database software environment (e.g. xbase compatible).

Querying/analysis functions

These are but a few of the sorts of queries that must be possible in the framework of the geologic database.

identify regions of overlap between polygons with different attribute features (Laramide granodiorite with sericitic alteration)

select arcs based on adjacency of polygons with different attributes features (intrusive contacts between limestone and mafic igneous rocks)

identify arcs bounding polygons meeting some criteria (faults cutting rocks < 15 Ma)

select point data based on features of enclosing polygons. (mines in Miocene silicic volcanic rocks)

select polygons based on proximity to points or arcs and properties of the polygon (analyzing earthquake-induced liquefaction potential)

Derivative map functions

The database needs to be general enough to support production of derivative maps to investigate ore deposit distribution and genesis, geologic hazards, environmental concerns, tectonic analysis, and geologic history.

ability to define derivative map units based on any combination of properties (combine formations of a group into a single map unit; make map showing average grain size of surficial units.)

ability to select subsets of points/lines/polygons within and area for inclusion (filter out minor faults...)

Map Layout/printing functions

The system will be used to support map layouts to print maps on demand.

Record symbolization system used for classification objects.

Specify standard geologic map symbols and line ornaments

Define and use standard color sets (using CMYK, to match our existing map color schemes).

record all parameters used to define a particular map layout/visualization

PHILOSOPHY OF GEOLOGIC DATA MODEL

A geologic data model can serve as a framework for collecting and recording observations and knowledge about the earth. Such a model needs a solid foundation on the basic data (observations, measurements, relationships) used for analyzing geology. These fundamental data entities can be divided into 3 classes: spatial data (location of contacts and points..), quantitative measurements (bedding orientation, fracture density, cobble count, isotopic age determination, quartz c-axis fabric...), and descriptions (rock, fossil, surface, soil, contact relationships, geometric relationships....). These data classes should form the basis for the root tables of the model.

Spatial data is typically recorded on a base map as the location of boundaries of rock bodies. With the addition of the topographic information on the base map, the lines drawn on the map can be thought of as 3-D traces of the intersection of a geologically identifiable surfaces with the earth's surface. This data is one basis for interpretation of the 3-D geometry of rock bodies.

Quantitative measurements can be scalar (magnitude of magnetic field, isotopic date from a rock...), unit vectors (orientation of surfaces...), vector with magnitude (slip vector for fault, gravity or magnetic field...), or arrays (cobble counts, fracture orientation distribution...). These measurements serve a wide variety of

purposes, but are generally associated with a point or area with small dimensions relative to the map scale. The values obtained for these features are ideally independent of the observer collecting the data.

Descriptive data can be associated with a point of observation, or can be a generalized description applied to a surface (fault, contact...) or rock volume (arc or polygon on a map). For most of the commonly described geologic features, there are common systems of classification that denote certain characteristic features. Rock and fossil names are examples of such classification systems. If a contact is described as intrusive or depositional, certain features of the contact will be assumed. Thus many observations made in the field are summarized by identifying an observed object as a member of a pre-defined class. Depending on the classification system used and the experience of the geologist making the observations, descriptive 'data' is commonly subjective. The database must include a set of standard classification systems (lithology, stratigraphic ages, structural data types, fossils, mineral names, formal stratigraphic units) to make the terminology as unambiguous and consistent as possible.

Part of the geologist's work in mapping an area is to develop a system of classifications for rocks and contacts (etc...) that can effectively summarize observations and simplify the task of describing the endless variety of things that are actually observed. The key to the classification system developed in each area is the extraction of key features to be used as criteria for class membership in that area. Typically, many of the observed rocks and contacts will readily match previously established criteria for classification developed in nearby areas, or found to be useful in almost all situations (identification of bedding in sedimentary rocks...). In these cases, classifications from the standard classification system definitions can be used. Classifications developed for a specific area (e.g. new map units) must be fully documented in the database, preferably built up from basic data entities. Recording a full description of the classification system used to pigeon hole observations is an essential role of a geologic database.

A standard printed geologic map records the spatial information contained in the location of rock body boundaries on the map. Classification of contact types is generally binary—'fault' or 'contact'. Geologic reasoning (consideration of rock type, e.g. granite, and rock age for adjacent rocks) further allows geologists to study a map and identify contacts as a conformable depositional contact, non-conformity, angular unconformity, intrusive contact, or etc. Accompanying text sometimes describes the nature of contacts that do not conform to any of these simple classifications. Relationships such as overlying and underlying are determined by geologic reasoning and analysis of the geometry of contacts. The system used to classify rocks into the mappable units shown on the map is described in accompanying text. The classification of measurements of orientation of features (bedding, foliation, contacts...) shown on the map is generally indicated by a set of symbols representing standard feature types, with little or not explanation of the criteria used to assign membership to a particular class (in many cases little explanation is necessary). In areas of complex structure, much of the orientation data collected in the field is not shown on the map because of space limitations.

The printed geologic map thus contains a great deal of information that is not explicitly stated, and probably contains only a subset of the data and observations made by the geologist in the field. One of the attractions of a digital geologic database is the possibility of including a record of as much information as there is time or need to record about an area. The geologic data model must facilitate the storage, retrieval, and analysis of information at whatever level of detail it is available—from detailed field notes to a 50-year old reconnaissance map.

DATA MODEL SUMMARY

The logical structure of the model consists of a set of entities that correspond to a physical object or particular location in the real world, and a set of classifications that correspond to idealized concepts that might be applied to many entities, and a set of relationships between these objects. Table 1 provides examples of entities and classifications to which those entities might be assigned. Entities and classifications are both referred to as objects. Sets of entities may be aggregated to define compound entities and sets of classifications may be aggregated to define compound classifications.

Table 1. Entity Objects and corresponding Classification Objects

Entity	Classification that might apply to entity
a particular rock sample	granodiorite
bedding orientation a particular outcrop	bedding
a stratigraphic section	Formation
a particular contact	thrust fault
a microprobe analysis of a particular mineral grain	pyroxene

Relationships between objects are assigned using correlation tables. Relationships are conceptually like verbs that connect objects. The types of relationships define the set of correlation tables required.

Named relationship types are:

Hierarchy: A relationship that defines one object (the child) as a subset or more specific instance of another object (the parent).

Classification: A relationship that defines an object as an instance of a certain classification object.

Correspondence: A relationship that associates two objects

Proportionment: A relationship that defines one object as a proportional component of another.

Disposition: A relationship that defines the arrangement, positioning, or distribution of an object with respect to another object in space or time.

A correlation table identifies two objects and assigns a relationship between them. Relationships between objects may be defined in the context of a compound object (relationship between Formations in a Group). Objects that can be connected by hierarchical relationships are grouped into types.

Table 2. Examples of relationships

Hierarchy:

object type	object	relationship	related object
classification	curvilinear bedding	is a child of	bedding
classification	Formation	is child of	Supergroup

Classification

object type	object	relationship	related object
Class to compound class		approximate bedding	is grouped with all bedding
entity to class	a particular rock boundary trace	is	conformable depositional contact
entity to class	fabric measurement from outcrop XX	is	pressure solution cleavage

Correspondence

object type	object	relationship	related object
entity	lithologic entity X	forms a bed in	Stratigraphic section ZZ
entity	lithologic entity X	forms the rock in	Polygon 23456
entity	fabric measurement (strk,dip)	was obtained at	point 246787

Proportionment

object type	object	relationship	related object
entity	quartz of a particular sort	forms 35% of	lithologic entity X
classification	Formation A	is a member of	Group AAA
entity to class	lithologic entity X	is 20% of	Map unit AA

Disposition:

object type	object	relationship	related object	context
entity	lithologic entity A	is interbedded with lithologic entity B		in map unit AA
classification	Stratigraphic Age 1	continuous through	Stratigraphic Age 2	age of rocks in Map Unit Z
entity	lithologic entity X	overlies	Lithologic entity Y	Stratigraphic section ZZ
classification	Map Unit A	overlies	Map Unit B	

A geologic map is a visualization derived by selecting a set of entities from a particular spatial domain or map extent, assigning each entity to a classification object, assigning graphical elements to symbolize classification objects, and mapping the symbolized objects to a 2-D display based on their distribution in physical space. The set of classification objects used to define a visualization is called a classification scheme. The distribution of the graphical elements that symbolize the classification objects on the 2-D display conveys spatial relationships between objects. A visualization is uniquely defined by the selection of entities to classify, map extent, classification scheme, graphical elements for symbolization, and mapping between real space and 2-D display. The data model includes information that records the components of a particular visualization derived from the database. Such a visualization may be represented by a printed map that has been digitized, or a map derived from the database as a result of a particular analysis.

To make data sets from different sources as unambiguous and consistent as possible, a useful data model must provide a well-defined system of descriptive terms, entities, classifications, and relationships. Development of a set of such objects that will serve to describe everything that might be observed is clearly not possible. However, definition of a basic ‘vocabulary’ that can describe and classify a great deal of earth science is clearly possible—after all, if this couldn’t be done, geologists could not communicate their findings to other geologists. This vocabulary can be automated as a set of dictionaries and database-defined entities and classifications. Two major dictionaries are proposed: a dictionary of descriptive terms and a dictionary of relationships. Lookup functions could be facilitated by classifying the words in these dictionaries into ‘types’ according to the sort of situations to which they apply. A hierarchy from most general to most specific words in these dictionaries would be also be defined by the database.

A group of database-defined entities serve as basic constructs for defining database-defined classifications that are entity based. For example, formal stratigraphic units are defined by a type-section entity. Other database-defined classifications are based on idealized entities, for example a particular lithology is defined in terms of some idealized rock with a particular set of characteristics; there is no specific rock specimen that defines ‘granodiorite.’ Still other database-defined classifications are purely descriptive or interpretive, for example the classification of a planar fabric in a rock body as bedding, pressure solution cleavage, or schistosity. To deal with these various situations, the database needs to include several tables of entities that are the foundation of classification systems, and several tables of classifications, some founded on the database-defined entities, and some purely descriptive. A set of database-defined classifications specifies the types of entities and classifications that are recognized by the data model.

The data model must also allow for situations that do not fit within the pre-defined framework in a satisfactory manner. This requirement is accommodated by the inclusion of free-text fields for additional information.

Table 3. Dictionaries

Table	function
descriptive words classifications	a collection of terms and definitions that can be used to describe entities and classifications
relationships represent	a collection of phrases and definitions of the relationship the phrase is intended to represent

DATA MODEL DETAILS

The data model based on this philosophy includes 5 domains:

Entity Domain

Classification Domain

Relationship Domain

Visualization Domain

Symbolization Domain

Entity Domain

The Entity Domain consists of measurements and observations representing physical entities in the real world, e.g. a particular, discrete rock body, hand sample, bedding exposed and measured in a particular place, an isotopic age from a sample collected in a specific place, or the location of a boundary between rock bodies. Entity domain objects can be broadly classified into spatial entities, and description or measurement entities. Compound entities are defined by grouping individual (singular) entities that together are related to a particular physical entity. Examples of compound entities are stratigraphic sections, and a description of a rock as a compound entity composed of mineral-grain entities. Different data structure definitions are necessary for different types of entities because each type may have a unique set of characteristics.

Database-defined entities of the description or measurement type serve as the basis for defining database-defined classifications. Examples include ‘type-section’ compound entities that contain information about the type section of a formal stratigraphic unit, and ‘stratigraphic time boundary ages’ singular entities that define the accepted age for boundaries between stratigraphic time intervals (e.g. the age of the Cambrian-Ordovician boundary). A formal stratigraphic unit classification object is related to a ‘type section’ entity to be fully specified. The stratigraphic time scale is built up of ‘stratigraphic age’ classification objects defined as the time intervals between related ‘stratigraphic time boundary ages’ entities.

Spatial Entities

Spatial entities record physical locations. Typically the geologist locates points and lines in the field. Points record locations at which measurements or observations were acquired. The lines record the projection onto a 2-D map surface of the intersection between a geologic boundary surface and the earth’s surface. As pointed out by Bain and Giles [1997], maps are sometimes drawn depicting the traces of intersections between geologic boundary surfaces and other surfaces in the earth, e.g. an unconformity, or a particular elevation. A cross section could be thought of as a map showing intersections between a particular vertical plane and geologic boundary surfaces. Expansion of the geologic data model to account for such different ‘intersection surfaces’ (‘geological level’ in the terminology of Bain and Giles [1997]) can be accommodated by adding to the model a description of the geometry of the ‘intersection surface’ for a particular set of spatial objects. Polygons are defined by linking lines or arcs that bound an area. Attributes are typically assigned to polygons to describe characteristics of rocks within the bounded area.

Each spatial entity should be attributed with an ‘accuracy estimate’ field that provides a length estimate of how well the feature is located. Either by definition the units of this estimate should be meters, or a unit field will have to be included. This accuracy must not exceed the numerical precision of the hardware/software implementation of the database. Each spatial entity also needs to have identifiers for the source of the data.

Observation or Measurement Entities

These are the data entities that abstract properties of real-world geology and are used to analyze and classify rocks and structures. Observation entities are descriptions associated with a point of observation, or can be a generalized description applied to a surface (fault, contact...) or rock volume (arc or polygon on a map). For most of the commonly described geologic features, there are common systems of classification that denote certain characteristic features. Rock and fossil names are examples of such classification systems. If a contact is described as intrusive or depositional, certain features of the contact will be assumed. Many observations made in the field are summarized by identifying an observed entity as a member of a pre-defined class. Depending on the classification system used and the experience of the geologist making the observations, descriptive 'data' is commonly subjective. The database must include a set of standard classification systems (lithology, stratigraphic ages, structural data types, fossils, mineral names, formal stratigraphic units) to make the terminology of descriptive data as unambiguous and consistent as possible.

Quantitative measurements can be scalar (magnitude of magnetic field, isotopic date from a rock, thickness of a stratigraphic interval...), unit vectors (orientation of surfaces...), vector with magnitude (slip vector for fault, gravity or magnetic field...), or arrays (cobble counts, fracture orientation distribution...). In many cases the units of measurement will need to be specified as well, either by default or explicitly. Quantitative measurements serve a wide variety of purposes, but are generally associated with a point or an area with small dimensions relative to the map scale. The values obtained for these features are ideally independent of the observer collecting the data.

Classification Domain

Classifications are modeled as classification objects, which includes a name for objects assigned to the class, and a definition of the unique, distinguishing characteristics of objects assigned to the class. Compound classifications are defined by aggregating a group of classification objects into a new classification object. Common geologic classifications are database-defined, and these may constitute most of the classifications used. The most common situations where new classifications must be defined are for areas lacking formal stratigraphic units in which mappable units must be defined based on described lithologic entities, and for derivative maps that aggregate classifications into more general map units.

Database-defined classifications:

In order to make the database useful across sub disciplines, and to facilitate data transfer between users, a set of standard classifications should be defined (data dictionaries; see Giles et al. [1997]). These standard sets of words and definitions include (among others—open to discussion):

Minerals

Lithology

Stratigraphic age

Structural data types

Soil Taxonomy

Fossil Names

Metamorphic zones

Alteration classification

Formal Stratigraphic nomenclature

Map projections

These database-defined classifications are constructed in terms of the same set of entities as any classification object of the same type in the database. The mechanism for maintaining the database-defined classifi-

cation system for a standard data model needs to be carefully planned. Certain types of classifications may have different definitions in different parts of the country. The model could even allow for different database-defined classifications to be used for the same types of entities in the same area as long as the classification system used is defined, and there is a mapping between classification systems used (which of course could be problematic.....)

A set of database-defined classifications describes the types of entities, classes, and relationships defined by the database. The descriptive terms and relationships contained in the word lists are also considered database-defined classifications because they represent abstract concepts. Finally, word dictionaries for descriptive and relationship terms are considered database-defined classifications because they represent abstract concepts as opposed to physical entities.

User defined classifications:

Within a particular area of study, every geologist confronts a real world that conforms to varying degrees with previously established classifications. When the conformity is not good, new classifications must be defined and clearly documented (i.e. described, fit in the existing hierarchies as appropriate, and relationships to other existing classes and entities set up). The most common example is the definition of mappable units when working in a previously unmapped area, or studying an area in great enough detail that new, more specific map units need to be defined. In addition, to allow flexibility, allowance must be made for free text descriptions of basic data entities (i.e. 'description' or 'name' fields) not constrained by the standard classification system. Use of these fields must be in addition to the standard classifications, not instead of the standard classification. Otherwise the ability to search for features will be lost.

Relationship Domain

The relationship domain contains information about hierarchy, classification, correspondence, proportionment, and disposition. A hierarchy is a series in which each element is graded or ranked. Hierarchical relationships define one object (the child) as a subset or more specific instance of another object (the parent). Classification relationships define an object as an instance of a certain classification object. Correspondence relationships associate two objects. Proportionment relationships define one object as a proportional component of another; typically in the description of the component entities in some compound entity (mineral grains in a lithologic entity) or classification object (lithologic entities in a map unit). Disposition relationships define the arrangement, positioning, or distribution of an object with respect to another object in space or time.

Hierarchical relationships can be defined only between entities on the same type of tree or classifications on the same type of tree. The type of a tree is determined by the type of the root object. Lines are defined by points, polygons are defined by lines; for the spatial data objects most of this data management is taken care of by the GIS software environment, and need not form part of the geologic data model. A generalized lithologic entity may be described in more detail by child lithologic entities. A classification may aggregate several other child classifications. Hierarchy is described in two ways. The position of an object within the hierarchy for a particular object type is given by an integer 'level', with 0 representing the root of the hierarchy or most general/inclusive object. A tree defines parent-child relationships by associating a child object with a parent object. A single tree consists of a root object and all the child objects related to that root.

A classification relationship implies that the entity or classification object that is classified is an instance of the abstract idea represented by the classification object. Class membership relationships are between entities and a classification object or between classification objects and a compound classification object. In a compound classification, the component entities do not combine to form the whole, but rather indicate objects that are directly included in the class. An example would be the classification 'bedding' that includes the structural measurement types 'approximate bedding', 'irregular bedding', 'planar bedding', and 'bedding determined by 3-point construction'. Compound classifications imply a hierarchy that should also be recorded in the data structure.

Correspondence relationships are used to associate description or measurement entities with other objects. This association may be one to one, as when a given measurement was made at a particular location or a rock description represents a particular outcrop or outcrop area. The entity-entity correspondence relationship plays the role of the Singular Object correlation tables of Johnson et al. [1998]. This construct requires that each ‘coverage’ of spatial objects be thought of as a separate entity_type, consistent with the concept the like things are put in separate coverages. The entity_type must be unique for each coverage, and corresponds to the ‘Cover_ID’ of Johnson et al. [1998]. Assignment of any entity to a spatial object, which links data with location, is done through the correspondence relationship ‘located at’. Correspondence relationships might associate a surface character entities with a particular lithologic entity as part of its description, associate a protolith lithologic entity with a lithologic entity describing a metamorphic rock, or associate many fossil entities (species) with a single lithologic entity. Finally, correspondence describes the relationship when a descriptive dictionary term is associated with an object.

Proportionment relationships describe situations in which a number of entities are associated with another entity or classification object, and each entity is represented to a certain degree in the associated object. Each component entity is assigned a fractional quantity (>0 and ≤ 1) to indicate what part of the total consists of this component. These quantities are expressed as percentages, and the sum over all the components of an object must be 100%. Proportionment relationships are commonly used for rock description. A given lithologic entity may be defined as a compound entity consisting of various percentages of particular grain entity constituents. A map unit classification object might consist of 60% of a particular sandstone lithologic entity and 40% of a particular shale lithologic entity. In these cases the component entities are based on specific constituent grains or rock outcrops characteristic of the class, and define it by example.

Disposition relationships may be between entities or classifications of different types, and are of two kinds—relationships strictly between two entities, and relationships that are specific to a particular context. The most common disposition relationships describe contacts between lithologic entities like ‘intrudes’ or ‘overlies conformably’. By convention the ‘related entity’ is always the older object in the relationship; a text field contains description of the particulars of the relationship. A context-dependent relationship between two entities is specific to their occurrence as part of a compound entity, or as components of a classification object. An example of the first case would be the correlation of a lithologic entity with a grain entity that describes clasts in a conglomerate lithologic entity; this correlation would indicated that clasts of one lithologic entity are present in another lithologic entity. An example of a relationship between entities in the context of a classification object would be a relationship between lithologic entities that define a map unit classification object; this correlation might describe interbedding of a sandstone lithologic entity with a shale lithologic entity in the map unit classification ‘Bright Angel Shale’ (also a formal stratigraphic unit, which is a child of ‘map unit’). A context-dependent relationships between classification objects could also describe the age distribution of lithologic entities within the time interval between assigned ages if more than one age is associated with a classification object.

Visualization Domain

A visualization is defined by:

Data sources and coverages

Classification Scheme

Map extent, which also defines the map projection

A change in any one of these three factors will result in a new visualization of data from the database, and conversely, specification of all three factors is necessary to duplicate a single visualization. A visualization is conceptually similar to a ‘map’, but the use of a different term is essential to divorce the user’s thinking from the pieces of paper we call maps.

The concept of a data source will become quite complex as the database evolves from a series of tiled

coverages, each digitized from a published map, into a collection of points and lines with different lineage, along with classifications from a variety of sources. A data source may be an update based on new field work over an extent that is a small part of an existing spatial dataset; all the points, arcs, or polygons within the extent of the update area may not be changed. If the database is distributed over a network, different files (in different places) might contain the arcs for various data sources in the same map extent. Most GIS implementations store points, lines, and polygons in separate files, which would then be components of a data source. Different data sources may originate in different map projections.

A data source is defined as the intellectual source of information objects, independent of their map projection or file structure. A particular data source will probably include several separate data sets—points, arc, polygons, possibly more than one of each for conceptually distinct sorts of spatial entities. A particular data source may exist in several map projections. A coverage is a particular file containing spatial data, e.g. points or lines of a particular sort in a particular map projection, with a particular processing history (with respect to automated GIS functions like splining, generalizing, etc.). Each coverage is assigned an ‘entity type’ unique to that coverage. The ‘entity type’ definition must contain the information necessary to locate its data file and identify the source of the spatial data.

The necessary metadata tables to document a particular visualization meant to convey a particular set of observations and interpretations are as follows:

Coverage Source Table correlation: Links to coverages of spatial data to use for the visualization. For a data set developed from field observations directly into a geologic database, the ‘source’ would be a reference to itself. In the future when data sets accumulate containing features from a variety of sources (a large-scale map with many updates attributed to different sources), the source table correlation will need to contain links for every source attached to any record from any table used to construct the visualization.

Map Area Table: This defines a particular map extent used to spatially select objects from the coverages identified by the Coverage Source Table to include in the visualization. It is questionable whether the definition of a particular map extent merits an associated authorship. I think not. A Name, source Org and Source ID, Projection and Resolution should be defined. Part of the database should be a coverage (coverages?) of polygons defining the extents of each Map Area in this table. The projection of a visualization is specified in the Map Area Table. The source coverages must be in the same projection.

Classification Scheme—this is the heart of the map, it is used to logically select spatial objects to display and to determine how they will be displayed, thus dictating what geologic information is conveyed by the map. A classification scheme is a compound classification that selects a group of classifications to be symbolized on a particular visualization. The classification scheme is thus defined in the classification object table.

Symbolization Domain

A standard set of graphical elements and a scheme to determine which graphical elements are used for which classification objects. USGS OFR 94-525 defines a pretty good set of graphical elements and relates them to the most common classification objects. Since any map visualization will include an explanation relating graphical elements and classification objects, standardizing symbolization is a convenience to map viewers, but provides no added analytical or archival functionality in the database. Including a set of symbolization tables will make automated map visualization generation possible.

Rules for symbolizing spatial objects based on their classification, text label strings for polygons and constructing the map explanation are considered part of the symbolization domain. The map explanation scheme table stores information about symbolization for a particular classification scheme. Titles for graphical elements in the explanation are specified. Labels for polygons (e.g. text strings for rock unit labels) could be database-defined for database-defined classification objects, but for greater flexibility, polygon label strings for appropriate classification objects are assigned in the map explanation scheme. The map explana-

tion table contains a link assigning a cartographic object (graphical element) to use for symbolizing each classification object. Rules for the map explanation are cartographic conventions for converting the hierarchical structure of the classification objects used in this classification scheme to an appropriate graphical layout with grouping and headings for each type of object that is symbolized.

INTERCHANGE FORMAT

An important aspect of a standardized database is the development of a standard interchange format. This format should allow all of the user-defined information in a particular dataset to be transferred to another database with the standard logical format. The interchange format would assume that any database-defined constructs (dictionaries, entities, classifications) would exist in any environment to which the data was to be imported. The interchange format does not have to be elegant—it has to be complete, and in a form that can easily be moved between any software or hardware environment. The ESRI export file format (.e00) is an example of this sort of interchange data structure. Such a structure implemented for a standard geologic database would of course have to have a publicly available definition. The definition of an interchange format would allow various physical implementations of the standard model to be used, as long as the content of the implementation was consistent with the standard (use of dictionaries, standard classifications, and constructs for describing entities). For a distributed database designed for real-time query response, the export to interchange format and import from interchange format to local structure would have to be transparent to the user. This aspect of the standard design is perhaps the most critical for a distributed database, but remains to be developed.

IMPLEMENTATION ISSUES

Because database software packages at the Arizona Geological Survey with sufficient power to implement the model are relational, the implementation considered here is based on a relational database. The major issue in implementing this data model is the trade off between grouping the entities, types and relationships into single tables of objects that can be described with a consistent set of fields, and separating objects into tables that isolate particular types of objects in individual tables. For example, all scalar measurement quantities could be stored in a single table, keyed to entity_ID and entity_type; fields would include a measured value, uncertainty, method of acquisition, source identification, and units of measurement. Such a table might include thickness measurements, isotopic ages, rock density determinations, etc. Alternatively, separate tables, with similar fields in each one, might contain each of these kinds of measurements. Another decision of this sort is whether classification schemes, which are compound classifications (the scheme is a classification object whose components are a group of classification objects), should be defined in a separate table or in the classification object table. If a visualization is considered a classification, then the Coverage Source Table becomes a correlation relationship between entities (the coverages) and a classification (the visualization); the precedence value would correspond to the sequence# in the classification correlation table. More thought and experimentation is necessary to determine the optimum set of tables to use.

Another important implementation issue is the degree to which attributes are embedded in the GIS-related spatial object data tables (pat and aat in Arc/Info terms). This adds a great deal of redundancy to the data set, but makes inspecting the data related to a particular spatial object simpler. The important point as I see it is that databases implemented using a standard data model use the same vocabulary (dictionaries of terms), and contain the same kinds of information. Whether the complete lithologic description is attached to every polygon or kept in a separate table and related to polygons through a key is a local implementation issue, as long as the information content is the same in both cases, and the structure of the embedded data is such that it can be exported in a form that can be imported into a more relational implementation.

Any database implementation should be constructed such that the tables containing user-defined objects are maintained separately from tables containing database-defined objects, and that these objects tables are kept in a separate database from the database containing the forms, queries, reports and code that define the user interface. This allows better cross-platform compatibility because the data tables are all that would need to be transferred. Also, updates to the interface could be implemented simply by changing the 'front-end' database, and linking the new interface to the underlying tables.

Relationship tables

Relationships assignments can be grouped into several types that demand specific data structures. A relationship can involve one entity, e.g. when a hierarchy level number is associated with an object. A relationship can involve two objects. Classification relationships, parent-child relationships, and associations between dictionary terms and objects are of this type. A relationship can involve two objects and a relationship term. Correspondence, proportionment and some disposition relationships are of this type. Finally, a relationship can involve two objects, a context object, and a relationship term. Many disposition relationships are of this type. Except for the assignment of descriptive dictionary terms to entities, all of these are in general many-to-many relationships, and assignment must be made through an intervening correlation table. Because entities and classification objects are uniquely identified by their Type and Id, if the entity types and classification types were all unique, all relationships could be described through a single set of correlation tables. Because of the difficulty in ensuring the uniqueness of all entity type and classification type identifiers (whether those are strings—names—or ID numbers), it is simpler to have a set of entity-entity, classification-classification, and entity-classification correlation tables for each of the appropriate types of relationships. Preserving the conceptual distinction between entities and classification objects is also consistent with the goal of maintaining as much separation as possible between observations/measurements and interpretations.

The set of relationship tables included in the implementation diagrammed with this report keeps the entity and class relationship tables separate. The rank and parent-child relationships are described in one hierarchy table. The level# of an object in these tables corresponds to the number of parent-child relationships between the parent and child objects. For objects that are the root of a hierarchy, the parent and child are the same, and the level is 0. Parent-child relationships are defined by records for which the level is 1. The level# of an object is dependent on the choice of the root object; the maximum level#, which is the ultimate rank, is the level# when the root object is at level# 0. Parent-child relationships are always between objects on the same type of tree.

Hierarchical trees

Should hierarchical trees associate a child object only with its immediate parent object (next lower rank level), or should the tree associate a child object with all parent objects down to the root of the hierarchy. Association with only a parent object is more efficient from a storage point of view, but requires software constructs to trace up or down through the hierarchy to determine all the children of a low level object, or the parent objects more than one level down the hierarchy from an object. The disadvantage of adding records to associate a child with all parents back to the root is that updating a parent-child relationship requires modifying more than one record. If the hierarchies are well thought out to begin with they should not need modifying very often. In view of the cheapness of storage, recording all the parent objects seems advantageous.

Map Explanation

The map explanation is the text material included on printed maps that describes the rock units, and explains the significance of graphical symbology. The map explanation is constructed based on information in the Map Explanation Scheme Table. More thought and testing is necessary to determine a good way to record a standard layout of headings, indentations, and grouping of Map Units and Symbols. The hierarchy defined by the Classification Object Hierarchy and Tree tables could be used to establish groups and head-

ings, but there must be some means to customize which headings are used. A separate set of hierarchy and tree records might be allowed as part of a classification/map explanation scheme to set up the layout of the explanation (more thought necessary here). Many geologic maps include descriptive text not specifically related to definition of rock units or structural features. Much of the information in such text can usually be entered into the data structure by careful reading and analysis, but in many cases it would be useful to be able to preserve text blocks for placement on map visualizations. No implementation of this feature is sketched out, but this could be done by attaching text files to ‘containers’ on the visualization layout (more thought necessary here too!).

Other

Should it be possible to choose alternate symbolization schemes? This could be implemented by making the visualizationID a key in the Map Explanation Scheme, instead of the Classification SchemeID. Then to change the symbolization used for a visualization (say you want it in Turkish?) the same classification scheme, map extent and coverages could be used with a different visualization ID. The alternative would be to generate a clone of the classification scheme (same component classification objects), and associate these with the different symbolization via the Map Explanation Scheme table. The second approach is less efficient and obvious (need discussion here...).

IMPLEMENTATION EXAMPLES

In this section, implementation of some standard objects used for geologic description, analysis and visualization is described in terms of the proposed model.

Stratigraphic section

A stratigraphic section is a compound entity constructed of a series of measured intervals, each constructed as a proportional classification object with component lithologic entities. The section is an entity because it represents a sequence observed in a particular physical location. The stacking of the section is recorded by disposition relationships that record the measured interval that overlies each measured interval. Thickness scalar measurements are associated with each measured interval. This same construct could describe a section in a cliff or drill hole, related through an entity-entity correspondence relationship with a point entity, or section measured along a traverse, related through an entity-entity correspondence with an arc entity.

Description of Mappable Units

The approach taken here follows the procedure commonly used to define mappable units in a new map area. Each variety of rock is described in outcrop, and based on these descriptions, a set of characteristic features are defined and serve to recognize that mappable rock unit as distinct from other rock units. A formal stratigraphic unit must have a type section, which because it is in a particular place, is an entity. For informal map units, some particular set of outcrops serve as a reference section (at least in the mind of the geologist as the mapping is done), and a description of the characteristic features of rock in these outcrops serves to define the mappable unit by example. A lithologic entity is a lithologically homogeneous (at some scale of description) rock type. It may be interbedded with other lithologic entities within a map unit, or mixed with other lithologic entities on a meter scale in a migmatite. The lowest-level lithologic entity is a single hand sample; typically descriptions of map units are built up of component lithologic entities that are generalized from a number of rock samples, which may or may not be described separately. In the case where rock sample descriptions are included in the data set and related to sample locations, the rock samples might be correlated with a compound lithologic entity that then becomes a proportional component in a map unit definition. In another case, the lowest-level rock descriptions may be a generalized lithologic entity that includes some variability. Later, more detailed mapping might define map units based on more specific lithologic entities that are children of the generalized entity. A single lithologic entity might be subdivided into different sets of more specific lithologic entities in different areas or by different geologists; thus, it would be the root of more than one hierarchy of child entities. The data model must allow for this sort of ‘richness’ of the data as the database grows.

Structural measurements

Strike and dip or trend and plunge of fabrics in rocks are recorded in 2 component parts: location as point spatial entity, and the measurement as a unit vector quantity entity. An entity-entity correlation relationship associates the measurement with the point, and a classification correlation assigns the measurement to a structure type classification object. This allows for dealing with a great variety of situations, including measurement of several fabrics of different sorts in one location, and, through the relationship tables, describing the relationships between the fabrics and their spatial disposition in the outcrop. It also allows for different classification of a fabric; one observers bedding may be interpreted as cleavage by another geologist.

Geochronology

Detailed information pertaining to isotopic ages should be maintained in a separate database. Isotopic age scalar measurements included as entities in the tables in this geologic map section of the database would be based on a query against the full isotope geochronology database.

Cooling age is one type of radiometric age; to assign a blocking temperature interpreted for the cooling age, use an entity-entity (radiometric age-temperature) relationship to assign blocking temperature for date.

Age assignments

Lithologic entities or classification objects (map units) may be assigned several ages (due to uncertainty or polyphase history). This may be done via a single age correlation, minimum age and maximum age correlation relationships, or correlation with a number of ages. Disposition relationships between the age assignments define the distribution of the rocks in the unit between the age assignments (continuous or discrete periods between, two distinct periods of formation, etc...)

Cartographic Object (symbol) definitions

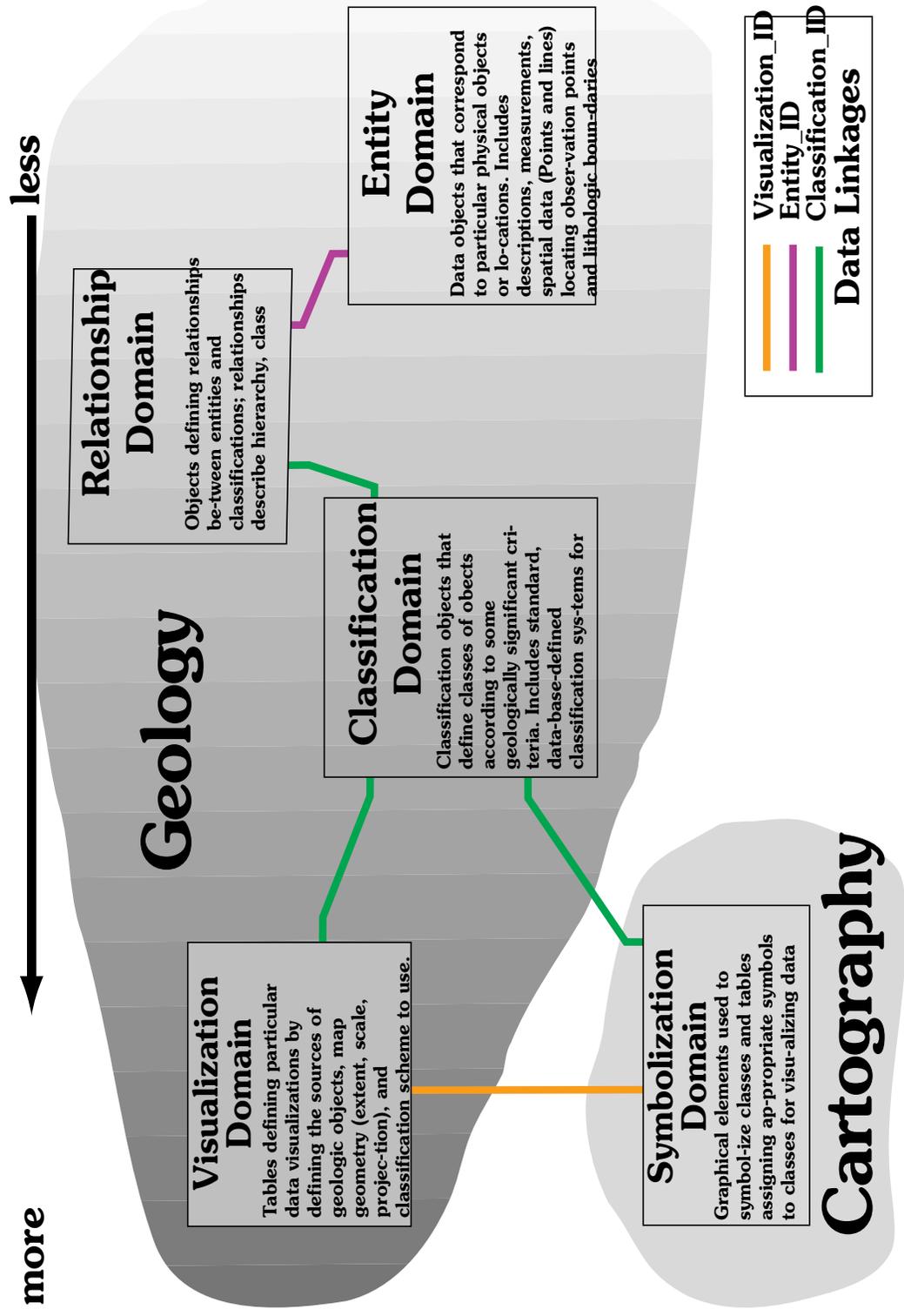
Display priority should be built into the cartographic object definitions or map explanation scheme—that is, it is part of the symbolization domain. This is necessary when an object is symbolized with a pattern superimposed across other patterns or colors. A symbol would be defined as ‘overprinting’ if it needs to be applied on top of other symbols.

REFERENCES

- Bain, K.A., and Giles, J.R.A., 1997, A standard model for storage of geological map data: *Computers & Geosciences*, v. 23, no. 6, pp. 613-620.
- Giles, J.R.A., Lowe, D.J., and Bain, K.A., 1997, Geological Dictionaries—Critical elements of every geological database: *Computers & Geosciences*, v. 23, no. 6, pp. 621-626.
- Johnson, B. R., Brodaric, Boyan, and Raines, G. L., 1998, Digital Geologic Maps Data Model, V. 4.2: <http://ncgmp.usgs.gov/ngmdbproject/standards/datamodel/model42.pdf>, U. S. Geological Survey.
- Matti, Jonathan C., Miller, Fred K., Powell, Robert E., Kennedy, Stephen A., Bunyapanasarn, T. P., Koukladas, Catherine, Hauser, R. M., and Cosette, P. M., 1997a, Geologic-point attributes for digital geologic-map databases produced by the Southern California Areal Mapping Project (SCAMP), Version 1.0: U. S. Geological Survey Open-File Report 97-859, 37 pages.
- Matti, J. C., Miller, F. K., Powell, R. E., Kennedy, S. A., and Cosette, P. M., 1997b, Geologic-polygon attributes for digital geologic-map data bases produced by the Southern California Areal Mapping Project (SCAMP), Version 1.0: U. S. Geological Survey Open-File Report 97-860, 42 pages.
- Matti, Jonathan C., Powell, Robert E., Miller, Fred K., Kennedy, Stephen A., Ruppert, K. R., Morton, G. L., and Cosette, P. M., 1997c, Geologic-line attributes for digital geologic-map databases produced by the Southern California Areal Mapping Project (SCAMP), Version 1.0: U. S. Geological Survey Open-File Report 97-861, 103 pages.

Basic Framework of Geologic Data Model v1.0

Stephen M. Richard, 7/98



Entity domain

correspond to physical objects or locations in real

Geologic Entities; Any of these can be related to other entities to define compound entities or to define classification objects by example. Entities can be classified as members of a Classification object. Many kinds of disposition relationships can be assigned to pairs of entities.

Singular

May be singular or Compound

Compound

Description/Observation Entities

Rock constituents are context dependent:
 1. sedimentary rocks, descriptions of grain type in carbonate rocks, sand grains in sandstone, clasts in conglomerate
 2. igneous and metamorphic rocks: mineral phases that form the rock.
 constituents are correlated with lith entities via the classification correlation

Thickness measurement 1/measurement
*Entity_ID *EntityType_ID *source_ID method type thickness units comment

Rock constituent entity 1/rock constituent
*Entity_ID *EntityType_ID constituent type min diameter max diameter Units shape color description frameworkY/N source_ID

Radiometric Age 1/date
*Entity_ID *EntityType_ID *source_ID date uncertainty method comment

Orientation measurement 1/measurement
*Entity_ID *EntityType_ID *source_ID type method modifier strike dip comment

Surficial geologic entity 1/surficial geo entity
*Entity_ID *EntityType_ID armor dissection morphology etc. source_ID

soil entity 1/soil entity
*entity_ID *entity_type description etc. source_ID

bedrock geologic entity 1/lithologic entity
*Entity_ID *EntityType_ID rock_name Rock_Class Color induration outcrop geomorph depoSetting texture/fabric grain_sorting/texture description source_ID

Stratigraphic Section 1/described section
*Entity_ID *EntityType_ID *source_ID date uncertainty method comment

To Be Defined: Metamorphic zones, Alteration zones, etc

Spatial Objects

Spatial Objects are stored in GIS-implementation dependent file structures. In order to make the Key for each spatial object entity correspond to the keys for other entities, the cover_ID plays the role of Entity_ID in these tables. (Each table represents multiple coverages)

Points 1/point
* Entity_id * cover_id source_org source_id location_accuracy

Arcs 1/arc
* Entity_id * cover_id source_org source_id location_accuracy

Polygons 1/polygon
* Entity_id * cover_id source_org source_ID

Database Defined Entities

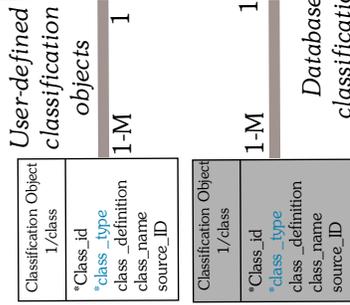
Entities used to define standard classifications that are part of the data model; lithologic classes, soil classes, other classification systems are defined in terms of description constructs used to describe underlying entities, but these are considered part of the database and the detailed definition records and source citations need not be explicitly included.

Formal Unit 1/Formal Unit
*Entity_ID dB_entity_type type_section etc. source_ID

Stratigraphic Time Age boundaries 1/named interval boundary
*entity_id dB_entity_type Age uncertainty older_strat_name younger_strat_name source_ID

Visualization Domain

Classification Objects



Type definitions



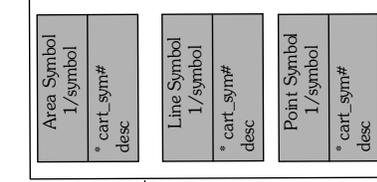
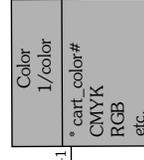
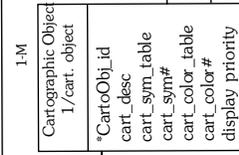
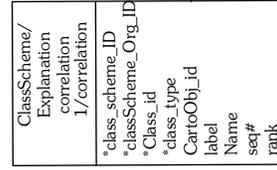
These tables define the types of entities, classifications, and relationships that exist in the model.

Classification Domain

Relationship Domain

View of spatial objects must be clipped to extent of MapArea_ID extent

Map Explanation Scheme



These tables are used to control the graphical display tools of the GIS software environment

Symbolization Domain

Should symbol selection be keyed to the Classification Scheme or the Visualization?

In order to selectively display Spatial Objects within the map extent, spatial objects could be classified using the Visualization_ID, a classification object of type Visualization in the classification relationship tables.

Entity

An entity can enter into the definition of only one classification for a particular visualization

Relationship Domain

For each table there is also an equivalent table of database-defined relationships

Correspondence, Disposition

Hierarchy

Entity Hierarchy	1/parent-child pair
*Entity_type	
*Entity_ID	
*ParentEntity_id	
rank	
level#	

Entity-Entity

Entity-Entity Relationships	1/relationship
*rel_ID	
Entity_ID	
Entity_type	
rel_Entity_ID	
rel_Entity_type	
relation_type	
relation	
comment	
Source_ID	

Entity-context	Entity Relationship
Entity Relationships	1/relationship
*Context_rel_id	
cEntity_type	
cEntity_ID	
entity_type	
Entity_id	
rel_Entity_type	
rel_Entity_id	
relation_type	
relation	
Source_ID	

Entity-Class

Class-context	Entity Relationship
Entity Relationships	1/relationship
*Context_rel_id	
Class_type	
Class_ID	
entity_type	
Entity_id	
rel_Entity_type	
rel_Entity_id	
relation_type	
relation	
Source_ID	

Class-Class

Class Relationship	1/relationship
*Class_rel_id	
Class_type	
Class_id	
rel_Class_type	
rel_Class_id	
relation_type	
relation	
Source_ID	

no context

context

Classification, Proportionment

Compound entity	classification correlation
1/compound entity	
*Entity_ID	
*Entity_type	
*subEntity_id	
*SubEntity_type	
percent	
range of %	
quality	
seq#	
comment	
Source_ID	
Source_org	

Class definition	correlation
1/correlated class	
*Class_id	
*Class_type	
*Entity_id	
*Entity_type	
percent	
range of %	
quality	
seq#	
basis	
comment	
Source_ID	
Source_org	

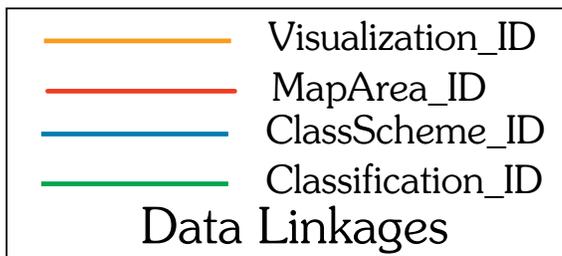
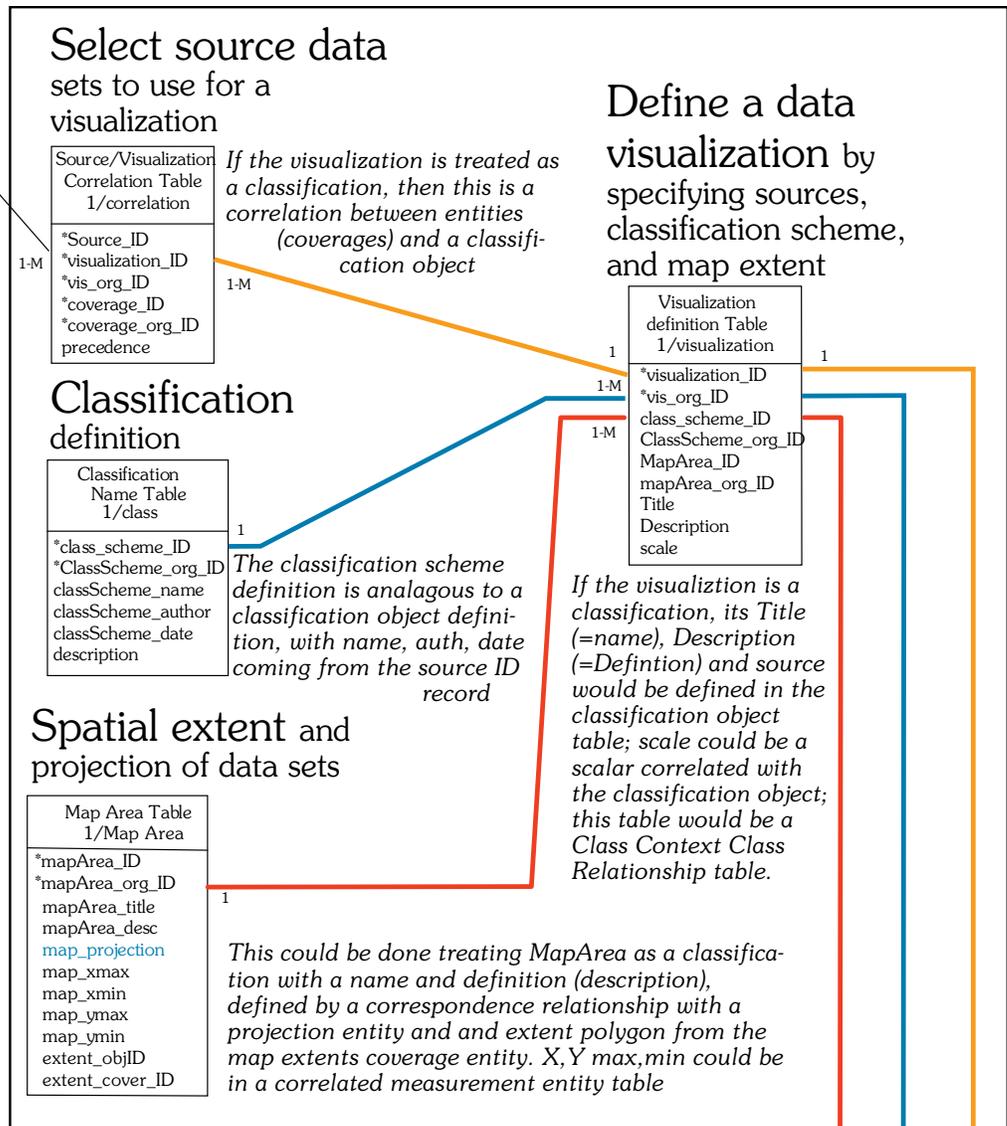
Class/proportional	entity correlation
1/correlated entity	
*Class_id	
*class_type	
*subClass_id	
*subClass_type	
percent	
quality	
seq#	
comment	
Source_ID	
Source_org	

Compound classifications can only be defined for a particular visualization using classification objects that are not symbolized in that visualization. A classification object can enter into the definition of only one compound classification for a particular visualization

Classification

Visualization Domain

Standard FGDC metadata for data sources



View of spatial objects must be clipped to extent of MapArea_ID extent polygon

Classification Scheme: Defines the set of classification objects to display and the symbolization to use for display

Spatial objects selected on visualization_ID