

# Geoscience Terminology Development for the National Geologic Map Database

By Stephen M. Richard<sup>1</sup>, Jonathan Matti,<sup>2</sup> David R. Soller<sup>3</sup>

<sup>1</sup>Arizona Geological Survey  
416 W. Congress # 100  
Tucson, AZ 85701  
Telephone: (520) 770-3500  
Fax: (520) 770-3305  
email: Steve.Richard@azgs.az.gov

<sup>2</sup>U.S. Geological Survey  
e-mail: jmatti@usgs.gov

<sup>3</sup>U. S. Geological Survey  
908 National Center  
Reston, VA 20192  
e-mail: drsoller@usgs.gov

## ABSTRACT

Science language vocabularies are being constructed for use in the U.S. National Geologic Map Database (NGMDB), to provide geologic data to users in government, business, academia, and the general public in a consistent data structure, using consistent descriptive terminology or science language. These terms and definitions are used to classify observable or inferred facts or events, and to assign values for properties in descriptions. The vocabulary makes pre-defined terms available to apply in descriptions without having to reconstruct the entire description denoted by the vocabulary term. Terms within a controlled vocabulary may be *atomic* or *analyzable*. The atomic terms are irreducible in the database environment; they are either numbers, or terms defined by narrative text. The text definition is meant for a human user, and is not amenable to automated analysis. Analyzable terms are specified by combinations of properties and relationships according to some formal description structure. Analyzable terms may be compared, using the values of the properties in their formal description. The atomic terms in a vocabulary may represent complex concepts that could be represented by description schema. Implementation of formal descriptions for terminology associated with a concept allows users to develop alternate or more specific terminology for concepts that may interoperate with other systems by using automated description comparisons (Description Logic).

The unifying feature of the NGMDB is a common conceptual model and terminology system. To function as a part of the NGMDB, individual databases must conform

to, or interoperate with, the common conceptual model and shared terminology. The simplest route to NGMDB conformance is to use the standard science vocabularies. If desired, users can build their own vocabularies, but in order to be NGMDB conformable, user-defined vocabularies will need to: 1) define the scope of their vocabulary; 2) provide text definitions of each vocabulary term; 3) build formal definitions rooted in the standard science vocabulary, using description schema from the common conceptual model, and a standard, well documented syntax; and 4) provide a correlation of each term with the most specific subsuming term from the standard vocabulary.

## INTRODUCTION

Science language vocabularies are being constructed for use in the National Geologic Map Database (NGMDB), a distributed, federated information system that is planned for integrating geologic data provided by state and federal geological surveys. The goal is to provide geologic data to users in government, business and academia in a consistent data structure, using consistent descriptive terminology. A conceptual model referred to as NADM-C1, proposed as a standard data structure for geologic map data, has been developed by the Data Model Design Team of the North American Data Model Steering Committee (NADMSC) and is under review (Data Model Design Team, 2003). The National Geologic Map Database Project will use this model as a foundation for database development. A Science Language Technical Team of the NADMSC has developed a

draft science language that is currently in review as well (<http://geology.usgs.gov/dm/steering/teams/language/charter.shtml>), and the NGMDB will use this vocabulary with extensions as necessary. The use of a clearly defined vocabulary, with definitions developed to avoid ambiguity in the use of terms, is essential for consistent communication of science concepts to a wide variety of data users with varying geologic expertise.

Implementation of any geoscience information system requires a collection of science language terms and definitions to populate descriptions of rocks, geologic units, and structures (Giles and others, 1997). A collection of terms, their definitions, and relationships between the terms (especially hierarchical relationships) are referred to here as a controlled vocabulary or terminology system (de Keizer and Abu-Hanna, 2000, de Keizer and others, 2000). The data model implemented by a database dictates the kinds of controlled vocabularies required to actually populate the database. For example, the NADM-C1 models 'WeatheringCharacter' as a property of a geologic unit. This property is specified by a term from a science language vocabulary, thus requiring a controlled vocabulary of terms that specify different weathering character values. Controlled vocabularies are used to populate fields in a database in order to make possible the clear and unambiguous communication of database content. The fundamental difficulty is the variability in usage of many common geoscience terms. Consider the various usages of the term 'shale':

- "Shale is a laminated or fissile claystone or siltstone....To those claystones which are neither fissile nor laminated but are blocky or massive, the term mudstone may be applied." (Pettijohn, 1957, p. 341)
- "Shale...is rock that splits parallel to bedding into thin flakes or plates." (Williams and others, 1954, p. 316);
- "In this book, the author uses the term shale for all sedimentary rocks composed dominantly of mud-size (<0.6 mm) particles." (Boggs, 2001, p. 153).
- "A widely used and often loosely used field term for mudstone which shows a conspicuous fissility on weathering. It is somewhat unsatisfactory in that the weathering state must play a part in its recognition and it cannot be consistently used in comparing rock at outcrop with, say, that of a borehole core" (Collinson and Thompson, 1989, p. 54).
- According to Potter and others (1980, table 1.2), "mudshale" and "clayshale" are terms restricted to rocks that are laminated (<10 mm) and that have specific ranges of clay-size constituents: 33-65% (mudshale) and 66-100% (clayshale);
- "Shale is the term used for fissile mudrock and, more generally, for the entire class of fine grained sedimentary rocks that contain substantial quantities of clay minerals." (Blatt and others, 1980, p.

374). These authors emphasize fissility as a core requirement for the definition of shale (also see Folk, 1968).

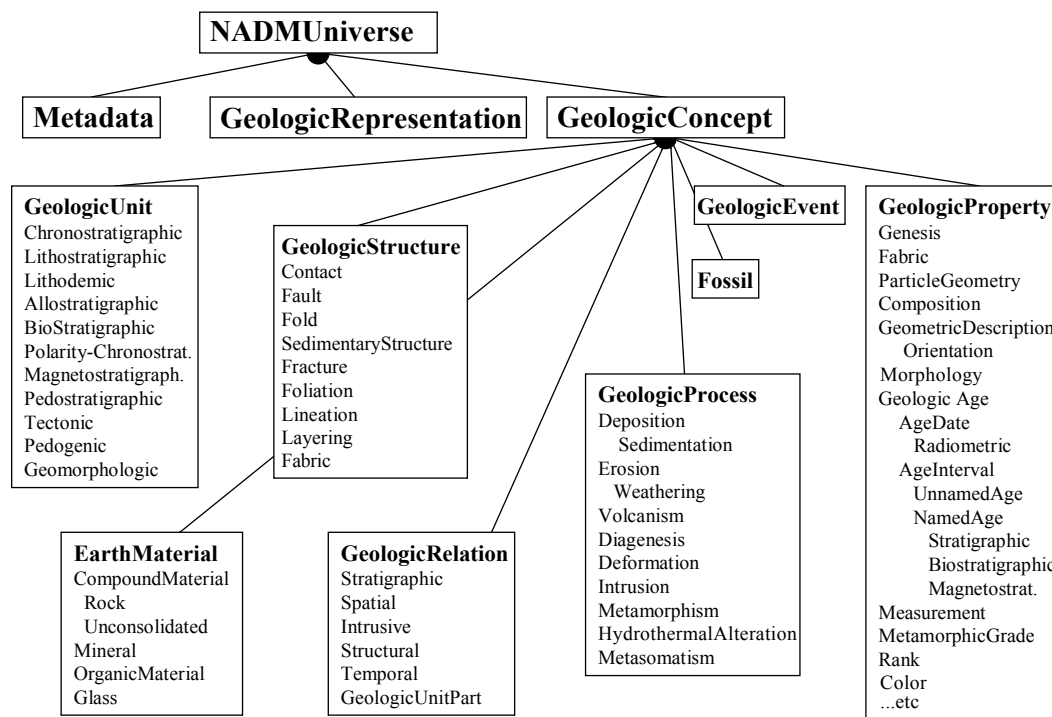
It should be apparent that the use of shale as a rock name in a database, without some connection with the definition of shale the user is implying, does not communicate unambiguously what kind of rock is being described. Standardized terminology is being developed in order to avoid these sorts of problems.

In this paper, the NADM conceptual data model (NADM-C1) is first reviewed because it defines the framework within which the NGMDB science language is being developed. Then the design of the NGMDB science language is described.

## NADM CONCEPTUAL MODEL

The NADM-C1 for geologic map information, version 1.0 (Data Model Design Team, 2003) is designed to be a technology-neutral conceptual model that can form the basis for a web-based interchange format using evolving information technology (such as Extensible Markup Language, or XML), and for the design of database implementations that can interoperate using the interchange format. The intended purpose is to allow the sharing of geologic information independent of logical and physical implementations. NADM-C1 is a model of geoscience concepts and the relationships between them, with special emphasis on concepts related to information presented on geologic maps (Richard, 1999 and in press; Brodaric and Gahegan, 2000; Brodaric and Hastings, 2002; Soller and others, 2002).

The top level of NADM-C1 is 'NADMUniverse,' representing all concepts in the model. Three subclasses of NADMUniverse are represented: GeologicConcept, Metadata, and GeologicRepresentation (fig. 1). The metadata part of NADM-C1 is outside the scope of this paper and will not be discussed. Concept names in the text written using the Courier font refer to NADM-C1 model elements, and concept names in plain text refer to geoscience concepts in general. The term 'concept' as used here represents the notion of any mental phenomena that human beings use in their internal representation of the world. Webster's dictionary (<http://www.m-w.com/>) uses the terms 'idea' and 'object of thought' to convey the meaning of 'concept.' GeologicConcept includes concepts specific to the domain of geoscience knowledge. Figure 1 shows a generalized version of the modeled GeologicConcept hierarchy. The top-level concept that pertains to geologic maps and geologic information (loosely defined) is called GeologicConcept. The names in the boxes beneath the major concept names represent additional concepts that inherit from GeologicConcept. In other words, GeologicStructure is a GeologicConcept and Contact is a GeologicStructure and, therefore, a GeologicConcept. Table 1 defines major geologic concepts in NADM-C1.



**Figure 1.** High-level geologic concept hierarchy for NADM-C1.

**Table 1.** Major sub-concepts of GeologicConcept in NADM-C1.

Concept	Scope and rationale
EarthMaterial	A naturally occurring substance in the earth. EarthMaterial represents substance, and is thus independent of quantity or location. Ideally, an EarthMaterial is defined strictly on the basis of physical properties, but because of standard geologic usage, genetic interpretations commonly enter into the definition as well. Does not include melted rock (magma or lava). Many concepts related to water or petroleum have not been modeled in this version.
Fossil	The remains, trace, or imprint of a life form that has been preserved in an EarthMaterial, and that demonstrates evidence of having been changed from its original biogenic form. Fossil is distinguished from biologic remains or biogenic sedimentary structure based on evidence of having been converted incipiently or substantially into a modified version of the original biogenic form or structure. Although the passage of time is implicit in the definition of Fossil, no constraint is placed on the amount of time necessary to become a fossil.
GeologicEvent	An identifiable event during which one or more geologic processes act to modify a geologic entity (Earth material, geologic unit, or geologic structure). A GeologicEvent may have geologic properties such as a specified GeologicAge and Geologic Environment. An example might be a cratonic uplift event during which erosion, sedimentation, and volcanism occur.
GeologicProcess	A function, possibly complex, that acts on one geologic entity to produce another geologic entity at a later time. Process is time independent; some GeologicProcesses are observable today in the field or in the laboratory, others can only be inferred from observing the results of the process. Processes take one or more of EarthMaterial, GeologicUnit, or GeologicStructure as input and have one or more of EarthMaterial, GeologicUnit or GeologicStructure as output.
GeologicProperty	An inherent feature used to characterize a GeologicConcept. Some examples include physical properties, color, age, grain shape, metamorphic grade, and weathering character.

GeologicRelation	Any of a wide variety of relationships that can exist between two or more GeologicConcepts. For example, the GeologicRelation 'intrudes' is a relationship between an intrusive igneous rock and some host rock. Includes spatial, temporal, sequence, correlation, and parent/child relations. Many of the relationships in NADM-C1.0 (particularly attribute links and parent-child links) are not explicitly modeled as kinds of GeologicRelation, but are represented as associations in the model schema.
GeologicStructure	A configuration of matter in the Earth based on describable inhomogeneity, pattern, or fracture. The identity of a geologic structure is independent of the material that is the substrate for the structure. Properties like 'clast-supported', 'matrix-supported', and 'graded bed' that do not involve orientation are considered kinds of GeologicStructure because they depend on the configuration of parts of a rock body. Includes sedimentary structures.
GeologicUnit	A naturally occurring body of material distinguished from adjoining material on the basis of content (lithologic or fossil), inherent attributes, physical limits, geologic age, or some other property or properties (adapted from North American Commission on Stratigraphic Nomenclature, 1983, p.22; <a href="http://www.agiweb.org/nacsn/code2.html">http://www.agiweb.org/nacsn/code2.html</a> ). Corresponds to 'stratigraphic unit' in the North American Stratigraphic Code. Commonly used properties include composition, texture, included fossils, magnetic signature, radioactivity, seismic velocity, and age. Sufficient care is required in defining the boundaries of a unit to enable others to distinguish the material body from those adjoining it (North American Commission on Stratigraphic Nomenclature, 1983).

---

GeologicRepresentation includes concepts related to the representation of geoscience knowledge, and this paper deals particularly with one representation concept—GeologicVocabulary. A geologic vocabulary is a collection of controlled concepts, referred to here as terms. Each of these is associated with a preferred name, and terms are usually organized in some logical fashion such as a hierarchy. The preferred name for a controlled concept in any particular vocabulary provides a standard means of identifying the associated GeologicConcept in that context, and can be thought of as a proxy for the collection of property values and relationships specified in the definition of the controlled concept. Examples of geologic vocabulary include a collection of standard rock types, a stratigraphic lexicon, or a geologic time scale.

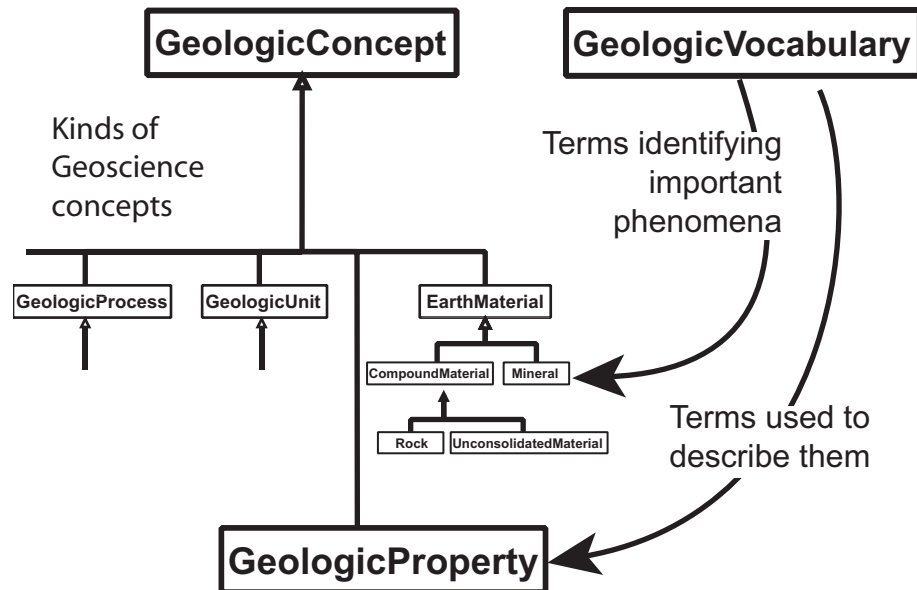
A description is a collection of property values and relationships that apply to some thing (in the most general sense), specifying the nature of the thing such that it may be identified. A term from a science language vocabulary implies the properties and relationships specified in its normative description. The normative description may be a simple body of text, in which case the controlled concept is not computationally analyzable. Other normative descriptions may be specified by a description scheme that establishes the properties and relationships that may be assigned values to build a description. Such descriptions may be analyzed using software inference tools to determine subsumption and class membership relationships. For example, a rock is defined as 'a compound material that is consolidated'. The description scheme for compound material specifies that (among other things) a compound material has a 'consolidation degree' property. This property is specified by a term from the consolidation degree terms vocabulary. Given a collection of compound

material descriptions, all instances of rock could be identified by selecting for those with a consolidation degree property value that is 'consolidated' or some term that is a child of 'consolidated' (such as 'strongly indurated').

Geologic vocabularies play a central role in the application of NADM-C1 for storing geologic information. Terms from geologic vocabularies are used to classify observable or inferred facts or events (phenomena), and to describe them by assigning values for various properties (fig. 2). Controlled vocabularies of rock names, kinds of geologic structures, or geologic units (stratigraphic lexicon) are used to classify observations to be recorded in a database. The vocabulary makes the definitions of controlled concepts available to apply in other descriptions without having to reconstruct the entire description denoted by the term. For example, in the description of a geologic unit, a constituent may be identified simply as 'granite'—a term from a standard lithology vocabulary, which has an associated normative description that follows the description scheme for a compound material as specified in the NADM-C1 model. Because the controlled term 'granite' is linked to a description, no other data entry is necessary in order to infer that the properties of 'granite' (consolidation, color, composition, etc.) apply to some part of the geologic unit. The NADM-C1 model identifies numerous geologic properties that are used in description or definition of geologic concepts. Each of these properties requires an associated vocabulary of terms to specify possible values for that property in a database.

As an example, consider the following text from a published geologic report:

"The diabase is typically dark gray or green-



**Figure 2.** Relationship between NADM-C1 conceptual model and science language. GeologicVocabulary is a model element representing a controlled vocabulary or terminology system that contains the words used to enumerate important phenomena and to specify property values in formal descriptions; for example, a lithology lexicon is a collection of terms (science language, vocabulary) that correspond to different kinds of compound Earth materials.

ish gray. It [grain size] ranges from aphanitic to coarse-grained, and the coarser grained diabase has ophitic texture.” (Peterson, 1962)

In this sentence, the term diabase identifies a rock type that must be defined in some controlled vocabulary if it is to convey a precise meaning. The description here specifies three properties—color, grain-size, and fabric. The values for these properties is given using terms “dark gray” or “greenish gray” for color, “aphanitic” and “coarse-grained” specifying a grain-size range, and “ophitic” specifying a fabric or relationship between mineral grains in the rock. These terms would have to be defined in a controlled vocabulary in order to have unambiguous meaning in a database.

## DESIGN OF NGMDB TERMINOLOGY

The NGMDB terminology system design consists of three parts. First is the actual list of terms that encompass the range of concepts that will be represented in the information system. Second is a specification of how descriptions are constructed to define concepts within the system. These description schema are defined by a conceptual data model that identifies the objects, properties, and relationships inherent in concepts to be represented in a database. The NGMDB science language uses description schema from the NADM-C1 (Data Model Design Team, 2003)

model, along with some extensions to that model. For example, the description of the thickness of a lithostratigraphic unit includes the following properties: 1) reported minimum thickness; 2) reported maximum thickness; 3) typical thickness reported or interpreted; 4) source of typical thickness (field geologist, person who parsed the map legend, or other); 5) reported unit of measure; 6) original free text description or, for new data, a summary text description; 7) comments and/or processing notes from the person who parsed the legend; and 8) reference to a spatial extent (such as a rock body, borehole, outcrop, some part of a quadrangle). The description scheme also specifies the cardinality for each attribute (required, optional). The third part of the terminology system is the collection of definitions for terms included in the system.

The NGMDB terminology will attempt to follow the criteria of non-vagueness, non-ambiguity, and non-redundancy (de Keizer and Abu-Hanna, 2000). Non-vagueness prescribes that definitions of vocabulary terms must be sufficient to identify a particular concept. Non-ambiguity prescribes that definitions must address exactly one term. Non-redundancy specifies that a description has only one corresponding most-specific term. Because the NGMDB is intended to be useful to a wide variety of users, definitions of terms must be intelligible to non-geologists. In short, definitions of terms need to be clear, unambiguous, and non-overlapping.

The NGMDB science language is a concept vocabu-

lary. The identity of a controlled concept (represented by a term from the vocabulary) is based on its definition, not the actual word used. The preferred name or term associated with the concept in the vocabulary is a label for the concept used in human communication. Each controlled concept has these properties:

- Unique identifier
- Name (a term)
- Definition
- Tracking (information on source of term)

Hierarchical arrangement of concepts is fundamental to many aspects of human thinking (Murphy and Lassaline, 1997). Most of the concepts that geologists use have a hierarchical structure, from very general to very specific; and for any observation or interpretation, terms consistent with the geologist's level of confidence are used. Hierarchical relationships between complex concepts are non-unique. For example, bedding partings in a sedimentary rock are a kind of planar structure and a kind of sedimentary structure. Multiple hierarchies may be defined, each valid and useful in some context. Implementation of the science language should include support for multiple explicitly defined hierarchies (Brodaric and others, 2002). For terms that have formal definitions constructed in the information system, description logic inference tools can determine hierarchy based on subsumption relationships between the normative descriptions.

## Terms

Terms (controlled concepts) within a controlled vocabulary may be *atomic* or *analyzable*. The atomic terms are irreducible in the database environment; they are either numbers, or terms defined by narrative text. The text definition is meant to convey the meaning of a term to a human user, but is not amenable to automated analysis other than a simple string or number comparison of the identifier for the term. In the example given above, atomic terms specify color (dark gray) and fabric (ophitic). Another example would be terms used to describe outcrop character, such as "bouldery" or "ledgy".

Analyzable terms are specified by a linked collection of properties and relationships according to some formal description structure (description schema, see below). Analyzable terms may be compared on the basis of the values of the properties in their formal description. In the above example from Peterson (1962), the grain-size property is specified using analyzable terms "aphanitic" and "coarse-grained". These terms correspond to some quantitative range of particle diameters in the rock, commonly specified by a minimum and maximum diameter for the size-range represented by each term. The description associated with an analyzable grain size term

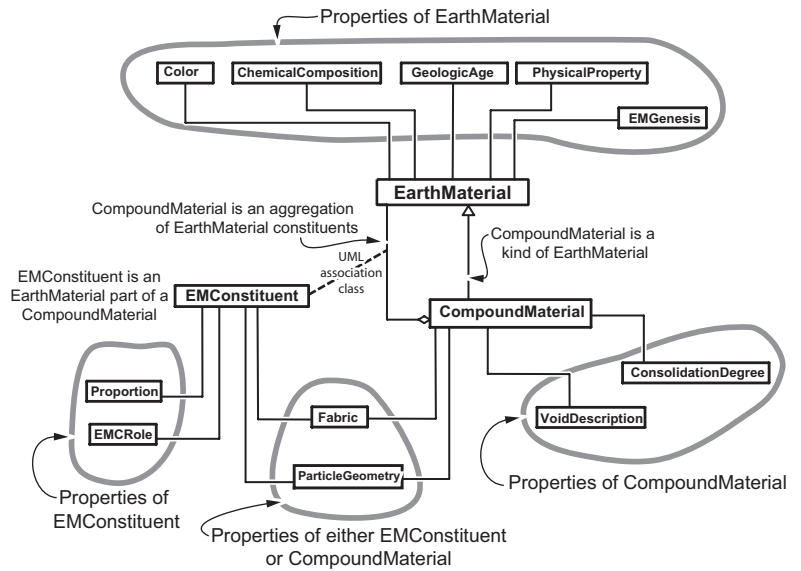
might thus consist of a minimum diameter property and a maximum diameter property, each specified by a length property. The length property might in turn be specified by a description that includes a quantity property and a measurement unit property. The atomic concepts in this description would be the numbers that specify quantity, and terms that specify measurement systems (millimeters, inches, etc..).

The atomic terms in a vocabulary may be conceptually simple, but in many cases, atomic terms may represent complex concepts that could be represented by description schema. The choice of when to model and implement the description of a complex concept or to represent the concept using an atomic term is based on the usefulness of having an analyzable description. Implementation of formal descriptions for terminology associated with some concept allows users to develop alternate terminology, or more specific terminology, for the concept that may be integrated with other terminology systems that use the same description scheme for that concept. The granularity (level of detail) of terms in a vocabulary must be considered in the design of a practical terminology system. A vocabulary consisting only of atomic terms is relatively simple to implement, but is difficult to extend and customize. Such a vocabulary would probably be limited to more general terminology because of the difficulty in defining specialized terminology in such a way that it will gain acceptance. All hierarchical relationships would have to be explicitly represented. Processing queries in such a system would be relatively simple, and could probably be done using Structured Query Language (SQL). Systems in which formal descriptions are constructed for terms would allow users to customize terminology without losing the ability to integrate their data with others. Hierarchical relationships based on subsumption could be determined using description-logic inference software. The descriptions would still have to be based on a common terminology, but at a conceptually simpler level, for which wide agreement on definition is easier to attain. The cost is that such systems are more difficult to implement, and the query processing might require logical analysis of descriptions well beyond the capability of SQL databases, necessitating a logic-based inference processor.

## Description Schema

A description schema is a specification of the properties associated with a concept, and its relation to other concepts. Particular instances of the concept are characterized by particular values for these properties and relationships. The NADM-C1 model includes description schemes for major geoscience concepts: materials, geologic units, and structures. Figure 3 shows a simplified version of the NADM-C1 description scheme for Earth-Material. This scheme is the template for description of

**Figure 3.** Schema for description of EarthMaterial from NADM-C1. UML notation (Rumbaugh and others, 1999) is used; multiplicities for association ends are not shown. EMConstituent is a UML association class, which represents a link between objects in which each link instance has individual attributes. Circles drawn with thick gray lines group properties that have the same range.



any material component of the Earth. An abstract class named CompoundMaterial represents materials that are aggregates of particles, and includes rocks and unconsolidated material. The scheme allows any material to be described by color, chemical composition, age, genesis (geologic history), or physical property attributes. A compound material (a rock) is described as an aggregation of constituent parts, each of which may be any Earth material. Some attributes are specific to compound materials, such as the degree of consolidation and description of the void spaces between particles. Some attributes are specific to constituents (an occurrence of an Earth material as part of a compound material), namely the proportion that the constituent forms in the aggregation, and a role property that specifies the relationship of the constituent to the whole aggregate material (for example, phenocryst, groundmass, matrix, framework, etc.). Finally, there are attributes that may apply to either individual constituents or to the aggregate as a whole, such as fabric and particle geometry. For instance, plagioclase may be a constituent as phenocrysts, with a grain size (part of particle geometry, see below) of 2-5 mm. Alternatively, a particular sandstone (compound material) may be described as fine-grained (a grain size attribute, part of the particle geometry group of properties).

Properties may be specified using atomic terms from a geologic vocabulary. For instance, consolidation degree in figure 3 may have an associated hierarchical vocabulary of atomic terms (fig. 4). Other properties may themselves have analyzable structure, with descriptions of property values in terms of other (generally simpler) properties. In this case, a vocabulary may be thought of as a collection of property descriptions, with a corresponding term for each description. The terms from this vocabulary effectively specify property values from the corresponding description.

As a demonstration of the relationship between an analyzable property description and a geologic vocabulary, figure 5 shows a detail of particle size description. Particle size is a component in the description of particle geometry, a property of either a compound material or a material constituent (fig. 3). Particle size may be specified quantitatively in a variety of ways. In figure 5, the simple approach of specifying a maximum and minimum grain diameter is the description scheme for a quantitative particle size description (QuantitativePSD). Each diameter is in turn a measured quantity, for which the description scheme specifies a numeric quantity and a measurement unit (determined by an atomic term from a vocabulary of measurement unit systems). The table in the lower left of the figure provides an example of four quantitative particle size descriptions. Several terminology systems have been proposed for specifying particle size in rocks. Terms from one of these systems may be used to populate a qualitative (terminological) particle size description (QualitativePSD in fig. 5). The table in the lower right of figure 5 includes several terms from one vocabulary for particle size (Wentworth, 1922). Each of these terms has a corresponding quantitative particle size description; thus they are analyzable terms.

## APPLICATION OF NGMDB TERMINOLOGY

The National Geologic Map Database is envisioned as a distributed, federated information system, with nodes controlled by state geological surveys, the U. S. Geological Survey, and perhaps by interested business and academic collaborators. Sheth and Larson (1990) discuss the general framework and some approaches to such information systems. The NGMDB system will be distributed be-

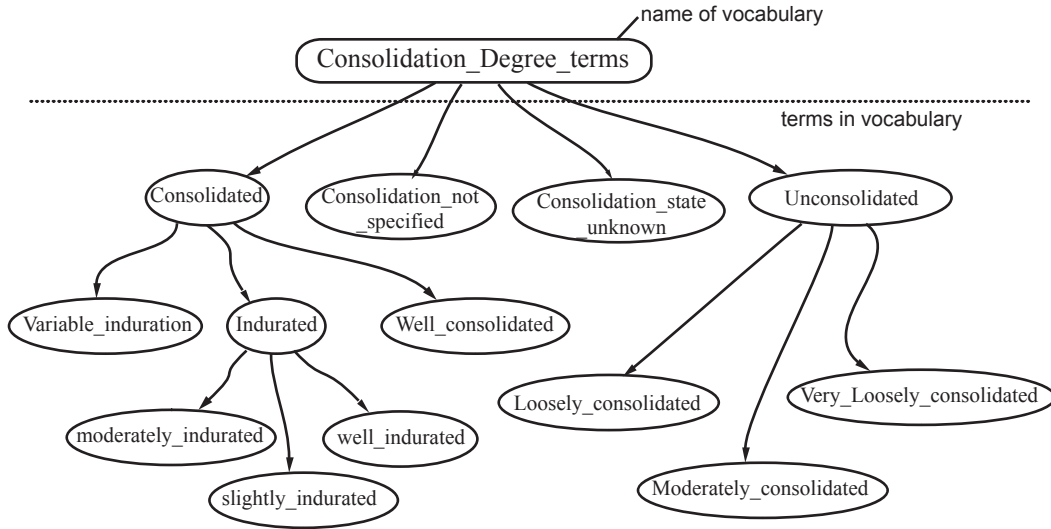


Figure 4. Vocabulary for consolidation degree property.

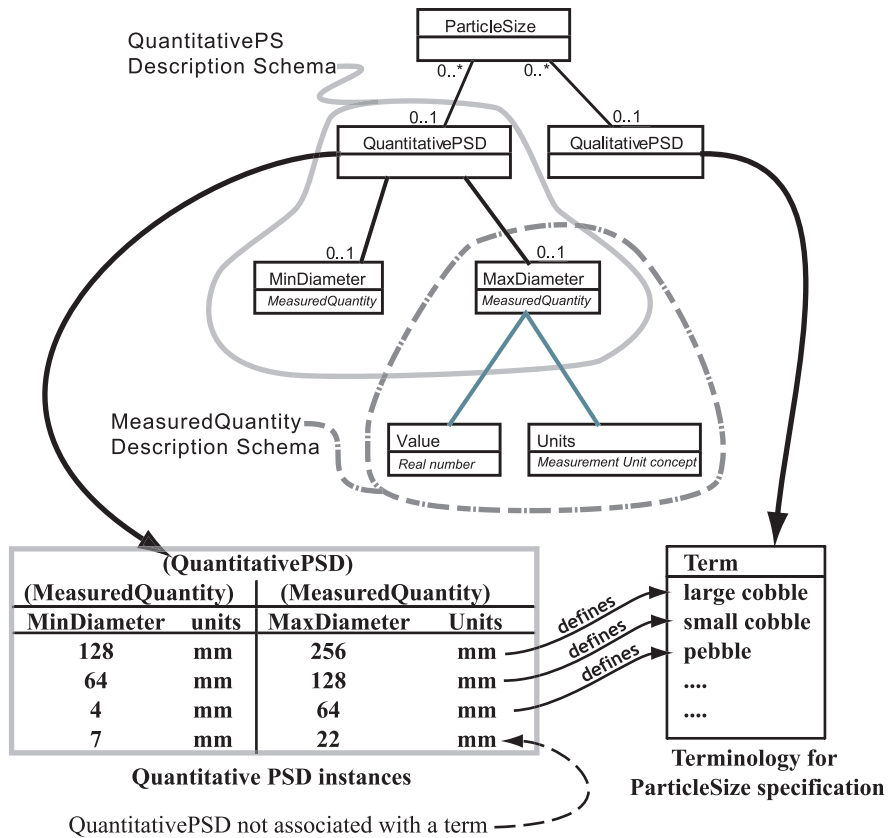


Figure 5. Particle size description. PSD is an abbreviation for 'ParticleSizeDescription'.



cause data are created and managed by different agencies. The system is federated in that all nodes on the system will interoperate using a common conceptual model and a shared terminology. Individual nodes may implement the conceptual model and standard terminology directly, or may implement something different and interoperate with the network using a software transformation between their local data model and the common, conceptual data model. The common conceptual model will include NADM-C1, with extensions for description of geologic units and inclusion of feature level metadata.

Because the unifying feature of the NGMDB is the common conceptual model and shared terminology, to function as a distributed system each agency's database must conform to, or interoperate with, the common conceptual model and shared terminology. The simplest route to NGMDB conformance is to use the standard science vocabularies. If desired, users can modify the NGMDB standard vocabulary or build their own vocabularies. In order to be considered NGMDB-conformant, user-defined vocabularies likely will need to:

1. Define the scope of their vocabulary or (terminology)
2. Define terms in a human-intelligible form (text)
3. Build formal definitions rooted in the standard science vocabulary, using description schema from the common conceptual model, and a standard, well documented syntax. A standardized format is necessary but has not yet been determined.
4. Provide a correlation of each term in the vocabulary with the most specific subsuming term from the standard vocabulary.

The NGMDB Science Language is now under development, and will be made publicly available to facilitate implementation of conformant databases. The NGMDB terminology is intended to include:

1. Vocabularies for all geologic properties identified in the NADM-C1 data model
2. Hierarchical vocabularies for classifying rocks and unconsolidated materials, geologic structures, and geologic processes (EarthMaterial, GeologicStructure and GeologicProcess from NADM-C1)
3. Terminology required for the NGMDB relational database implementation of NADM-C1 (for example, vocabulary for measurement units)
4. Terminology for additional phenomena, properties, and metadata deemed necessary by collaborators in NGMDB prototype projects (such as engineering properties).

Figure 6 shows the framework for some of the vocabularies that eventually will be needed for the National Geologic Map Database. Clearly, development of these vo-

cabularies is an ambitious goal that will take some time to achieve. Fortunately, other efforts are underway that share the same objectives, particularly the science language activities of the North American Data Model Steering Committee Science Language Technical Team (SLTT), this volume; <<http://geology.usgs.gov/dm/steering/teams/language/charter.shtml>>, and the Cyberinfrastructure For The Geosciences (GEON) project, a large, National Science Foundation-funded effort to develop cyberinfrastructure for geological sciences. Development of NGMDB science language is thoroughly integrated with science language development by the SLTT. Draft language developed by the SLTT is being incorporated directly into the draft NGMDB science language vocabularies. The National Geologic Map database project is actively exploring avenues of collaboration with GEON to accelerate development of formal geologic vocabularies for use in the construction of geologic information systems.

## REFERENCES

- Blatt, H., Middleton, G.V., and Murray, R., 1980, *Origin of sedimentary rocks* (2nd ed.): Englewood Cliffs, New Jersey, Prentice-Hall, Inc., 782 p.
- Boggs, Sam, Jr., 2001, *Principles of sedimentology and stratigraphy* (3rd ed.): Upper Saddle River, New Jersey, Prentice-Hall, Inc., 726 p.
- Brodaric, B. M., Boisvert, E. and Patera, A., 2002, A set-theoretic technique and software for managing multiple-classification in geological hierarchies, *in* Proceedings, International Association of Mathematical Geologists: annual meeting.
- Brodaric, Boyan, and Gahegan, M., 2000, Geoscience map data models, Open systems GIS and semantics, *in* Proceedings, GeoCanada2000-The Millennium Geoscience Summit, Calgary, Alberta, p. 7.
- Brodaric, Boyan, and Hastings, Jordan, 2002, An object model for geologic map information, *in* Richardson, D., and van Oosterom, P., eds., *Advances in Spatial Data Handling*, 10th International Symposium on Spatial Data Handling: Heidelberg, Germany, Springer-Verlag, 562 p.
- Brodaric, Boyan, Boisvert, Eric, Patera, Antonio, 2002, A set-theoretic technique and software for managing multiple-classification in geological hierarchies: 2002 Annual Conference of the International Association for Mathematical Geology, Berlin.
- Collinson, J.D., and Thompson, D.B., 1989, *Sedimentary structures* (2nd ed.): London, Unwin Hyman, 207 p.
- Data Model Design Team, 2003, NADM conceptual model 1.0, A conceptual model for geologic map information: North American Data Model Steering Committee, available at [http://geology.usgs.gov/dm/steering/teams/design/NADM-C1.0/NADMC1\\_0.pdf](http://geology.usgs.gov/dm/steering/teams/design/NADM-C1.0/NADMC1_0.pdf) (accessed 02/04/2004), 55 pages.
- de Keizer, N.F., and Abu-Hanna, A., 2000, Understanding terminological systems II: Experience with conceptual and formal representation of structure: *Methods of Information in Medicine*, v. 39, p. 22-29.
- de Keizer, N.F., Abu-Hanna, A., and Zwetsloot-Schonk, J.H.M., 2000, Understanding terminological systems I - Terminology and typology: *Methods of Information in Medicine*, v. 39, p. 16-21.

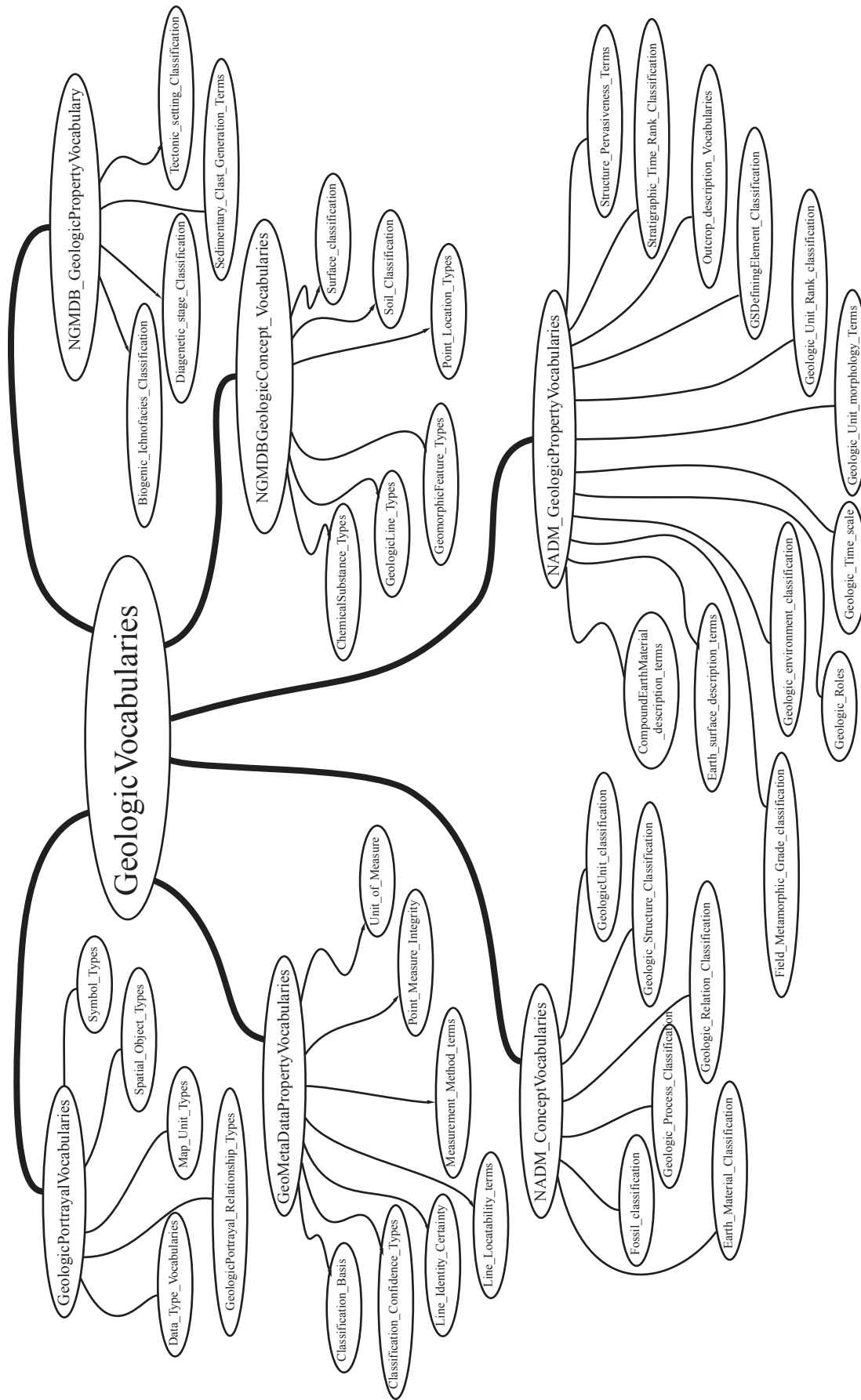


Figure 6. Framework of geologic vocabularies that will be important for the National Geologic Map Database.

- Folk, R.L., 1968, Petrology of sedimentary rocks: Austin, Texas, Hemphill's Bookstore, 170 p.
- Giles, J.R.A., Lowe, D.J., Bain, K.A., 1997: Geological dictionaries--critical elements of every geological database: Computers & Geosciences, v. 23, 621-626.
- Murphy, G. L., and Lassaline, M. E., 1997, Hierarchical structure in concepts and the basic level of categorization, *in* Shanks, David, and Lamberts, Koen, eds., Knowledge, concepts, and categories: Cambridge, Ma., The MIT Press, p. 93-132.
- North American Commission on Stratigraphic Nomenclature, 1983, North American stratigraphic code: American Association of Petroleum Geologists Bulletin, v. 67, no. 5, p. 841-875.
- Peterson, N. P., 1962, Geology and ore deposits of the Globe-Miami District, Arizona: U. S. Geological Survey Professional Paper 342, 151 pages.
- Pettijohn, F.J., 1957, Sedimentary rocks (2nd ed.): New York, Harper Brothers, 719 p.
- Potter, P.E., Maynard, J.B., and Pryor, W.A., 1980, Sedimentology of shale: New York, Heidelberg, Berlin, Springer-Verlag, 306 p.
- Richard, S. M., 1999, Geologic concept modeling, with examples for lithology and some other basic geoscience features, *in* Soller, D. R., Ed., Digital Mapping Techniques 1999, Workshop Proceedings, U. S. Geological Survey Open-File Report 99-386, p. 59-75, <<http://pubs.usgs.gov/of/1999/of99-386/richard.html>>.
- Richard, S. M., in press, Geoscience concept models, *in* Sinha, K, and Baru, C, Geoinformatics: Geological Society of America Special Paper.
- Rumbaugh, J., Jacobson, I., Booch, G., 1999, The unified modeling language reference manual: Reading, Ma., Addison-Wesley, 576 p.
- Sheth, A.P., and Larson, J.A., 1990, Federated database systems for managing distributed, heterogeneous, and autonomous databases: ACM Computing Surveys, v. 22, p. 183-236.
- Soller, D. R., Brodaric, Boyan, Hastings, J. T., Wahl, Ron, and Weisenfluh, G. A., 2002, The central Kentucky prototype: An object-oriented geologic map data model for the National Geologic Map Database: U. S. Geological Survey Open-File Report 02-202, 39 p., <<http://pubs.usgs.gov/of/2002/of02-202/>>.
- Wentworth, C.K., 1922, A scale of grade and class terms for clastic sediments: Journal of Geology, v. 30, p. 377-392.
- Williams, H., Turner, F.J., and Gilbert, C.M., 1954, Petrography: San Francisco, W.H. Freeman and Co., 406 p.