How Can Classificatory Structures Be Used to Improve Science Education?

Olha Buchel and Anita Coleman

There is increasing evidence that libraries, traditional and digital, must support learning, especially the acquisition and enhancement of scientific reasoning skills. This paper discusses how classificatory structures, such as a faceted thesaurus, can be enhanced for novice science learning. Physical geography is used as the domain discipline, and the Alexandria Digital Earth Prototype project provides the test bed for instructional materials and user analyses. The use of concept maps and topic maps for developing digital learning spaces is briefly discussed.

Olha Buchel (buchel@library.ucsb. edu) is Research Associate at the University of California at Santa Barbara, and **Anita Coleman** (asc@ u.arizona.edu) is Assistant Professor at the School of Information Resources and Library Science, University of Arizona, Tucson.

Knowledge structures, such as classification schemes and thesauruses, are not often thought of or used as pedagogical systems. Also known as classificatory structures or knowledge organization schemes, they are used to organize knowledge for retrieval in libraries. Along with other information systems in the library, the online public access catalog and bibliographic databases, these tools have generally been designed to meet the information needs of the user who is a researcher or professional librarian, not the novice learner. Trends in interdisciplinary study and research, the widespread availability of electronic information resources, and the interest of funding agencies in the development of educational digital libraries provide an incentive to investigate how one type of classificatory structure, the faceted thesaurus, can facilitate science teaching and learning.

First, we discuss how scientific reasoning as a general educational objective and learning outcome can provide a framework for the design of digital learning spaces. The goal of learning spaces is to facilitate the acquisition of scientific learning and reasoning skills in novice learners (Coleman 2001). Next, we explain the geosciences knowledge domain, with special emphasis on our test area, physical geography. We highlight some of the similarities between information organization for learning and knowledge organization in libraries. We find that concepts and relationships, classification, and vocabulary are critical components of both activities. Finally, we discuss the use of the faceted thesaurus as the foundation for digital learning spaces. We identify the enhancements needed for developing learning spaces in physical geography. The use of concept maps (Novak 2001) and topic maps (XML 2000; ISO/IEC 1999) is briefly discussed. The Alexandria Digital Earth Prototype (ADEPT) project provides the test bed for instructional materials and user analyses. ADEPT is supported by the National Science Foundation Digital Libraries Initiative Phase 2, and is a successor to the Alexandria Digital Library (ADL) project. For more information, see www.alexandria.ucsb.edu.

The Conceptual Framework

The conceptual framework rests on a synthesis of cognitive learning theories about scientific reasoning and concepts. Many theories discuss learning and define the characteristic steps of learning: component display (Merrill 1983), information-processing theory (Miller 1956), modes of learning (Rumelhart and Norman 1981), and mental models (Mayer 1989). From the perspective of cognitive processes that make up a scientific reasoning skill set, these theories show general agreement that scientific reasoning is closely related to problem solving and that it involves both inductive and deductive reasoning. A spatial representation theory of reasoning suggests that people do better when visualizing things, and therefore successful scientific reasoning includes observation and visualization skills (Leonard 1997). Ziman in his discussion of scientific research and knowledge provides a summary of many important aspects of science, such as patterns of fact, differentiating facts into categories, skills of observation, accuracy, relevancy, explanation, description, generality, and extensive reliance on and use of instrumentation, measurement, and models (Ziman 1984). From these theories and research studies, it is possible to derive an operational definition of the selective skills and critical steps in learning that are an integral part of the scientific reasoning process.

We define scientific reasoning as inductive and deductive thinking. Inductive thinking includes concept development, which is composed of concept acquisition, concept formation, and concept mapping. Deductive thinking includes hypothesis development, which comprises discovery, observation, model building, and evaluation/proof formulation based on empirical evidence. This definition can be correlated with educational objectives when teaching, and with learning outcomes while or after being exposed to learning activities such as lectures and laboratories. For example, training in concept mapping has been shown to facilitate acquisition of text information (McCormick and Pressley 1997). Therefore, in measurable terms of cognitive processes, scientific reasoning comprises:

- Concept acquisition (gathering facts, definitions)
- Concept formation and analysis (identification of associated properties, processes, phenomena, measuring equipments, observational methods)
- Concept mapping (specification or comprehension of relationships)
- Instantiation (illustration with examples and nonexamples)
- Generalization and categorization (definition of abstract general relevant properties, finding similarities with other members of the broader category)
- Problem formulation (designing an experiment; selection of parameters for manipulation)

- Hypothesis generation (discovery)
- Explanation (elaboration)
- Prediction (accurate, relevant calculation)
- Evaluation (interpretation of outcomes; use of empirical evidence)

This is neither a comprehensive, definitive, nor sequential list of scientific reasoning skills. It provides the theoretical framework for our design and prototype development of learning spaces. In this framework, concepts emerge as the foundational units for facilitating science learning. For a similar approach based on concepts, see Smith, et al. (2002).

There are a number of definitions of the word "concept" in the learning sciences. These definitions have been contributed by psychologists, educators, philosophers, linguists, and cognitive scientists and include the following:

- A concept is an idea or thought, more precisely the abstraction that represents or signifies the unifying principle of various distinct particulars (Barrow and Milburn 1990).
- Concepts represent the fundamental elements of all concept areas. In formal content situations, concepts are classes of objects, symbols, and events that are grouped together in some fashion by shared characteristics (Husen 1994).
- Most concepts are structured mental representations that encode a set of necessary and sufficient conditions for their application, if possible, in sensory or perceptual terms (Laurence and Margolis 1999).

These definitions show that concepts are not studied in an isolated manner; rather, they are studied in their associations or relationships with other concepts.

An important component of instructional design is the analysis of the concepts to be learned. Two basic types of analysis are: (a) content task analysis, which focuses on defining the critical characteristics of the concepts and the relationship of those characteristics according to superordinate and subordinate organizations, and (b) contextual analysis, which focuses on the memory and organization of the concepts (Husen 1994). Both types of concept analysis imply the specification of relationships among concepts. These relationships can be restated in the terminology of librarians: analyzing characteristics and superordinate and subordinate organizations of concepts, students define generic-specific and object-property types of relationships. During contextual analysis, students determine other types of relationships between concepts, which librarians know as associative relationships. Associations can also be defined between objects and processes, objects and events, tools and methods, etc.

Related to concepts, but not quite the same from the standpoint of educators, are the words "subject," "topic," and "class." The subject is a field of knowledge or established area of instruction, for example, the subject of mathematics. Subjects are often associated with disciplines. Teachers, scholars, and other workers tend to specialize in the study or use of a particular body of knowledge—or we could say subject/discipline. Subjects involve not merely their particular subject matter, but a particular kind of activity related to it. Topics, on the other hand, are subjects that invite treatment by a number of disciplines. The topic may involve a number of concepts from different subjects, or it may be led or dominated by a single subject. The topic web is a schematic, annotated way of planning the topic, showing how a variety of ideas, activities, subject areas, or skills are related to the core idea (Blake and Hanley 1995). Class is a group, set, or kind of things sharing common attributes. It is frequently associated with the set: a number of things of the same kind that belong or are used together.

Theoretically, a discipline determines its subject, identifies topics, breaks them down into classes and concepts, determines relationships, and teaches them via a number of activities. Physical geography is the domain used here to investigate how concepts are organized in classroom teaching and learning.

The Discipline of Physical Geography

Document analysis was used to examine the organization of content in one textbook as well as traditional classroom lectures supplemented by Microsoft PowerPoint presentations. We chose materials from introductory freshman and sophomore courses in physical geography. Informal interviews with two teaching faculty supplemented the formal document analyses. Curriculum materials used include Christopherson (2000).

Document analysis is a method of research that is used to study historical documents, usually primary source materials. This method can be used to investigate details like document type, date, creator, as well as answer questions like why was the document written and what can be inferred about the document creator and other pertinent subject matters. Education staff at the National Archives and Historical Administration have created and made publicly available via their Web site a number of document analysis worksheets for different types of documents, such as maps, text, etc. (NARA 2001).

Lectures and textbooks have been the primary tools of Western education for some time now. Lecture materials are unusual in that they can be considered as both primary and secondary source materials. Textbooks are clearly secondary source materials. Good lectures summarize, synthe-

size, and present a vast amount of material in a bite-sized chunk. Textbooks provide explanation, corroboration, and pointers to more materials on the subject. Both types of materials were examined in order to identify key aspects of their organizational structure.

A limitation of these analyses is the lack of observation of real learning activities, evaluation, and user studies of physical geography learning in students. However, we feel that such studies, while useful, should be preceded by a clear understanding of the nature of the discipline as perceived and presented by its expert teachers.

We found that organization by concepts is the preferred method for presenting learning material. Concepts are the building blocks in the educational process. Instructors teaching concepts also defined the terminology and explained the relationships among concepts. Thus, in geography teaching, a variety of resources for particular natural processes or phenomena are presented as terms selected and defined by the instructor; and relationships within and external to other processes, phenomena, tools, methods, classifications, theories, and states are explained, explored, and studied.

One of the key aspects of geography learning at the undergraduate level is vocabulary—terminological lists, lists of standard terms, and their definitions. Additionally, many geographic terms and concepts represent details of natural phenomena and require pictorial explanation. For example, alluvial fans and geologic folds are explained with text, verbal analogies, images, diagrams, maps, and photographs. Educators create personalized collections of images of natural phenomena, processes, and objects. These are accompanied by definitions, which sometimes are cross-linked with others, offered as a glossary, and used for presenting new material to students.

Scientific classifications are of great importance. The textbook contained about 70 classifications, ranging from objects (soils, rocks, minerals) to phenomena (hurricanes, tornadoes) to spatial and temporal divisions such as geologic time periods. Additionally, instructional materials pointed to a number of other classifications of objects and phenomena. For example, there are more than 2,000 coordinate systems alone.

A final aspect of organizing for learning in geography is the attention given to the expression (representation) of geographical concepts and their relationships using mathematics—measurements from instruments for specific concepts, equations that specify relationships. Through computation, most of these are ultimately transformed and represented as visualizations such as climographs, hydrographs, hypsographic curves, etc. These visual and mathematical representations are used extensively to promote basic scientific interpretation of complex phenomena and processes, often not possible by mere observation.

Real-world phenomena taught in physical geography are presented under disciplinary aspects of geography, geology, physics, biology, chemistry, astronomy, and other science and engineering disciplines. Synthesis of diverse perspectives is considered to be an outcome and strength of geographical knowledge and is consistently highlighted in presentations. This contrasts directly with the widely held view of science as essentially reductionist in nature. But it fits with the study of geography as an applied science, dealing with measurements, forecasting and modeling, and interpretation of natural phenomena.

Similarities and Differences between Organization for Learning and Knowledge Organization in Libraries

Summarizing, the two activities of organization for learning and organization for information retrieval in libraries appear to have many similarities. Concepts, relationships, and classifications are also key tools that are used to organize knowledge in libraries. Librarians, like educators, use the same concept-related terminology (but with somewhat different meanings), specify the same relationships, and are involved in similar processes related to concept analyses. How are library concepts different from concepts used by educators? Are terms, classes, facets, and subjects the same as concepts? How are concepts arranged in library classification schemes? How do librarians analyze concepts?

For librarians, a concept is a knowledge unit with similar characteristics. Often, the term "concept" is used interchangeably with words such as "term," "subject," "subject heading," "topic," and "facet." "Terms" are the main components of thesauruses, while "subjects" or "subject headings" comprise the subject heading lists, like Library of Congress Subject Headings (LCSH) or Dewey Decimal Classification. Unlike terms, which mainly include concepts from a specific domain related to phenomena, subject headings—human constructs—may include different types of concepts, specifically names, time periods, form, and topics. The main differences between terms and subjects can be described as follows:

1. In general, a term denotes a single concept, while a subject heading may consist of composites of terms, although it also may consist of a single concept (Dykstra 1988).
2. The guidelines for thesauruses give rules for establishing hierarchical relationships and for assigning associative and hierarchical terms. LCSH also has rules that are used when establishing new headings; however, composite headings are more difficult to relate than terms, and there remain many headings

and relationships that were established before the rules were made (Dykstra 1988).

The term "topic" is frequently used interchangeably with the terms "subject" and "subject heading," or "topical subject." Topic represents an aspect of the main subject other than form, place, or period—for example, headings: Libraries, Agriculture.

"Subject" is defined as any one of the topics or themes of a work, stated explicitly in the title or text or implicitly in its message. In library cataloging, books and other items are assigned one or more subject headings that represent their content to assist users in locating information by subject. In indexes and bibliographic databases, the subject headings assigned to documents are called descriptors. Topics and subjects in library classifications are associated with document aboutness. In library cataloging, subject analysis has traditionally been carried out on the summarization level that is finding the one overall subject concept that encompasses or can represent what the whole item is about. Alternatively, the 20% rule is invoked where 20% of the document is about the subject (Taylor 1999).

In library and information science, class is "the first order of structure in a hierarchical classification, at which level major disciplines are represented. A class may incorporate one or more divisions, which in turn may incorporate one or more subdivisions" (Library of Congress 2001). Examples of classes are the fundamental disciplines or what educators refer to as subjects that are the foundation of the main classification systems: mathematics, physical science, human science, history, art, and so on. Classes are usually divided and arranged according to principles of categorization, such as shared properties and exclusivity.

Relationships among concepts are specified to varying degrees by different types of classificatory structures. For example, thesauruses generally specify only three types of semantic relationships (equivalence, hierarchical, and associative) (NISO 1993). Library classifications are limited in how relationships can be constructed or how many can be specified by many factors, such as the type of scheme (faceted or enumerative, universal or special) and the hospitality of the inherent notation. In faceted thesauruses or faceted classifications, the relationships are structured with a central idea in mind—for example, object, process, or event. However, most library classifications attempt to preserve the principle of containing relationships. Containing relationships include:

1. Main class or basic subject in relation to all its subdivisions
2. Genus in relation to species
3. Whole in relation to part
4. Class in relation to its members (Langridge 1973).

The principle of containing relationships does not mean that each item must be more special than the one preceding it. Many items are neither more general nor more special than those adjacent. For example, a book on Mozart in general would precede one on Mozart's operas, etc. Such knowledge organization structures correlate well to topics in education, where concepts from different knowledge domains are related to one topic. These relationships are often based on contextual analysis, and the relatedness is based on proximity of concepts in the text. For instance, the concept "drainage basin" is highlighted as a heading in a textbook; other concepts like sheetflow, interfluves, gullies, and continental divides appear in text below this subheading. According to the inclusion relationships principle, drainage basin is a broad term, and sheetflow, interfluves, gullies, and continental divides are narrow terms.

Concept, subject, and facet analyses are the processes by which public knowledge structures are used and created. They are familiar activities to librarians, and distinctions between them are often not made. Concept analysis, usually done by indexers, uses an indexing language or thesaurus. Subject analysis as done in library cataloging is the process of assigning subjects (subjects are much broader than concepts or facets) from a controlled vocabulary list (or a thesaurus) to a document. Discourse communities interpret facet analysis in different ways. Classificationists, designers of classification schemes, perform facet analysis when they try to identify the fundamental classes needed or inherent in a subject. The definition of facet analysis we use in this study is based on the original work of Ranganthan. It is also currently used by the Facet Analysis Theory (*Facet Analytical Theory* 2001) project to create subject-based portals for the Web. Facet analysis is the "rigorous process of terminological analysis where the vocabulary of a given subject is organized into facets and arrays, resulting in a complex knowledge structure with both semantic and syntactic relationships clearly delineated" (Broughton 2002, 137).

These analysis techniques are used to solve the disambiguation problems of semantics, which are well-known problems in information retrieval. When a user searches using a word or phrase, do the records that are retrieved with the same words or phrase really correspond to what the user meant? Classificatory structures take care of semantic problems such as synonyms and homographs in many different ways—for example, thesauruses use qualifiers and parenthetical statements. They also specify relationships between terms. Semantics therefore refers to the meaning of the term, both its dictionary definition as well as all the associations to it. Definitions are called the "reference" or "denotation," and associations are called "connotation." Definitions are limited and often standardized by community consent and use, but connotations may be infinite since they are determined by personal experience. Thesauruses

select associations and include them in three kinds of semantic relationships; indexing languages try to describe many more associations. However, we find that in geography teaching many more associations (connotations) need to be specified and described for the novice learner. This is one important difference. Another difference is that definitions (denotation) must also be provided. Given these surface similarities, we decided to find out how, if, and what knowledge organization system could be used to facilitate physical geography science learning. Our specific questions: How can library classification schemes and thesaurus-type knowledge structures be used for educational purposes? How can differences between organization for learning and for information retrieval be reconciled?

There are a number of earth sciences thesauruses that include physical geography and these are described briefly in appendix 1.

To answer the question of whether a thesaurus can be used for educational purposes, a comparison of concepts in two types of information resources was carried out. Concepts from physical geography texts and the major thesaurus in the geosciences, *GeoRef*, were compared. We also examined the information system *GEOBASE*, a database that indexes materials in physical geography.

Physical geography is one of the subjects in geosciences; it may be considered a marginalized knowledge domain because it does not have a major classification scheme or a thesaurus devoted only to it. We found that documents on physical geography have just a linear list of terms rather than a thesaurus for collection indexing and retrieval. The *GeoRef* thesaurus does contain some terms used in physical geography, but many of them are not included. According to our preliminary estimates, 65% of concepts explained in the textbook cannot be found in *GeoRef*. Examples of concepts not found in *GeoRef* are angle of incidence, angle of repose, atomic number, atomic weight, autumnal equinox, available water, average global temperature, azimuth. A full list of concepts not in *GeoRef* is available on the ADEPT server at http://piru.alexandria.ucsb.edu/~buchel/concepts/p4.html.

To answer the question of how a library classification scheme can be used to support science learning, we examined two widely used classification schemes: Library of Congress Classification and the Dewey Decimal Classification. We found that many physical geography concepts are excluded from these major, albeit general, classification schemes. We speculate that this is so because concepts are neither subjects nor topics. Examples of geography concepts not found in these tools are open and closed systems, law of basin areas, leeward. However, these terms are needed if geographic information resources such as maps and datasets are to be more adequately described for information retrieval in a library catalog that supports science learning.

Some universal classification schemes are employed primarily outside the United States, but many of these are based on subjects and disciplines similar to LCSH and DDC, so geology concepts would be separated from geography, resulting in educational limitations.

Another disadvantage of current universal classification schemes is their use of the principle of containing relationships, not the principle of building relationships around the main idea—object, process, or event—as the guideline for knowledge organization. Thus, they do not represent the concept relationships that reflect the order of things in science.

Scientific classifications differ from library classifications and subject thesauruses in many ways. One significant difference is that library classifications are based on the literary warrant and link topics to subtopics. Many scientific classifications (chemistry, physics, biology, medicine) belong to disciplines that have widely accepted classifications and categorizations. Most of the concepts in these disciplines might naturally fit into thesaurus hierarchies based on the is-A relationship, often referred to as broad term (generic) - narrow term (generic) (BTG-NTG) relationship. The proliferation of special library classifications and thesauruses in many science disciplines shows that many more relationships and deeper subject intension is often needed than provided by general-purpose schemes. Physical geography, however, is a discipline that is somewhat unique. Geographers and geomorphologists do not have unanimous approaches to classifying real-world phenomena or processes; rather they have multiple classifications based on various criteria. All are considered equally important for teaching the science of geography.

Limitations of Existing Knowledge Structures

As early as 1944, Swank pleaded for a critical discussion that recognized the interrelationships between classification, library catalogs, indexes, and bibliographies (Swank 1944). While in the digital world these tools are certainly merging and can be merged, our analysis shows that other critical interrelationships that need to be considered for the development of digital learning spaces are the ones that integrate knowledge structures and reference sources. This means that we should explore the merging of knowledge structures such as classification schemes and thesauruses with reference works such as encyclopedias and dictionaries. Included in this list are glossaries, gazetteers, and terminology lists. Reference tools like Xrefer (www.xrefer.com) and Atomica (http://atomica.com) already support limited thesaurus+definition+encyclopedia linkages. However, we do not discuss the integration of these sources further in this paper. We focus only on the limitations of classificatory concepts such as hierarchy, semantic relationships, and order as they are currently implemented or used in knowledge structures. The enhancements needed are also discussed.

The thesaurus' hierarchical relationships—generic, instance, and partitive—are not enough to describe the full granularity of how phenomena or processes or objects are analyzed in physical geography. To illustrate this point, let us consider types of "atmosphere," specifically, its NTG relationships. Figure 1 shows them listed in alphabetical order.

All these concepts represent types of atmosphere. They are all linked to atmosphere with one type of relationship, NTG. However, for a specialist in the field, these concepts differ in their relationship to the concept atmosphere; namely, they are based on different classifications that consider different criteria. Following the faceted thesauruses' practices, such as those in the Art and Architecture Thesaurus (A&AT, available at www.getty.edu/research/tools/vocabulary/aat), these different classifications can be maintained in different "nodes," which simply mean different sets of subtypes. Nodes themselves are treated as nonindexing terms and are shown in angular brackets in figure 2.

By default, the conventional thesauruses and thesaurus construction software, such as Multi-Tes (available at www.multites.com/), arrange narrower concepts in alphabetical order. Such an arrangement is not a satisfactory

Atmosphere
　　NTG
　　　　heterosphere
　　　　homosphere
　　　　ionosphere
　　　　mesosphere
　　　　ozonosphere
　　　　stratosphere
　　　　thermosphere
　　　　troposphere

Figure 1. NTG Relationships for "Atmosphere"

Atmosphere
　　NTG
　　　　<composition criterion>
　　　　homosphere
　　　　heterosphere
　　　　mesosphere
　　　　stratosphere
　　　　thermosphere
　　　　troposphere
　　　　<function criterion>
　　　　ionosphere
　　　　ozonosphere

Figure 2. Nodes for "Atmosphere"

foundation for building learning spaces in physical geography. In real geography learning, concepts are arranged in a certain order, which is by no means alphabetical, and differs depending on the concept. It may be chronological or based on disciplinary logic. For example, geological time periods are arranged in a chronological order that is based on macro time scales. Other times, arrangements are based on micro time scales (seconds or minutes or even less). The need for ordering concepts exists in library classification schemes too (for example, the DDC has preference order, and Colon Classification recommends a citation order). However, general thesauruses usually do not specify a citation or preference order. The exception is faceted thesauruses, such as the A&AT. Our findings suggest that ordering of concepts in sets (or nodes) is desirable for enabling learning.

It would also be useful to add multiple ordering of concepts—for example, as the concepts are organized in the textbook. In the literature, concepts can be repeated as subtypes of one concept under different classifications. For instance,

1. N. Lancaster (1995) differentiates the following dune types:
   - Crescentic dunes (barchans and crescentic ridges)
   - Linear dunes
   - Star dunes
   - Parabolic dunes
   - Nebkhas
   - Lunettes
2. E. D. McKee (Christopherson 1999) differentiates:
   - Crescentic dunes (barchans, transverse, parabolic, barchanoid ridges)
   - Linear dunes (longitudinal, seif)
   - Star dunes
   - Other dunes (domes and reversing)
3. L. Aufrere (Christopherson 1999) classifies dune in the following way:
   - Longitudinal
   - Oblique
   - Transverse

In these classifications, linear, crescentic, and star dunes appear in two classifications. To incorporate all these classifications into a thesaurus, we would have to list some of the terms as subtypes of sand dunes twice or more. A sample concept map including all sand dune classifications is available at http://piru.alexandria.ucsb.edu/cmaps/dunes/Sand%20Dunes.html. The classifications are taken from Nicholas (1995).

Some universal classification schemes are employed primarily outside the United States, but many of these are based on subjects and disciplines, so geology concepts would be separated from geography, resulting in educational limitations.

Another peculiarity of concept arrangements in textbooks is that they are not in the alphabetical order usually found in thesauruses. Very often orderings carry additional semantic information—for instance, they show chronological sequence or evolution. Thesauruses should reflect such scientific orders. These orderings do not necessarily have to be displayed for the users; they can be used only by librarians or in the background. The users will just see the sequence of concepts in a way the phenomena or their types are arranged in nature, or in chronological order or the order in which scientists usually arrange them.

Ranganathan suggested the following possibilities for order in array within a facet in addition to alphabetical order (Langridge 1973):

- Increasing quantity. Types of polygons could be arranged in this order: triangle, quadrilateral, pentagon, hexagon, heptagon, octagon, nonagon, decagon, hendecagon, dodecagon, etc.
- Later in time. Writers in literature could be arranged according to their date of birth.
- Later in evolution. Living things could be arranged in this way.
- Spatial contiguity
- Increasing complexity. Methods, instruments, machinery could be arranged in this way.
- Canonical order. This means a traditional order, such as arithmetic, algebra, geometry.
- Favored category or literary warrant. This order could give precedence to the subjects in the array about which most had been published.

Subject intension is depth of the subject, the micro-topics. If relationships that support scientific theories, classifications and categorizations, and concepts on a level of even micro-topics are available, they can be used for the construction of concept maps. Concept maps are gaining quick popularity as a favored instructional material in many disciplines (Novak 2001). There are a number of software packages that allow students and instructors to construct concept maps. *IHMC software* is available at http://cmap.coginst.uwf.edu and *Inspiration* is available at www.inspiration.com. However, our experience at ADEPT indicates that many instructors do not have the time to build concept maps and organize their materials using them. Another reason why concepts maps are not widely constructed in science is because it is very difficult to build a concept map from scratch and show the complexity of relationships on one plane, as opposed to multidimensional space. However, instructors are often willing to use concept maps in instruction if they are constructed, maintained, and organized by other responsible entities, such as libraries.

Convergence

We conclude that a faceted thesaurus based on scientific classifications of disciplinary-specific facets such as objects, processes, phenomena, and methods can provide the foundation for developing digital learning spaces in physical geography. It must be constructed with great care given to the contents, categorization, and quality of the hierarchies. For example, polyhierarchies must specify roles and have more detailed associative relationships. An ALCTS committee has been investigating the area of subject relationships; see www.ala.org/alcts/organization/ccs/sac/rpt97rev.html for their report. Hierarchies must also link types and parts of objects to objects, processes to processes, etc. Besides overcoming as many of the limitations mentioned above as possible, other specific enhancements are also needed and are described below.

Facets: Facets are "clearly defined, mutually exclusive, and collectively exhaustive aspects, properties or characteristics of a class or specific subject" (Maple 1997). Our analysis derived similar facets in physical geography that were originally assembled as universals in the context of all disciplines (Dahlberg 1978):

- objects (for instance, landforms, hydrologic bodies, rocks, soils)
- properties, attributes
- processes and activities (fluvial processes, eolian processes, atmospheric processes)
- instruments
- theories, principles, classifications
- applications
- disciplines (geophysics, geostatistics, geochemistry, etc.)

Names for Relationships: Another desirable enhancement that should not be overlooked is more detailed specification of the associative relationships among concepts. Relational structures—indicated usually by an abbreviation RT for related term or by AS, often used for associative relationships—are neither sufficient nor explicit. While the task of decoding what is hidden behind the abbreviation would not seem to be complex for the more experienced users of a specific domain, for the novice domain users, in our case, students, it will be difficult to understand the nature of RT relationships without specification. Therefore, for students a short explanation of the associative relationships is necessary.

Labels for Nodes: Node labels are organizational devices that are often used to arrange hierarchical displays (Milstead 1998). An example of node labels for concepts of atmosphere types is shown in the next section. Node labels are enclosed within angled brackets.

Support for Orders in Arrays of Related Concepts: As discussed previously, concepts in the relationships can be ordered not only alphabetically, but also in chronological, evolutional, canonical, and other sequences. Agreement is needed for how objects should be ordered in geosciences. For instance, in our example with atmospheres, the flexibility of defining sequential orders lets us arrange concepts of atmospheric types in the natural order as is shown in figure 3.

Using Topic Maps for Digital Learning Spaces

Many technical and theoretical approaches offer solutions to interlink concepts; semantic nets, ontologies, topic maps, and concept maps are some of them. The use of the ISO Topic Map standard, XTM, as a tool for building and displaying digital learning spaces is briefly explored in the following paragraphs (XML 2000).

Topic maps provide "a standardized notation for interchangeably representing information about the structure of information resources used to define topics, and the relationships between topics" (XML 2000; ISO/IEC 1999). The topic definition in the standard is similar to the librarian's definitions of topics or subjects. However, it is possible to use topic maps for linking concepts and specifying relationships between them. In other words, it is possible to construct information resources such as concept maps based on classificatory knowledge structures, like the faceted thesaurus using the XTM standard. In doing so, no significant distinction between concepts and topics is made.

As examples of topic maps, XTM and faceted thesauruses have many features in common:

- Both link concepts.
- Both treat concepts and relationships separately.
- Both use controlled vocabularies.
- Most relationships in both standards have symmetrical counterparts.

```
Atmosphere
    NTG
            <composition criterion>
            homosphere
            heterosphere
            <temperature criterion>
            troposphere
            stratosphere
            mesosphere
            thermosphere
            <function criterion>
            ozonosphere
            ionosphere
```

Figure 3. Natural Order for "Atmosphere"

- Both serve as backbones in information systems that allow linking the information resources to the concepts; the only difference is that topic maps do this internally within the topic map, and thesauruses do this externally—surrogate bibliographic records carry linkages to the concepts in the thesaurus.
- In both standards hierarchies play an important role in information organization. The first step of organizing concepts into classes and subclasses of concepts in topic map construction is similar to the task of defining a genus and species, a primary task in building a thesaurus.
- Classes in topic maps and hierarchies in thesauruses facilitate a systematic approach to cross-referencing concepts from different facets; from these an ontologist will have a clear picture in terms of the categories for characterizing the concepts.

Standard thesaurus relationships can be mapped to topic map terminology as shown in figure 4.

The advantages of topic maps are that they are in XML format and have a number of software packages—for authoring, navigating, and displaying. These software packages help users and creators visualize the contents of the concept space and display explicitly concepts and relationships among concepts. The visualizations can be shown to users as concept maps. Concepts immediately related to a particular concept can be viewed as a concept map for the particular concept. The relationships in topic maps also have a much richer structure than the relationships in thesauruses. While thesauruses have only a predefined set of relationships, the XTM standard gives full semantic freedom in the specification of relationships. This is both an advantage and a disadvantage. Since concepts and relationships are important for librarians and educators and may vary based on discipline, different learning spaces can be designed for different domains.

Relationships in topic maps have two components: role-defining topic and role. Role-defining topic can be expressed as a verb; for example, written by, has, originates. The idea of including verbs in a concept space is interesting, since most library knowledge organization schemes have always tried to avoid verbs. In science education, verbs are important information elements and are used quite exten-

sively: modify, move/transport, measure, capture, produce, protect, damage, form, originate, re/distribute, dissipate, accumulate, gravitate, occupy, cover, resist, protect, dissolve, decrease, increase, exist, generate. For the role-defining topics we initially recommend implementing isAssociatedWith, Has-Constitute/Comprise, IsStudiedIn-Studies, Uses-IsUsedBy, IsCausedBy-Causes using XTM. However, a more complex typology of relationships can also be specified.

Roles refer to the related concept and explain its category. For roles, names of basic categories—or facets—such as, processes, methods, tools, properties, objects; or their more detailed subcategories: landforms, landscapes, hydrologic bodies, vegetation, fluvial processes, and other can be applied. Roles are the elements that can make the topic map model especially attractive to educators and novice learners. Figure 5 shows the list of relationships and roles needed for teaching and learning about atmosphere.

Due to the complexity of the relationship structure in the XTM standard, the relationships have to be specified in both directions: from concept a to concept b and from concept b to concept a. From this perspective, the thesaurus appears to be a more efficient model; the symmetrical relationships (in our case from b to a) are derived from the relationship between a and b. The weakness of both standards though is that they lack the predefined behavior of relationships. While in thesaurus the behaviors are described in the NISO standard, no action is taken upon it by the existing software packages. For instance, knowing classes of concepts and the directionality of relationships, one could impose the constraints that would disallow such entries as Hamlet wrote Shakespeare, or erosion is made of sand. The constraints that would disallow such entries are necessary if we are going to use the help of the scientific community in building the relationships. Additional research of relationship behavior in concept spaces is highly desirable. Details are available in Hill et al. (2002).

Thesaurus	Topic Map
BTG-NTG - General term	\<instance\>
USE - Use this term	\<baseNameString\>
UF - Use this term instead	\<variantName\>
SN - Scope note	\<scope\>
NTG-BTG - Narrower term	\<instanceOf\>

Figure 4. Thesaurus—Topic Map Relationships Mapping

IsStateOf
IsMethodFor
IsClassificationFor
HasClassification
HasAssociatedMethods
HasStates
CanBeMeasuredWith
IsForMeasuring
HasProcesses
IsProcessOf
HasProperties
IsPropertyFor
IsAssociatedWithGeosystem
IsUsedFor

Figure 5. Typology of Associative (Related) Relationships for Atmosphere

The basic idea of topic maps is that all concepts are grouped around one central concept—topic via the inclusive relationship—*instanceOf*—instance. Each relationship can have a reference to a specific role-defining topic. The examples used in the XTM standard indicate that the topics for the relationships can be expressed as verbs. For instance, in creating a topic map about streams we could say: topic streams includes instance—*has*. Further, all related-to-stream concepts that can be linked with the verb has can perform specific roles in the relationship—in this specific case, all the components will have a role—*Part*. Different types of stream parts can be disambiguated by the more detailed specification of roles: *LongitudinalPart*, *CrossSectionalPart*, *Component*. An experimental visualization of the described set of relationships is shown in figure 6.

While the topic map standard offers a nice data model, it does not describe the theoretical details of construction for concept spaces: what relationships can be used, or how the relationships should be specified. Such principles must be developed. We believe that the underlying principles that are well suited for topic map construction are the principles of faceted thesauruses. These include strictly defined rules about specification of hierarchical relationships, use of controlled vocabularies, and construction of associative relationships. As a general organizational principle, we also propose using the "natural order of things" approach, described earlier. This means that we will structure relationships according to the order of things as perceived in science/nature. Streams will be related to streams, parts of streams, stream properties, and other objects related to the streams: e.g., terraces, fluvial processes, and physical, chemical, and biological properties. When this approach is followed, a subject ontogeny is maintained as well and becomes a part of public knowledge structures.

Topic maps can provide the conceptual framework for developing the content domain and scientific reasoning processes of student learning that can be associated with information resources. Because the standard builds a struc-

tured semantic network over heterogeneous and topically diverse resources, it allows easy and selective navigation to the requested information as well as easy maintenance. The interface for topic maps can be either conventional with hyperlinks, or graphical, the so-called hyperbolic browser interface that represents concepts as the nodes in hierarchies. Hyperbolic navigation where the user can rearrange nodes and bring areas into focus with the mouse has been shown to be a better interface for understanding of complex, heterogeneous data sets.

Hyperbolic visualizations with Empolis or Ontopia allow users to enter a search term/concept phrase (Empolis is available at www.empolis.com and Ontopia at www.ontopia.net). The system responds with a concept map for the term/phrase. Each node of this map has other concepts associated with it with specific roles such as *has-parts, isCausedBy-Process*, and so on. Each node can be activated as a live hyperlink to provide entry into other thesaurus maps that extend or narrow the relationships and concepts as appropriate. Each node also has occurrence roles, the information resources associated with each concept.

The use of topic maps for educational purposes may need certain improvements to visualizations, such as:

1. support for multiple ordering
2. display of hierarchical relationships
3. more interactive features that allow a user to select and save existing learning spaces. A click on a concept node should display the information about a concept—its scope and variant notes.
4. aggregation of resources by form/type (including micro-types such as charts, images, etc.) and number of resources (e.g., the number of models available for this concept)
5. visualization limited to one concept with immediate relationships, not the whole body of concepts

Conclusion

Summarizing, we find that classificatory notions, such as hierarchy, concepts, classes, facets, and vocabulary can be used to provide the digital learning spaces needed to support science education in physical geography. To do so, the critical places where the enhancements need to take place in a faceted thesaurus are described: terms to be more rigorously defined and maintained as concepts and facets, scope notes to include nodes that make the relationships explicit, semantic relationships to be extended with other disciplinary classifications and associations, and display of concepts and relationships in multiple orders. Other enhancements needed, such as combining definitions and illustrations with thesaurus terms and better understanding
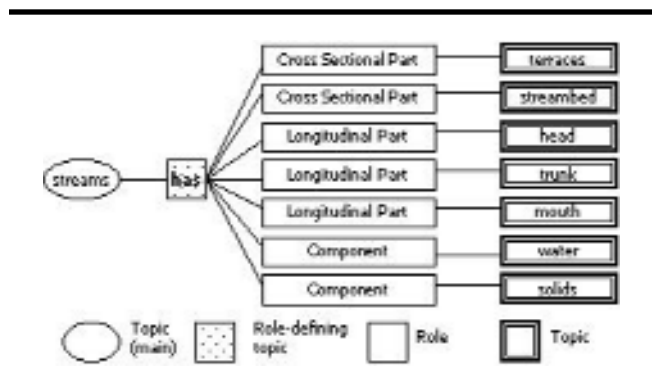

Figure 6. Visualization of Relationships for Stream

of the behavior of relationships in concept spaces, have been identified and further investigation is recommended. Development alternatives include concept maps using concept mapping software or ISO topic map authoring. These approaches, specifically XTM, have been compared with existing principles and protocols for thesaurus construction and maintenance, and the advantages and disadvantages have been highlighted.

Works Cited

Barrow, Robin and G. Milburn. 1990. *A critical dictionary of educational concepts: An appraisal of selected ideas and issues in educational theory and practice.* New York: Teachers College Pr., Columbia University.

Blake, David, and Vincent Hanley. 1995. *The dictionary of educational terms.* Aldershot, Hants, England; Brookfield, Vt.: Arena.

Broughton, V. 2002. Facet analytical theory as a basis for a knowledge organization tool in a subject portal. Paper presented at the Seventh International ISKO Conference, Challenges in knowledge representation and organization for the twenty-first century: Integration of knowledge across boundaries, Granada, Spain, July 10–13.

Chistopherson, Robert W. 2000. *Geosystems: An introduction to physical geography*, 4th ed. Upper Saddle River, N.J.: Prentice-Hall.

Coleman, A., et al. 2001. Learning spaces in digital libraries. Paper presented at the fifth European Conference ECDL 2001, Sept. Darmstadt, Germany. Also published as *Lecture notes in computer science* 2163. Berlin: Springer-Verlag.

Dahlberg, Ingetraut. 1978. *Ontical structures and universal classification.* 1st ed., Sarada Ranganathan Endowment for Library Science series. Bangalore, India: Sarada Ranganathan Endowment for Library Science.

Deighton, L. C., ed. 1971. *The encyclopedia of education.* New York: Macmillan.

Dykstra, Mary. 1988. LC subject headings disguised as a thesaurus. *Library Journal* 113, no. 4: 42–46.

Extensible Markup Language (XML) 1.0, Second Edition. Accessed Mar. 15, 2001, www.w3.org/TR/REC-xml.

Facet analytical theory in managing knowledge structure in the humanities 2001. Accessed July 1, 2002, www.ucl.ac.uk/fatks.

Greenblatt, Ellen. 1995. *Authority control: A basic glossary of terms,* Dec. 15. Accessed July 17, 2002, http://ublib.buffalo.edu/libraries/units/cts/ac/def.html.

Hawes, Gene R., and L. S. Hawes. 1982. *The concise dictionary of education.* New York: Van Nostrand Reinhold.

Hill, Linda, et al. 2002. Integration of knowledge organization systems into digital library architectures: Position paper for the 13th ASIS&T SIG/CR Workshop, Reconceptualizing classification research. *Proceedings of the 13th ASIS&T SIG/CR Workshop Classification Research*, Nov. 17.

Husen, T., and Torsten Neville Postlethwaite, eds. 1994. *The international encyclopedia of education,* 2d ed. New York: Elsevier Science.

ISO/IEC 13250 topic maps: Information technology—Document description and markup languages. 1999. Accessed October 20, 2001, www.y12.doe.gov/sgml/sc34/document/0129.pdf.

Lancaster, Nicholas. 1995. *Geomorphology of desert dunes.* New York: Routledge.

Langridge, Derek. 1973. *Approach to classification.* North Haven, Conn.: Linnet Bks.

Laurence, Stephen, and Eric Margolis. 1999. Concepts and cognitive science. In *Concepts: Core readings*, eds. Eric Margolis and Stephen Laurence, 1–82. Cambridge, Mass.: MIT Pr.

Leonard, Brian M. 1997. *Introducing the measurement of shape in freshman human geography.* Accessed July 17, 2002, http://ericae.net/ericdb/ED405192.htm.

Library of Congress. 2002. Subject cataloging manual: Subject headings, 5th edition with updates. In The Cataloger's Desktop, Issue 4 [CD-ROM]. Washington D.C.: Library of Congress.

Maple, A. 1997. Faceted access: A review of the literature. Accessed July 1, 2002, www-sul.stanford.edu/depts/music/mlatest/BCC/BCC-Historical/Bcc95/95WGFAM2.html.

Mayer, R. E. 1989. Models for understanding. *Review of Educational Research* 59, no. 1: 43–64.

McCormick, C. B., and M. Pressley. 1997. *Educational psychology: Learning, instruction, assessment.* New York: Longman.

Members of the TopicMaps.Org Authoring Group. 2000. *XML topic maps (XTM) processing model 1.0.* Accessed October 20, 2001, www.topicmaps.org/xtm/1.0/xtmp1.html.

Merrill, M. D. 1983. Component display theory. In *Instructional design theories and models*, ed. C. Reigeluth. Hillsdale, N.J.: Erlbaum.

Miller, G. A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63: 81–97.

Milstead, Jessica. 1998. *NISO Z39.19: Standard for Structure and Organization of Information Retrieval Thesauruses.* June 29, 2002, www.jelem.com/z39.htm.

NARA. 2001. *Digital classroom: Teaching with documents: Document analysis worksheets.* Accessed Mar. 10, 2001, www.archives.gov/digital_classroom/lessons/analysis_worksheets/worksheets.html.

NISO. 1993. *Guidelines for the construction, format, and management of monolingual thesauri.* Bethesda, Md.: NISO Pr.

Novak, Joseph D. 2001. *The theory underlying concept maps and how to construct them.* Accessed July 17, 2002, http://cmap.coginst.uwf.edu/info.

Rumelhart, D., and Norman, D. 1981. Analogical processes in learning. In *Cognitive skills and their acquisition*, ed. J. R. Anderson. Hillsdale, N.J.: Erlbaum.

Smith, T. R., et. al. 2002. Structured models of scientific concepts for organizing, accessing, and using learning materials. In *Challenges in knowledge representation and organization for the 21st century: Integration of knowledge across boundaries.* Proceedings of the Seventh International ISKO Conference, Granada, Spain, July 10–13, ed. Maria J. Lopez-Huertas, 232–39. Berlin: Ergon.

Swank, R. 1944. Subject catalogs, classifications, or bibliographies? *Library Quarterly* 14, no. 4: 316–32.

Taylor, Arlene G. 1999. *The organization of information.* Englewood, Colo.: Libraries Unlimited.

Ziman, J. 1984. *An introduction to science studies: The philosophic and social aspects of technology.* Cambridge, Mass.: Cambridge Univ. Pr.

Appendix
Thesauruses for Physical Geography

GeoRef—The thesaurus contains more than 27,000 terms, with several standard symbolic relationships, such as broader term, narrower term, related term, and use for. It also includes usage notes, dates of addition, and coordinates for selected place names. The GeoRef thesaurus is primarily used for indexing documents of the GeoRef database, the premier database from the American Geological Institute. It is the most comprehensive database in the geosciences and continues to grow by more than 80,000 references a year. The GeoRef database covers the geology of North America from 1785 to the present and the geology of the rest of the world from 1933 to the present. GeoRef is available at www.silverplatter.com/catalog/gref.htm.

GEMET—The General Multilingual Environmental Thesaurus (GEMET) has been created by merging different national and international thesauruses. The present Version 2000 of GEMET presents 5,298 descriptors, including 109 top terms, and 1,264 synonyms in English. The 5,524 terms belonging to the parental thesauruses and not included in GEMET constitute an accessory alphabetical list of free terms. GEMET provides a complete numerical equivalence (all the descriptors have an equivalent) with the included languages. GEMET is available at www.mu.niedersachsen.de/cds/Guided-Tour.htm.

Feature Type Thesaurus—A set of terms for categories of geographic places; these are terms to indicate the nature of a place. It has been designed for use with the Alexandria Digital Library Gazetteer. Feature Type Thesaurus is available at www.alexandria.ecsb.edu/gazeteer/FeatureTypes.

CIESIN Indexing Vocabulary was developed to index data resources and data sets related to human interactions in global change. Metadata records containing CIESIN indexing terms appear in the CIESIN Gateway, the Global Change Master Directory, and the Earth Observing System Data and Information System Information Management System. The CIESIN indexing vocabulary is available at www.ciesin.org/metadata/documentation/vocab/index.html.

In **NASA's Global Change Master Directory** terms are grouped into the following categories: data center, instrument, location, platform, and project. Particularly interesting to educators would be the instrument and location facets. NASA's Global Change Master Directory is available at http://gcmd.gsfc.nasa.gov/Aboutus/sitemap.html.

USGS Thesaurus is currently under development. It includes a high-level set of categories that will interface easily with the category structures currently in use within the USGS. Methods and sciences facets from this may be especially useful. The current version of the USGS thesaurus is available at http://alexandria.sdc.ucsb.edu/~lhill/usgs_terms/usgs/USGSMainPg.htm.

Other possible classifications, glossaries, and thesauruses can be culled from within physical geography, earth systems sciences, and other related disciplines like engineering. Some examples are:

- Glossary of Physical Oceanography and Related Disciplines (http://stommel.tamu.edu/~baum/paleo/ocean/ocean.html)
- NASA Thesaurus (www.sti.nasa.gov/thesfrm1.htm)
- Canadian Thesaurus of Construction Science and Technology (www.nrc.ca/irc/thesaurus/ctcst-search-form.html)
- EI Thesaurus (www.ei.org/eicorp/eicorp?menu=eithesaurusmenu&display=eithesaurus)
- INSPEC Thesaurus (www.iee.org/publish/support/inspec/document/thes)