Environment and Planning B: Planning and Design 2008, volume 35, pages 169 – 186

DOI:10.1068/b3344

In search of classification that supports the dynamics of science: the FAO Land Cover Classification System and proposed modifications

Ola Ahlqvist
Department of Geography, The Ohio State University, 154 N Oval Mall, Columbus, OH 43210, USA; e-mail: ahlqvist.1@osu.edu
Received 5 May 2006; in revised form 23 August 2006; published online 1 October 2007

Abstract. Classification of geographic phenomena is often a black box to anyone outside the immediate group involved in the classification process. There is a growing need for compatibility between datasets to map, evaluate, and monitor areas in a consistent manner. The FAO (Food and Agriculture Organization of the United Nations) Land Cover Classification system (LCCS) is a proposed method to enable interoperability for land-cover data and an attempt to open the classification black box for scrutiny. The FAO LCCS is used to demonstrate some of the strengths and weaknesses of feature-based classification methods and how some important improvements, based on theoretical developments in geographic information science, can extend LCCS to become a 'boundary object' that supports representation, negotiation, and analysis of dynamic and heterogeneous classification systems. The suggested improvements also include an outline of how future classification activities could be developed into a distributed web-based ontology infrastructure.

Introduction

There is a growing need for detailed and accurate information on land cover and land-cover change at all geographic scales. It is, however, problematic to establish a common language between different maps, datasets, and their thematic legends. This hinders useful application, in particular for comparative analysis, and it also complicates data validation (Herold and Schmullius, 2004). There are some proposed solutions to address this problem, based on parameterized approaches to soil classification and land-cover classification from national and international bodies (Di Gregorio, 2005; FAO, 1998). That work and current efforts to promote this type of classification system to an international standard serve as the motivation for this paper to discuss parameterized approaches to classification and interoperability.

The aim is to put recent theoretical developments from work on cognitive and practical aspects of formal geographic ontologies into a workable solution for current global efforts to create useful and comprehensive information about the environment. In the first two sections I outline how development of land-cover classification is situated in a continuum from mental categorization, through group negotiations and standardization, to end-user interpretation. Land-cover classification activities and work on homogenization of existing terminologies are also presented. The FAO (Food Agriculture Organization of the United Nations) Land Cover Classification System (LCCS) is then introduced as an example of a practical classification exercise that currently brings the land-cover community together around a 'boundary object' to negotiate common understanding, with the aim of producing global datasets capable of being reconciled at different scales and places. The last section picks up identified shortcomings in the current LCCS and other (semi)formal, first-order logic ontologies for geographic entities. Based on a recently developed method for concept representation (Ahlqvist, 2004) I suggest some improvements that create richer and more precise descriptions of land-cover categories.

Categorization, classification, and standardization

The oldest and, according to many studies, the most effective method to communicate knowledge is through personal contact and conversation (Vickery, 1975). In the early stages of a science, terms are typically developed from common language, thereby creating a casual and sometimes intuitive terminology. Language and conservation is a medium where meaning and shared understanding emerges from common use and collective interaction around singular language elements, or terms, each of which maintains an individual meaning (Gärdenfors, 2000; Shrage, 1995). A shared terminology as a means of knowledge communication is, therefore, an important part of most sciences and dates back to early civilizations when craftsmen, architects, physicians, and scribes needed written accounts to support development of, for example, irrigation, metal mining, buildings, arithmetic, geometry, and medicine. However, as a science develops and matures, the need for effective communication of knowledge makes increasingly codified and systematically organized reporting necessary. Historically we witnessed this during the scientific revolution when, as a response to rather chaotic nomenclatures, scientists such as Carl Linneaus and Antoine Lavoisier started to develop a common language for their respective sciences (Vickery, 1975). In history we find ample evidence that hierarchical thinking, classification, and standardization is not something that came as a response to the digital age where information needs to be organized for computer handling. It may just be the other way around—that the design of digital solutions was driven by the highly hierarchical and standardized thinking of contemporary science.

This short historiography illustrates a flow from everyday categories and words through classification and standardization. I have no intention here to go deeper into the philosophical or psychological sciences and theories of how people categorize and form knowledge, but instead I follow one of Pawlak's (1991) propositions: that knowledge is deeply seated in the abilities of humans to classify anything—(apparently) real things, states, processes, moments of time, and all other more or less abstract concepts we can think of. A useful summary of cognitive theories related to classification from an information theoretic approach can be found in Bjelland (2004), who describes categorization as a mental cognitive process to construct order from day-to-day impressions. Since individual impressions will be highly experiential and subjective, a mental *category* or *concept* is likely to be subjective and dynamic in response to further experience. Mental categories have to be concretized in order to be talked about, to be negotiated, and to create shared meaning. It is "the vagueness, instability, and subjectivity of mental concepts that cognitive theories of categorization attempt to explain, and classification attempts to overcome" (Bjelland, 2004, page 26). Thus, the categorization process can be regarded as an individual, mental-level precursor to a social process of classification at a group level. The *classification process* then aims at structuring a specific knowledge domain in order to create consistency and stability in communication between individuals. A *class* (or *term*) is the result of this classification process and serves as a tangible vehicle for communication of meaning. In the case of spatial information the mental category or concept refers directly to a *feature* in the real world whereas the class or term refers to the real-world feature indirectly through the associated mental construct. So, even if subjectivity is reduced through a classification process, it gets reintroduced by the interpreter in the use of a class or term. This triadic semiotic relationship (see figure 1) between the category or concept as a mental construct, the class or term as a group expression of multiple individual meanings, and the real world that is referred to has been elaborated upon extensively (for example, MacEachren, 1995; Mennis, 2003; Peuquet, 1994; Salgé, 1995).

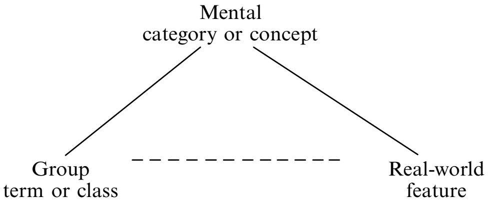


Figure 1. The triadic semiotic relationship between mental understanding and group understanding of a real-world feature. A direct relation exists as an association between category and feature as well as between category and term, but only an indirect relation exists between group term and feature through the individual, mental understanding. Adapted from Bjelland (2004).

If classification is a way to create order and stability for the communication of knowledge, it is important to emphasize that classification is still a dynamic, ordered, and sometimes only semiordered structure inundated with ambiguity and vagueness. Operationally, though, classification often makes an unproblematic leap from concept to class, eliminating any traces of concept ambiguity by stating mutually exclusive and crisp classes. In many ways classification and standardization are separate sides of the same coin (Bowker and Star, 1999) in that standardization takes classification one step further and fixes a classification system. Science has a lot to gain from developing accepted standards to ensure repeatable experiments, exchange of findings, and so forth, but the downside is that a further entrenchment of a classification system runs the risk of that system becoming stale and out of phase with contemporary thinking. A somewhat *oligarchic* power structure emerges when (groups of) experts develop and write dictionaries, encyclopedias, handbooks, ontologies, or standards all dictating the proper meaning of a language. We need to recognize that no classification system can reflect accurately either the social or the natural world (Bowker and Star, 1999). There are always multiple ways to conceptualize and communicate knowledge, so that there are inevitably many-to-many relations between categories and thus inherent ambiguity in any categorization. Above, I define the classification process to be about reconciling differences in the perspectives of different organizations, professions, nationalities, and so on for the communication of knowledge. This does not necessarily mean a formal process. On the contrary, a 'common use' view of classification can generate dynamic communities of practice (Bowker and Star, 1999) across disciplines. Gärdenfors (2000) argues that this, in a sense, creates a more democratic power structure, analogous to prices in a free market, where a single term cannot dictate meaning but instead meaning emerges from the social process and interaction between instances of individual meaning. Thus we see that classifications arise out of social communication needs but they serve purposes; not only do they reflect the ideas of a certain community or institution but they can also be the end result of negotiating and reconciling individual, group, and institutional differences. So, there is the necessity to define classifications and standards for making collective progress, and there is a problematic inertia that these impose on the necessary dynamics of science and practice. We would obviously like to take the best out of these two worlds—a structure that supports the needs of separate communities of practice and that is able to link across borders while maintaining an agreed identity. It has been argued that classification may work as a *boundary object* (Star, 1989) in situations such as these to mediate and support negotiations around which similarities and differences can be articulated (Harvey, 1999; Harvey and Chrisman, 1998; King and Star, 1990).

The boundary-object interpretation of classification means that classes can be customized to user requirements but also have common identities across users. This is achieved by allowing the boundary objects to be weakly structured in common use, and by imposing stronger structures in the tailored individual-user situation. In this way a boundary object is both ambiguous and constant. The following two design criteria from Bowker and Star (1999) illustrate some key desired characteristics of a classification process that supports the dynamics of science.

(a) *To recognize the balancing act of classification.* Classification schemes always represent multiple constituencies. They can do so most effectively through the incorporation of ambiguity—leaving certain terms open for multiple definitions across different social worlds. They are in this sense boundary objects.

(b) *To render the voice retrievable.* As classification systems get ever more deeply embedded into working infrastructures they risk getting black boxed and thence made both potent and invisible. By keeping the voices of classifiers and their constituents present, the system can retain maximum political flexibility. This includes the key ability to be able to change with changing natural, organizational, and political imperatives.

These suggestions recognize that we need to account for and represent the ontological diversity in different conceptualizations. There are also complex historical, institutional, and political issues incorporated into any conceptualization (Bowker, 2000). Although these dimensions are important for a full understanding of a given conceptualization they are not addressed in any depth here. The two points above are used loosely in order to analyze and propose some modifications to existing techniques for representing declarative knowledge. In the following outline of how classification is situated in the flow from geographic category to spatial data we will see how these criteria are essential to the use of a classification system. I have outlined above how classification follows the mental process of categorization, and that it is, in one way or another, a social process between stakeholders, where concepts are reconciled and specified in a shared terminology sometimes called ‘controlled vocabulary’ or ‘universe of discourse’ (see figure 2).

Following this we usually find stages of individual-level processes where users of a classification apply the shared terminology to identify and verify the identity of real-world objects, often as part of a data-production activity. That process completes the triadic semiotic relationship between category, term, and feature outlined above. This is a stage where subjectivity gets reintroduced and can get embedded in a dataset. The next stage, when other people use the data, includes a subjective interpretation process. At all stages the design criteria listed above are key to making correct identification and interpretation of objects and terms. From the data user’s point of view it is not only important to be able to look into the intended meaning of a term from the specification listed in a terminology description, but also into how the data producer interpreted and applied his or her understanding of the terminology. This is the second criteria of Bowker and Star (1999), to render the voice retrievable. Obviously it is not enough to supply a term together with a short written description, but rather to declare as much as possible about what goes into separating one term from another, saturating the specification with descriptive characteristics that can form a rich formal narrative.

This background will now serve as a foundation for an examination of land-cover classification in general, the existing FAO LCCS in particular, and finally an elaboration on how an LCCS-style classification can be modified to achieve ‘living’ classifications that support the dynamics of science.

Land-use and land-cover categories, classification, and standardization

Modern approaches to spatial land-cover classification can be traced back to the century-long tradition of vegetation classification, early work with terrain classification systems in the 1950s and 1960s, and land-use mapping in the 1960s and 1970s

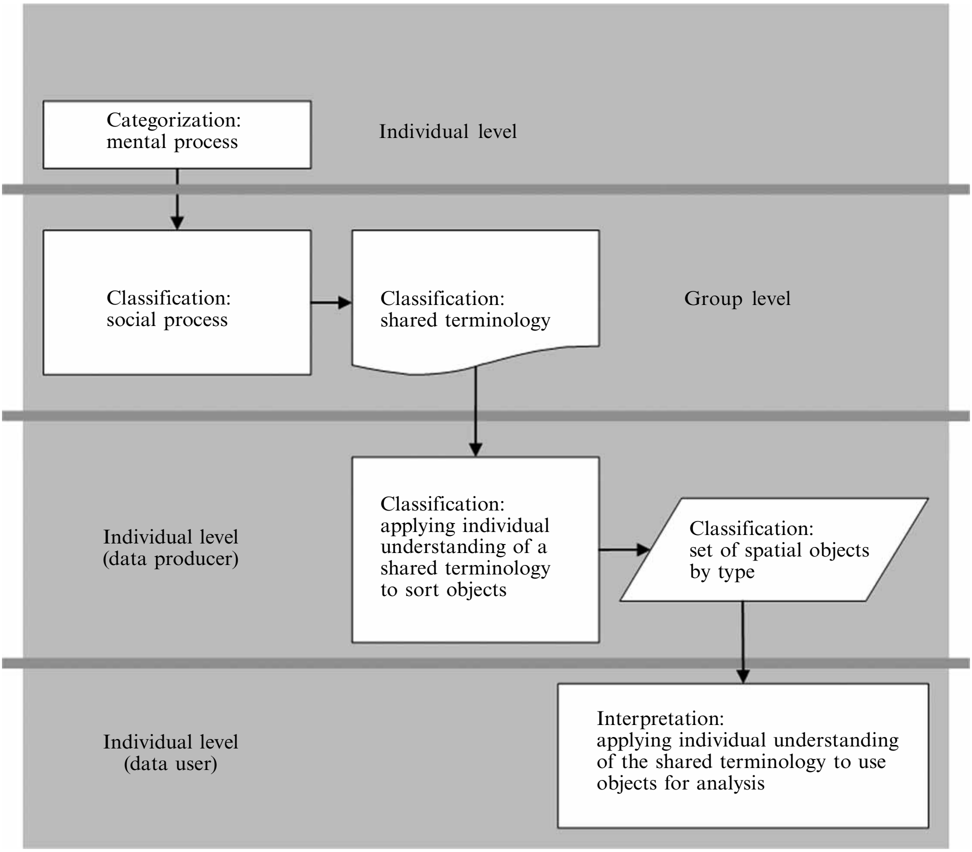


Figure 2. Outline that situates classification and the different meanings of classification in the overall process from mental understanding to the use of a term as an information carrier in geographic information such as land-cover data. Several feedback loops are present but, for clarity, these are not illustrated.

(see Anderson et al, 1976; Beckett et al, 1972; Jennings et al, 2004). In the latter half of the 20th century, researchers, public agencies, and private organizations recognized that accurate, meaningful, and current data on land use were essential for the escalating need to monitor changes in the environment (Clawson and Stewart, 1965). During the early 1970s the term 'land cover' started to intermix with the previously dominant term 'land use', and there is still an unfortunate mix of land use and land cover in terminology, taxonomy, and data sources. Clawson and Stewart (1965, page 14) defined land use as "man's activities on, under, or over the land or, inclusively, activities making use of land". Land cover on the other hand was defined (Burley, 1961 in Anderson et al, 1976, page 7) as "the vegetational and artificial constructions covering the land surface". In more up-to-date definitions (Di Gregorio, 2005) land cover is the observed (bio)physical cover on the earth's surface, and land use is the arrangements, activities, and inputs people undertake in a certain land-cover type to produce, change, or maintain it. Irrespective of the way we happen to define these two facets of a geographic landscape the data collected on land use and land cover were, and still are, mostly in the form of spatially referenced areal units with nominal terms that rely on the basic process of classification as described above.

Comber et al (2005) pointed to the many problematic semantic issues with land-cover information, mainly from a perspective of producing and using satellite-based data. Their analysis also highlighted the social constructivist character of land-cover information and the associated problem of reconciling different classification systems indicated in the previous section. This problem has been recognized for some time, and one of the recurring themes in land-use and land-cover monitoring initiatives is the effort to harmonize classification systems for landscape analysis. For example, the first expert meeting on harmonizing land-cover and land-use maps on a global scale was hosted by the United Nations Environment Programme (UNEP) and the FAO in 1993. This was in direct response to a growing need for standardization and compatibility between datasets to map, evaluate, and monitor wide areas in a consistent manner (Di Gregorio, 1991; FAO, 1995; Thompson, 1996). Some other examples of standardized nomenclatures and data-gathering methods include the National Vegetation Classification Standard (Vegetation Subcommittee, 1997), the Nordic Landscape Monitoring (Groom, 2005), the CORINE Land Cover (CEC, 1995), GLC2000 (Bartholomé and Belward, 2005), and AFRICOVER (Kalensky, 1998) to mention just a few.

Development of digital datasets also triggered other types of standardization work such as the Open Geospatial Consortium and the International Standards Organization Technical Committee 211 to enable interoperability between providers and users of geographic data. Those standards were concerned mainly with specifying syntactic and schematic aspects of interoperability, such as developing common exchange formats or standardization of projection, spatial reference systems, and measurement units. Nevertheless, Comber et al (2005) argue that, despite these efforts to create working spatial-data infrastructures, the semantic problem is still largely unresolved. In most initiatives the semantic aspect of data attributes is still left as unformatted text descriptions.

The semantic problems of data exchange have become apparent to the data-modeling community as well (Bishr, 1998) and early efforts based on *ontology* (Gruber, 1993; Guarino, 1995) have been suggested for solving discrepancies in conceptualizations. Ontology is originally the branch of metaphysics that deal with the nature of being. The term has, during the last ten years or so, been used in the geographic information science literature, where its meaning ranges from the metaphysical science of being, to the more computer scientific view that ontology is a formal specification of a common terminology in which shared knowledge can be represented. Several approaches to constructing ontologies for knowledge representation are possible. Perhaps the best well-known example of representing a terminological system in a formal way is WordNet (http://wordnet.princeton.edu), where terms can be partially ordered based on linguistic relations such as synonym, antonym, or hyponym (Miller, 1995). A different well-known formal description of terminology is Cyc (http://www.cyc.com), in which terms are attached with axioms such as facts, rules of thumb, and other assertions (Lenat, 1995). Another related activity to address semantic issues is the Semantic Web (http://www.w3.org), which seeks to develop languages and syntax for expressing information in a machine-processable form. One of the main benefits of creating these formal specifications is that they can help with information integration, taking data from different places and points in time and putting them together in a comprehensive information base (Fonseca et al, 2003). It should be noted that ontology development and integration can be seen as a formal parallel to the social-classification process described previously. We can make a rough distinction between two types of approaches to the use of ontology for data-integration purposes (Lutz and Klien, 2006). *Single ontology approaches* require that all terms reside in the same standardized taxonomic tree or concept network and use some measure of distance within that structure to estimate term similarity. *Hybrid ontology approaches* on the other hand can make comparisons across separate systems on the basis of a standardized set of descriptive primitives. Single ontology approaches focus on the standardization of terms and are readily compatible with the outcome from a classification process. Hybrid ontology approaches focus on the standardization of descriptive elements and can thereby enable a description of the classification process as it relates to how different conceptualizations compare with each other.

The application of ontology to land-cover classification work is related mainly to single ontologies as an explicit and formal specification of the outcome of a social classification process—that is, as a shared terminology structured according to a specified syntax (Gruber, 1993). Of specific interest to work that seeks to harmonize and compare heterogeneous conceptualizations, such as the land-cover terminologies mentioned above, is the capability to measure and compare similarity between categories in different classification systems. Such similarity measurements serve as a general method to establish semantic interoperability of information services (Schwering and Raubal, 2005). Semantic similarity assessment can be made in a number of ways, depending on the chosen representational format. A common technique is to follow a feature-based classification approach in which similarity is evaluated by comparing a list of discrete descriptive terms for each land-cover type. This corresponds roughly to a hybrid ontology approach as defined above. Based on the number of common terms the land-cover similarity is evaluated as a ratio of common attributes over total number of attributes. For some other examples see Hahn and Chater (1997) and Jones et al (2003). Evaluations of semantic similarity between terms have been used to compare classes as part of a classification and interpretation process (Feng and Flewelling, 2003; Kavouras and Kokla, 2002; Rodriguez and Egenhofer, 2004), and also to make geographic analysis using heterogeneous data (see Ahlqvist, 2005b; Fritz and See, 2005). Other issues specific to land-use and land-cover data have been addressed through the use of semantic similarity—for example, how incompatible classes raise significant problems for landscape-change analysis (Comber et al, 2004), and in accuracy assessment of land-cover maps (Fritz and See, 2005). Several of these studies have argued for similarity evaluations based on methods other than a straightforward feature-matching process.

The FAO LCCS

One of the standardization efforts mentioned previously was the UNEP and FAO initiated efforts to harmonize and standardize data needed for the implementation of UNCED's Agenda 21 (http://www.un.org/esa/sustdev/agenda21.htm). FAO had previously completed the International Reference Base for Soil Classification initiative, (now replaced by the World Reference Base for Soil Resources), which established a framework through which existing and ongoing soil classification work could be harmonized (FAO, 1998). The process that eventually became the LCCS initiative followed a similar path to reach an international agreement on the major land-cover types to be recognized at a global scale, as well as on the criteria and methods to provide a common scientific language for communication.

The current version of the FAO LCCS (Di Gregorio, 2005), is intended to meet specific user requirements for any land-cover classification initiative anywhere in the world, and it is designed to support mapping exercises independent of the scale used. The strategy chosen to meet those requirements is to develop a set of standard diagnostic criteria that describe local categories in such detail that they allow for comparison with other existing classifications. Hence, land-cover classes are defined by a set of feature descriptions terms *classifiers*. This follows the idea of a hybrid

Table 1. The original Anderson et al (1976) land-use – cover class 'industrial' definition and its description by the Land Cover Classification System (LCCS) classifiers, codes, and label (after Herold and Schmullius, 2004).

Anderson class

13 Industrial. Industrial areas include a wide array of land uses from light manufacturing to heavy manufacturing plants. Identification of light industries—those focused on design, assembly, finishing, processing, and packaging of products—can often be based on the type of building, parking, and shipping arrangements. Light industrial areas may be, but are not necessarily, directly in contact with urban areas; many are now found at airports or in relatively open country. Heavy industries use raw materials such as iron ore, timber, or coal. Included are steel mills, pulp and lumber mills, electric-power generating stations, oil refineries and tank farms, chemical plants, and brickmaking plants. Stockpiles of raw materials and waste-product disposal areas are usually visible, along with transportation facilities capable of handling heavy materials.

Surface structures associated with mining operations are included in this category. Surface structures and equipment may range from a minimum of a loading device and trucks to extended areas with access roads, processing facilities, stockpiles, storage sheds, and numerous vehicles. Spoil material and slag heaps are usually found within a short trucking distance of the major mine areas and may be the key indicator of underground mining operations. Uniform identification of all these diverse extractive uses is extremely difficult from remote sensor data alone. Areas of future reserves are included in the appropriate present-use category, such as agricultural land or forest land, regardless of the expected future use.

LCCS classifiers	**LCCS string**	**LCCS label**	**LCCS code**
B15 Artificial surfaces and associated area(s)	B15 – A1 – A4 – A12	Industrial and/or other area(s)	5003-8
A1 Built-up			
A4 Nonlinear			
A12 Industrial and other area(s)			

ontology approach with standardized descriptors allowing for heterogeneous user conceptualizations. The application of this classification system is divided into an initial hierarchical dichotomous phase, where eight major land-cover types are distinguished, and a subsequent modular-hierarchical phase, where a set of classifiers tailored to the previously identified major land-cover types is provided as potential descriptive criteria. Table 1 gives an example of how a land-cover class termed 'industrial' is described in the source manual (Anderson et al, 1976) and how it has been described, coded, and labeled through the LCCS classifiers.

Proprietary PC-based software (available at http://www.glcn-lccs.org) is available to create user-defined classes, such as the table 1 example, in a step-by-step process. It is also possible to supplement class definitions with optional attributes that may influence land cover—for example, climate, soils, and floristic aspects. In this way the LCCS generates mutually exclusive land-cover classes, which comprise: (1) a coded string of classifiers used (the LCCS string); (2) a standard LCCS term (the LCCS label); and (3) a unique numerical code (the LCCS code).

These data can be used to build an automatically generated LCCS map legend from the labels, or to link codes to a user-defined term in any classification terminology. For example, the Anderson class 'industrial' can be compared with, for example, the IGBP—DIS (International Geosphere—Biosphere Programme—Data and Information Service) (Loveland and Belward, 1997) land-cover class 'urban and built-up' areas, described in table 2.

The comparison uses the two LCCS strings that are codes for characteristic features of each land-cover class. The evaluation follows a set-based similarity matching procedure and calculates the ratio of similar attributes in the two classes to the

Table 2. The International Geosphere—Biosphere Programme—Data and Information Service (IGBP—DIS) land-use–cover class 'urban and built-up' definition and its description by the Land Cover Classification System (LCCS) classifiers, codes, and label (after Herold and Schmullius, 2004).

IGBP—DIS class

Urban and built-up. Land primarily covered by buildings and other man-made structures. Note that this class has not been mapped directly from AVHRR (advanced very high resolution radiometer) data. It is overlaid from the populated places layer from the Digital Chart of the World.

LCCS classifiers		**LCCS string**	**LCCS label**	**LCCS code**
B15	Artificial surfaces and associated area(s)	B15–A1	Built-up area(s)	5001
A1	Built-up			

total number of attributes in the referent class. IGBP—DIS (Urban and built-up) has two classifiers, B15 and A1, in common with the four classifiers, B15, A1, A4, and A12, used to define Anderson (industrial); the similarity between the two is estimated to be $2/4 = 0.5$. The LCCS procedure to compare similarities between classification systems corresponds to a hybrid ontology architecture in that it defines a shared vocabulary, which makes up the basic building blocks of the domain (Lutz and Klien, 2006). The classifiers act as standardized building blocks and can be combined to describe the more complex semantics of each land-cover class in any separate application ontology (classification system). The LCCS approach is, in this way, different from most other examples of standardized land-cover systems (for example, CORINE and IGBP), which follow a single ontology approach where all available semantic descriptions have been created with a very similar view on a domain and have to be shared by all users (Lutz and Klein, 2006). Early in the LCCS process, participants pointed out that a single ontology approach would reduce the application relevance of produced data (Herold and Schmullius, 2004). It was deemed to be more important to standardize the attribute terminology rather than the final categories, taking attention from the semantically problematic class name (for example, tropical rain forest) to focus on the descriptive features. It also enables users to compare categories within and across classification terminologies using the standardized descriptive features as a common language. In this way the LCCS is to some extent following one of the criteria of Bowker and Star (1999) 'to render voice retrievable' listed above, and it is also well aligned with current cognitive and information theoretic proposals to formally describe and compare categories (see Faucher, 2001; Gärdenfors, 2000; Mennis et al, 2000; Tversky, 1977). Conceptually the LCCS system follows a hybrid ontology approach in that it provides a set of characteristics to describe a land-cover class, and it relies on a feature-matching process to evaluate semantic similarity between land-cover classes.

The process of developing the LCCS collaboratively, involving many different stakeholders, further embraces the boundary-object view on classification presented earlier. Thus, we could think of the LCCS as a boundary object for cooperation and communication. The LCCS shared terminology also seems to have managed to balance the goal of a global generalized standardization of land-cover classification with the need for enough detail to ensure practical applicability. The LCCS is now being increasingly embraced by the land-cover community, and initial results are published from legend translations of, for example, IGBP, CORINE 2000, IPCC, GLC2000 and several other classification systems (presented in Herold and Schmullius, 2004). Nevertheless, the current LCCS shows a multitude of small but fundamental shortcomings that hinder its development into a fully dynamic classification system supporting the requirements of Bowker and Star (1999). Some examples of these shortcomings are as follows.

(a) Many existing definitions do not match the available classifiers exactly. For example, an IGBP woody vegetation class requires trees to be at least 2 m tall but LCCS classifiers use a 3 m cut-off for trees.
(b) Often classes are described only by a few classifiers (2 – 4) and this limits the ability to enable a sensitive analysis of category similarity. A vegetation class may not include any mention of details such as climate or tenure in its definition, but if these characteristics can be detailed by an expert they would help in rendering voice retrievability since it supports comparing this class to another class which includes those descriptors.
(c) The LCCS system is currently developed in 'traditional' working groups and uses proprietary software for its implementation. At this point it seems relevant to try to increase stakeholder participation and create open formats to enable a dynamic and ongoing discourse.
(d) Similarity between categories is based solely on a feature-based matching process in which all features are equally important. Current thinking on semantic similarity metrics indicates a need for additional metrics that also account for asymmetry in the evaluation (see Ahlqvist, 2004; Rodriguez and Egenhofer, 2004).

The next section elaborates on these shortcomings on the basis of ontology development and semantic interoperability in geographic information science, and also argues for some modifications to the current LCCS.

LCCS shortcomings and suggested modifications

The following suggestions are meant to improve further the existing LCCS and future developments of similar systems resulting in a classification-system approach that supports the dynamics of science as well as professional-user needs.

Unbounded classifiers

The design of classifiers in LCCS followed a requirement to define class boundaries that should be clear, precise, possibly quantitative, and rest upon objective criteria (Di Gregorio, 2005). In this way it follows a Boolean first-order logic for the classifiers (parameters), which creates an unnecessarily rigid and compartmentalized view on categories, albeit at a conceptually lower level. Many other attempts to overcome semantic heterogeneity (Devogele et al, 1998; Kavouras and Kokla, 2002; Rodriguez and Egenhofer, 2004) also rely on unambiguous identification of matching categories. One probable reason to maintain an underlying set of mutually exclusive and spatially exhaustive categories is that it is readily implemented in a standard database and compatible with a standard map output. Unfortunately this fails to represent many real-world situations where classes are frequently found to be vague and partly overlapping.

As an example of how a Boolean, first-order logic fails to accommodate category vagueness we look at the study by Lund (2006) who recently examined 720 different definitions from all over the world of the term forest. Of the 159 definitions listed that relate to forest as a land cover, those that cite a numerical threshold for tree-canopy cover give values that range between 10% and 80%. Some examples are shown in figure 3.

Examining the LCCS descriptive attributes (classifiers) that can be used to describe the characteristics of tree-canopy cover and tree height we note that they cover most of the domains over which the concept of forest varies amongst the listed examples. However, LCCS classifiers impose a restriction on tree height to be >3 m (B2, B5, B6, B7), or >2 m (B1) for woody vegetation, which limits the system from accurately and associated area(s)representing, for example, the Estonian definition that has a threshold of 1.3 m. What is also apparent from the figure is the almost continuous distribution values in these forest definition examples. The LCCS classifiers are limited to crisp ranges, which imposes a granularity that results in a loss of descriptive power. For example,

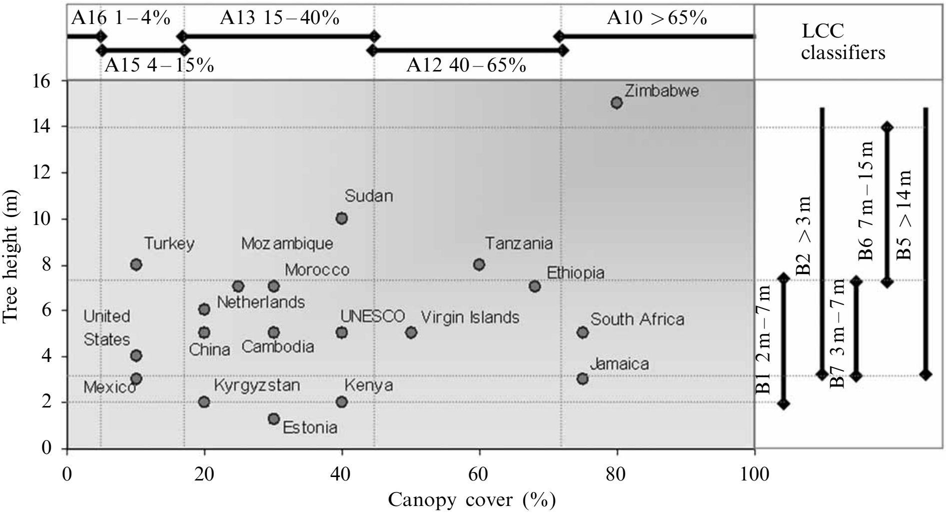


Figure 3. Criteria used to define 'forest' according to definitions found in Lund (2006) compared to the Land Cover Classification System (LCCS) canopy cover and tree height classifiers available for describing land-cover classes. Adapted from Comber et al (2005).

the difference between the UNESCO (>5 m height, >40% cover) and the Netherlands (>6 m height, >20% cover) forest definitions will not be recognized by the LCCS parametrization. The LCCS system is complicated further by the fact that the classifiers are arranged hierarchically so that some height and cover values for shrub cover are not accepted for tree cover. Furthermore, labels of life forms are 'hard-coded' into the attributes by stating that woody vegetation above a certain height is trees and below that height is shrubs. Accepting the dynamics of science, it seems most likely that the current LCCS classification rules that are relevant within the FAO framework will change over time, and that other cultures may look at vegetation domains in a different way. It is important to support not only the needs of a variety of end users but also to accept these dynamics.

Figure 3 clearly illustrates the need for a standard classification scheme to be flexible over a full range of possible attribute values. This motivates the search for a system design that does not impose a granularity at the attribute level and avoids reducing the measurement level to a nominal scale when ordinal or numerical information is available. To describe these different forest definitions accurately we need full freedom to express, for example, an Ethiopian definition of forest that sets thresholds at 7 m tree height and >68% canopy cover. The proposed solution is that a LCCS-style standard should only define the measurement scale and unit of measurement for the quantitative attributes and allow a user freedom to define any threshold values in that measurement domain. Comparison and analysis of definitions will still be possible as will be demonstrated in the next subsection. It is also possible to set up rescaling and transformation functions between measurement scales and units of measurement to enable users to select the measurement domain most appropriate to their context.

A related problem is that of classifiers that use nominal or ordinal values such as 'built-up' and 'nonlinear' in the description of the Anderson category industrial in table 1. These nominal terms are subject to the same problem of different interpretations as the land-cover categories they seek to describe, and that a parameterized solution is seeking to avoid. Although it is not always possible to retrofit descriptions such as these, it is desirable to substitute these nominal terms with quantifiable characteristics.

Table 3. Main life form labels in the Land Cover Classification System and their defining criteria (after Di Gregorio, 2005).

Life form	Tree cover	Shrub cover	Herb cover
Trees closed	closed	closed – absent	closed – absent
Trees open	open	closed – absent	closed – absent
Shrubs closed	sparse – absent	closed	closed – absent
Shrubs open	sparse – absent	open	closed – absent
Herbs closed – open	sparse – absent	sparse – absent	closed – open
Sparse vegetation	sparse – absent	sparse – absent	sparse – absent

Sometimes classification documentation includes quantitative guidelines as in the case of the LCCS description of 'life form'. This attribute is stated nominally as trees (closed or open), shrubs (closed or open), herbaceous, or sparse vegetation, but the description is based on a semiquantitative evaluation of the dominant type of vegetation in the uppermost vegetation layer and the height of the vegetation (see table 3). Thus, in this case, it may be possible to describe the life form semiquantitatively in terms of tree cover, shrub cover, and herb cover as ordinals or as cover percentages. Other descriptive attributes, such as spatial pattern and shape, which are found to be problematic because of their scale dependence (Di Gregorio, 2005) could possibly be reconciled on a quantifiable and scale-sensitive basis using pattern and shape metrics (see Edelman, 1999; Wästfelt and Arnberg, 2005).

Extended similarity assessment

Another important feature of a dynamic classification system relates to how it renders the voice retrievable by allowing a user to compare classes using the detailed class descriptions. In the feature-based approach to LCCS this process involves an information-reducing transformation of a numerical measurement scale to a nominal attribute in the similarity assessment. The current system therefore fails to recognize orderings within an attribute domain. As an example, this means that the inherent ordering of classifiers A10, A12, A13, A15, and A16 in figure 3 is not recognized in a feature-based similarity assessment. For example, classifiers A16 and A15 are as different (no match) as A16 and A10 (no match). However, the former pair of classifiers are very similar in canopy cover at 1% – 4% and 4% – 15%, as opposed to the latter pair of classifiers with very different canopy cover of 1% – 4% and >65%.

A suggested solution to this shortcoming is to have a similarity assessment process that can recognize ordered classifiers. The close correspondence with a semantic attribute classification and numerical clustering of multivariate data (Sokal, 1974) has led researchers to suggest similarity evaluations based on distance in a semantic data space (Gärdenfors, 2000; Nosofsky, 1986). Commonly used spatial distance metrics used for semantic similarity assessments are based on the Minkowski metric that can be easily adjusted to account for city-block, Euclidean, or other notions of semantic space. Different approaches have been suggested to enable calculation of distance between multidimensional conceptual regions (Ahlqvist, 2004; Schwering and Raubal, 2005). The example in figure 4 illustrates how two different aspects of similarity apply to two definitions of woodland, one used in Tanzania (Hyytiäinen, 1995) and the other in the USA (USDAFS, 1997).

The distance-based approach enables an *overlap* metric that measures the amount of shared features in the same way as the current LCCS, but without restricting measurement to predefined categorical intervals. It simply evaluates the overlap as a ratio of the overlap range (15 percentage units) over any chosen reference—for example,

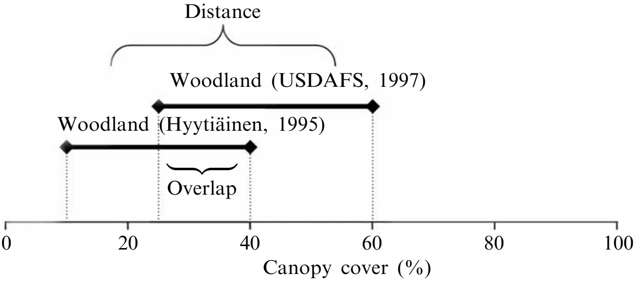


Figure 4. Two woodland definitions (Hyytiäinen, 1995; USDAFS, 1997) described as intervals on a numerical domain. Semantic similarity metrics 'distance' and 'overlap' are indicated.

the United States Department of Agriculture is 35 percentage units. The distance-based approach also enables some form of interval-based *distance* metric. One example is the dissemblance index (Kaufman and Gupta, 1985), which essentially measures the mean absolute difference (17.5%) between the lower bounds ($|10\% - 25\%| = 15\%$) and the upper bounds ($|40\% - 60\%| = 20\%$) of the intervals normalized by an attribute range—in this case 100 percentage units, giving a dissemblance index of 0.175 ($17.5/100 = 0.175$). Ahlqvist (2005a) recently argued for using combinations of those two metrics to evaluate the semantic relationship between concepts. The suggested methods for similarity evaluations can be implemented easily using fuzzy-set and fuzzy-number techniques demonstrated by Ahlqvist (2004) as a way to address the mentioned vagueness inherent in any conceptualization and its parameterization.

A rich formal narrative to create semantically rich class descriptions

The combined hierarchical–dichotomous–modular classification process of LCCS is argued to ensure use of the most appropriate classifiers and prevents the use of inaccurate classifier combinations (Di Gregorio, 2005). It is, for example, not possible to describe the water depth of a terrestrial land-cover type, but on the other hand it is currently not possible to describe a water seasonality aspect of forests using LCCS. In order to maintain a neutral system that is able to answer to many specific needs, it is important not to restrict the descriptive attributes to a subset as is the case through the narrowing down in the dichotomous phase of LCCS. To get a rich description that can be interpreted for many different purposes, it is important to create a rich formal narrative with as many descriptive attributes as possible. For coding purposes it may be practical to adhere to compound characteristics that collect a set of descriptive attributes with preset values, so that it is possible to say that, 'for characteristic X (say, built-up) this list of attributes $a_1, a_2, a_3, \ldots$ have the following values $x_1, x_2, x_3, \ldots$', essentially generating a class description that becomes reified in the new definition. In this way the dichotomous phase can be replaced by one, or a few, descriptive attributes.

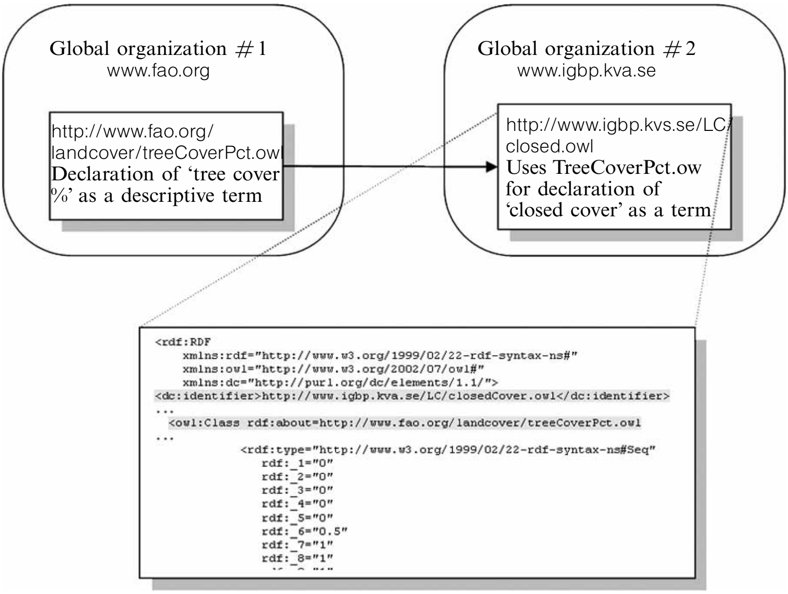
The set of possible descriptive attributes is, of course, infinite, but each application context will restrict the number of attributes to a manageable number necessary to discern between classes in the system. For example, land-evaluation applications concerned with soil conservation are likely to emphasize a very different set of descriptive attributes from that for an application interested in biodiversity management. Users who wish to use a taxonomy for some other purpose will be likely to retrofit additional attributes based on the already rich description provided in the original dataset and additional expert knowledge about the new application domain. This creation of a rich formal narrative is all part of rendering the voice retrievable for the intended meaning and to support user-driven terminologies that cross-reference each other. As modern mapping becomes increasingly multidimensional and dynamic we need rich class descriptions as part of metadata that can help to create a bettter interoperability not only between land-cover datasets but also with other types of geospatial data. A land-cover database does not require a unique Boolean formula, because each enumeration unit, be it a pixel or a polygon, has an identity and a class code or term can be derived on demand from a set of characteristic attributes. This design would ensure full flexibility, where different bases for classification and user requirements can be accommodated.

An open and dynamic syntax for exchanging formalizations of meaning

So far LCCS has not focused much attention on the issue of how to represent formally the developed definitions and translations. Exceptions are the proprietary software (Di Gregorio, 2005) and some initial work to define formally the currently proposed classifiers as a classification language based on formal language theory (Di Costanzo and Ongaro, 2004). Efforts to promote shared vocabularies in other contexts have emphasized the need to support the evolving and dynamic character of science through an open and flexible syntax for representing and contextualizing what data mean to people (Pike et al, 2003). The World Wide Web Consortium (http://www.w3.org/) and its semantic web activity focus on developing common formats for the interchange of data and on developing language for recording how the data relate to real-world objects. This activity is currently recommending the web ontology language (OWL) for publishing and sharing sets of terms, in order to openly support advanced web searching, software agents, and knowledge management. OWL is an extension of the extensible markup language (http://www.w3.org/XML/) and it provides a common description language with which to contextualize terms in any domain of interest by declaring classes of general things, which relationships can exist among those things, and the properties (or attributes) that those things may have. In a way this is similar to what the previously described WordNet and Cyc do, but a major difference is OWL's open, web-based foundation which enables users and producers of data to transparently exchange formalized meaning without the need to agree on specific software.

The current LCCS, together with suggested modifications, could be expressed in OWL as in the hypothetical example in figure 5. An organization such as FAO could, through a collaborative process, publish suggested parameters to use for land-cover descriptions. These form an initial shared ontology that other organizations can use to describe their own terminology or their interpretation of other terminologies. Ongoing development of existing and new land-cover terminology translations could then become a living continuous development and exchange that could adapt to new user needs in a flexible way. As a hypothetical example, the previously mentioned problem with bounded domains and categorized classifiers is replaced by a mapping between the LCCS unbounded classifiers specified on, say, http://www.fao.org to a definition such as the IGBP term 'closed cover' at http://www.igbp.kva.se that would be equally accessible to the user community for adoption or modification by publishing and linking to a modified definition.

In this example OWL serves as a metalanguage that enables a hybrid ontology to develop as a formally specified boundary object in a web-distributed classification process. Participation in the classification process is open to anyone, either through the adoption of published measurement domains and terms using those domains, or by suggesting additional measurement domains and terms by publishing and linking to existing definitions. Such a process allows for terminologies to appear as de facto standards, depending on which definitions tend to be used by a specific user community, and can potentially lead to the envisioned dynamic social classification process described at the outset.



```
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:dc="http://purl.org/dc/elements/1.1/">
<dc:identifier>http://www.igbp.kva.se/LC/closedCover.owl</dc:identifier>
...
  <owl:Class rdf:about=http://www.fao.org/landcover/treeCoverPct.owl
...
          <rdf:type="http://www.w3.org/1999/02/22-rdf-syntax-ns#Seq"
             rdf:_1="0"
             rdf:_2="0"
             rdf:_3="0"
             rdf:_4="0"
             rdf:_5="0"
             rdf:_6="0.5"
             rdf:_7="1"
             rdf:_8="1"
```

Figure 5. Outline of an open web-enabled ontology infrastructure based on the OWL specification language.

Conclusion

Categorization, classification, and standardization have been described as processes that bridge from individual-level, mental concepts to group-level, social interaction terminologies used for knowledge communication. From this background the FAO LCCS was described as a successful collaborative effort to balance the need for standardization and stability of terms with flexibility in individual use of those terms. Some important shortcomings were pointed out, and the following solutions were suggested to improve the currently proposed LCCS framework.

(a) Standardization should be restricted to unbounded domains of descriptive characteristics.

(b) Multiple continuous scale similarity assessments should be used to address different aspects of category relationships.

(c) A 'tell all you know' approach to terminology descriptions and translations creates rich formal narratives that support reuseability and render the voice retrievable.

(d) An open-format formalization for terminology description is necessary to leverage the LCCS methodology into a truly collaborative system flexible to the needs of science dynamics.

The proposed modifications and methodologies should be sufficiently general to be valid for any type of category description that uses descriptive features as a main component for category definition and object identification. The proposed classification framework should be applicable to any spatial data that use nominal categories as the main information carrier. Examples include information on soil, landform, gentrification, and exurbanization.

References

Ahlqvist O, 2004, "A parameterized representation of uncertain conceptual spaces" *Transactions in GIS* **8** 493 – 514

Ahlqvist O, 2005a, "Using uncertain conceptual spaces to translate between land cover categories" *International Journal of Geographical Information Science* **19** 831 – 857

Ahlqvist O, 2005b, "Using semantic similarity metrics to uncover category and land cover change", in *Lecture Notes in Computer Science* number 3799 (Springer, Berlin) pp 107 – 119

Anderson J R, Hardy E E, Roach J T, Witmer R E, 1976 *A Land Use and Land Cover Classification System for Use With Remote Sensor Data* US Geological Survey Paper 964 (USGS, Washington, DC)

Bartholomé E, Belward A S, 2005, "GLC2000: a new approach to global land cover mapping from Earth observation data" *International Journal of Remote Sensing* **26** 1959 – 1977

Beckett P H T, Webster R, McNeil G M, Mitchell C W, 1972, "Terrain by means of a data bank" *The Geographical Journal* **138** 430 – 449

Bishr Y, 1998, "Overcoming the semantic and other barriers to GIS interoperability" *International Journal of Geographical Information Science* **12** 299 – 314

Bjelland T K, 2004, "Classification: assumptions and implications for conceptual modeling", dissertation, Department of Information Science and Media Studies, Faculty of Social Science, University of Bergen, https://bora.uib.no/bitstream/1956/677//1/bjelland.pdf

Bowker G C, 2000, "Mapping biodiversity" *International Journal of Geographical Information Science* **14** 739 – 754

Bowker G C, Star S L, 1999 *Sorting Things Out: Classification and its Consequences* (MIT Press, Cambridge, MA)

Burley T M, 1961, "Land use or land utilization?" *The Professional Geographer* **13** 18 – 20

CEC, 1995 *CORINE Land Cover* Commission of the European Community, http://reports.eea.eu.int/COR0-landcover/en/land_cover.pdf

Clawson M, Stewart C L, 1965 *Land-use Information: A Critical Survey of US Statistics, Including Possibilities for Greater Uniformity. Resources for the Future* (Johns Hopkins University Press, Baltimore, MD)

Comber A, Fisher P, Wadsworth R, 2004, "Integrating land-cover data with different ontologies: identifying change from inconsistency" *International Journal of Geographical Information Science* **18** 691 – 708

Comber A, Fisher P, Wadsworth R, 2005, "What is land cover?" *Environment and Planning B: Planning and Design* **32** 199 – 209

Devogele T, Parent C, Spaccapietra S, 1998, "On spatial database integration" *International Journal of Geographical Information Science* **12** 335 – 352

Di Costanzo M, Ongaro L, 2004, "The Land Cover Classification System as a formal language: a proposal" *Journal of Agriculture and Environment for International Development* **98** 117 – 164

Di Gregorio A, 1991 *Technical Report on the Land Cover Mapping of Lebanon* FAO Project NECP/LEB/001/SAU, Food and Agriculture Organization of the United Nations, Rome, http://www.fao.org/docrep/008/y7220e/y7220e08.htm

Di Gregorio A, 2005 *Land Cover Classification System Classification Concepts and User Manual* Food and Agriculture Organization of the United Nations, Rome, http://www.fao.org/docrep/008/y7220e/y7220e02.htm

Edelman S, 1999 *Representation and Recognition in Vision* (MIT Press, Cambridge, MA)

FAO, 1995, "Planning for sustainable use of land resources. Towards a new approach" *FAO Land and Water Bulletin* No. 2, http://www.fao.org/docrep/V8047E/V8047E00.htm

FAO, 1998 *World Reference Base for Soil Resources* World Soil Resources Report 84, Food and Agriculture Organization of the United Nations, Rome, http://www.fao.org/documents/show_cdr.asp?url_file=/docrep/W8594E/W8594E00.htm

Faucher C, 2001, "Approximate knowledge modeling and classification in a frame-based language: the system CAIN" *International Journal of Intelligent Systems* **16** 743 – 780

Feng C-C, Flewelling D M, 2003, "Assessment of semantic similarity between land use/land cover classification systems" *Computers, Environment and Urban Systems* **28** 229 – 246

Fonseca F, Davis C, Câmara G, 2003, "Bridging ontologies and conceptual schemas in geographic information integration" *GeoInformatica* **7** 355 – 378

Fritz S, See L, 2005, "Comparison of land cover maps using fuzzy agreement" *International Journal of Geographical Information Science* **19** 787 – 807

Gärdenfors P, 2000 *Conceptual Spaces: the Geometry of Thought* (MIT Press, Cambridge, MA)

Groom G, 2005, "Strategic landscape monitoring and information: towards coordination of remote sensing in the Nordic Countries" *TemaNord* 2005:542, Nordic Council of Ministers, Copenhagen

Gruber T R, 1993, "A translation approach to potable ontologies" *Knowledge Acquisition* **5** 199 – 220

Guarino N, 1995, "Formal ontology, conceptual analysis and knowledge representation" *International Journal of Human Computer Studies* **43** 625 – 640

Hahn U, Chater N, 1997 *Concepts and Similarity*, in *Knowledge, Concepts and Categories* Eds K Lamberts, D Shanks (Psychology Press, Hove, East Sussex) pp 43 – 92

Harvey F, 1999, "Designing for interoperability: overcoming semantic differences", in *Interoperating Geographic Information Systems* Eds M F Goodchild, M Egenhof, R Fegeas, C Kottman (Kluwer, Dordrecht) pp 85 – 97

Harvey F, Chrisman N, 1998, "Boundary objects and the social construction of GIS technology" *Environment and Planning A* **30** 1683 – 1694

Herold M, Schmullius C, 2004 *Report on the Harmonization of Global and Regional Land Cover Products Meeting, 14th – 16th July 2004* (Food and Agriculture Organization of the United Nations, Rome), http://www.fao.org/gtos/gofc-gold/docs/GOLD_2D.pdf

Hyytiäinen K, 1995 *Land Use Classification and Mapping for the East Usambara Mountains, Usambara Catchment Forest Project* TP 12, Department of International Development Co-optation, Finnish Forest and Park Science, Vantaa

Jennings M, Loucks O, Peet R, Faver-Langendoen D, Glenn-Lewin D, Grossman D, Damman A, Barbour M, Pfister M, Walker M, Talbot S, Walker J, Hartshorn G, Waggoner G, Abrams M, Hill A, Roberts D, Tart D, Rejmanek M, 2004 *Guidelines for Describing Associations and Alliances of the US National Vegetation Classification* version 4.0, July, Ecological Society of America, http://vegbank.org/vegdocs/panel/NVC_guidelines_v4.pdf

Jones C B, Alani H, Tudhope D, 2003, "Geographical terminology servers—closing the semantic divide", in *Foundations of Geographic Information Science* Ed. M Duckham (Taylor and Francis, London) pp 205 – 222

Kalensky Z D, 1998, "AFRICOVER land cover database and map of Africa" *Canadian Journal of Remote Sensing* **24** 292 – 296

Kaufman A, Gupta M M, 1985 *Introduction to Fuzzy Arithmetic* (Van Nostrand Reinhold, New York)

Kavouras M, Kokla M, 2002, "A method for the formalization and integration of geographical categorizations" *International Journal of Geographical Information Science* **16** 439 – 453

King J L, Star S L, 1990, "Conceptual foundations for the development of organizational decision support systems", in *Proceedings of the Twenty-Third Annual Hawaii International Conference on System Sciences* Kailua-Kona, HI, 2 – 5 January (IEEE Computer Society Press, Los Alamitos, CA) pp 143 – 151

Lenat D B, 1995, "CYC: a large-scale investment in knowledge infrastructure" *Communications of the ACM* **38** 32 – 38

Loveland T R, Belward A S, 1997, "The IGBP-DIS global 1 km land cover data set, DISCover first results" *International Journal of Remote Sensing* **18** 3289 – 3295

Lund H G, 2006 *Definitions of Forest, Deforestation, Afforestation, and Reforestation* Forest Information Services, Gainesville, VA, http://home.comcast.net/~gyde/DEFpaper.htm

Lutz M, Klien E, 2006, "Ontology-based retrieval of geographic information" *International Journal of Geographical Information Science* **20** 233 – 260

MacEachren A M, 1995 *How Maps Work* Representation, Visualization and Design (Guilford Press, New York)

Mennis J L, 2003, "Derivation and implementation of a semantic GIS data model informed by principles of cognition" *Computers, Environment and Urban Systems* **27** 455 – 479

Mennis J L, Peuquet D J, Qian L, 2000, "A conceptual framework for incorporating cognitive principles into geographical database representation" *International Journal of Geographical Information Science* **14** 501 – 520

Miller G A, 1995, "WordNet: a lexical database for English" *Communications of the ACM* **38** 39 – 41

Nosofsky R M, 1986, "Attention, similarity, and the identification-categorization relationship" *Journal of Experimental Psychology: General* **115** 39 – 57

Pawlak Z, 1991 *Rough Sets: Theoretical Aspects of Reasoning About Data* (Kluwer, Dordrecht)

Peuquet D J, 1994, "It's about time: a conceptual framework for the representation of temporal dynamics in geographic information systems" *Annals of the Association of American Geographers* **84** 441 – 461

Pike W, Ahlqvist O, Gahegan M, Oswal S, 2003, "Supporting collaborative science through a knowledge and data management portal", 2nd International Semantic Web Conference, Sanibal Island, Florida, 20 – 23 October, http://flatbox.geog.psu.edu/codex/jsp/help/Pike_etal.pdf

Rodriguez M A, Egenhofer M J, 2004, "Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure" *International Journal of Geographical Information Science* **18** 229 – 256

Salgé F, 1995, "Semantic accuracy", in *Elements of Spatial Data Quality* 1st edition, Eds S C Guptill, J L Morrison (Elsevier, Amsterdam) pp 139 – 151

Schwering A, Raubal M, 2005, “Measuring semantic similarity between geospatial conceptual regions”, in *Lecture Notes in Computer Science* number 3799 (Springer, Berlin) pp 90 – 106

Shrage M, 1995 *No More Teams!: Mastering the Dynamics of Creative Collaboration* (Currency Doubleday, New York)

Sokal R, 1974, “Classification: purposes, principles, progress, prospects” *Science* **185** 111 – 123

Star S L, 1989, “The structure of ill-structured solutions: boundary objects and heterogeneous distributed problem solving”, in *Distributed Artificial Intelligence* Eds L Glasser, M N Hahns, Research notes in artificial intelligence (Pitman, London) pp 37 – 54

Thompson M, 1996, “A standard land-cover classification for remote-sensing applications in South Africa” *South African Journal of Sciences* **92** 34 – 42

Tversky A, 1977, “Features of similarity” *Psychological Review* **84** 327 – 352

USDAFS, 1997, “Guidance for conserving and restoring old-growth forest communities on national forests in the southern region”, report of the Region 8 Old-growth Team, United States Department of Agriculture, Forest Service, 1720 Peachtree Street NW, Atlanta, GA 30309

Vegetation Subcommittee, 1997, “Vegetation classification standard. Report FGDC-STD-005”, Federal Geographic Data Committee, US Geological Survey, Reston, VA, http://www.fgdc.gov/standards/documents/standards/vegetation/vegclass.pdf

Vickery B C, 1975 *Classification and Indexing in Science* 3rd edition (Butterworth, London)

Wästfelt A, Arnberg W, 2005, “Hybrid characterisation—or how to use a local spatial model for extracting meaningful information from satellite data”, in *Proceedings of the 10th Scandinavian Research Conference on Geographical Information Sciences* Eds H Hauska, H Tviete, Stockholm, Sweden, pp 1 – 12, http://www.scangis.org/scangis200J/papers/wastfelt.pdf