Metadata for Geographic Information

Paula Ahonen-Rainio

ABSTRACT. Users of geographic data may not be able to afford to purchase and implement a dataset that does not finally meet their needs. Therefore, metadata has a very important role in the information supply environment of geographic data. The development of national/local spatial data infrastructures recognizes the importance of metadata, as do the digital libraries providing spatial data.

The new ISO 19115:2003 standard of metadata for geographic information is introduced briefly. In particular, geographic information can be made available as digital maps (or images) that are meant for visual use, or as datasets meant for computational use. Metadata for digital maps is closely related to the metadata elements for conventional maps and can be enhanced by providing a sample map with the data. The case of computational use of geographic data is more complex. There are several details that may appear crucial when determining the fitness of data for an intended use. Understanding the importance of the crucial factors in each use case requires professional skills from the users of metadata.

KEYWORDS. ISO 19115:2003 standard, metadata for geographic information, metadata users, spatial data infrastructures, geographic data

---

Paula Ahonen-Rainio is affiliated with the Helsinki University of Technology, P.O. Box 1200, FIN-0215 HUT, Helsinki, Finland (E-mail: paula.ahonen@hut.fi).

Journal of Map & Geography Libraries, Vol. 2(1) 2005
Available online at http://www.haworthpress.com/web/JMGL

doi:10.1300/J230v02n01_03

INTRODUCTION

Digital information has a growing importance in economic and social activities in a modern society. Vast investments are made in information, and the majority of these datasets very often include some sort of spatial component. Because information may be re-used without deleterious effect on the information itself, increasing amounts of geographic data are provided for shared use outside the original data-supplier organizations. Given the substantial financial costs of these datasets and the relative ease and importance of re-using data instead of recreating it, development of spatial data infrastructures at local, national, regional, and global levels aims toward an efficient geographic information supply that meets the needs of society (Groot and McLaughlin 2000; Williamson et al. 2003).

The value of geographic data depends, first, on the awareness by potential users of the data's existence and, second, on the data's suitability for analysis and solving of spatial problems. The first is a problem that libraries have been attempting to solve for centuries, and that the Web has made obvious to many, if not all, non-librarians who use the Internet.

But the second problem is not one that is immediately obvious to all potential geographic dataset users, since many of these potential users know little or nothing about how geographic data are collected and presented. Geographic data provide a representation of the real world for use in many ways, but perhaps most often in visual display and in computational analysis. Any representation of the real world is an abstraction that simplifies, approximates, and filters some objects and phenomena, and may completely ignore others. The need to view the real world from different perspectives for different purposes has resulted in a wide variety of geographic datasets. And not only does the view of reality vary from dataset to dataset, there is considerable variation in the data models and data structures, the data quality, and the conditions and constraints that regulate the use of data.

In this context of variation, geospatial metadata that describe characteristics of geographic data are an essential asset primarily by making the data relatively readily usable, and then also by increasing the value to geographic datasets. Metadata may be defined as "a set of data that describes or gives information about other data" (OED). The "meta" prefix here, as with other technical terms (especially in computing), denotes data that operate at a higher level of abstraction. Metadata may be: embedded in the dataset described; in an external file and used without the actual data at hand; or for any one dataset located in both places. The content of metadata may vary from brief identifying information to an exhaustive description of a dataset. Though the above definition does not consider the forms of metadata, metadata generally are presumed to be formal and consistent so that not only human users, but also computers, can handle them.

The roots of metadata begin in the earliest collections of information but the concept has received much more attention in the digital era, as computer-readable metadata play a crucial role in database management and, more recently, in data transfer and the management of information services. The need for describing geographic datasets was acknowledged in the mid 1980s when the production of digital geographic data was established and ideas for national spatial data infrastructures started to emerge (DoE 1987; Moellering 1986; Vahala 1986). The early 1990s was a period when several geospatial metadata projects were launched (for a collection of examples, see Medyckyj-Scott et al. 1991).

To be effective in use, the contents of metadata–i.e., the metadata elements that are used for describing the characteristics of geographic data–must be in the same format for each dataset so that the descriptions are comparable. This need was recognized as a mutual interest by many geospatial metadata developers in the international field over a decade ago and has now resulted in the international standard, ISO 19115:2003, "Geographic Information–Metadata" (ISO 2003). The standard is based on experience gained in the development and use of previous metadata standards in Europe (CEN 1998), the United States (FGDC 1994), Canada (CGSB 1994) and Australia (ANZLIC 1995) and some standards for geographic data transfer and exchange that include metadata elements, such as the Digital Geographic Information Exchange Standard (DIGEST), the International Hydrographic Organization's Special Publication 57, and the United States Geological Survey's Spatial Data Transfer Standard (SDTS) (Danko 1997).

The realization of metadata should be determined by their use and intention. Generally speaking however, just as with library cataloging, and also with metadata for spatial data infrastructures, the resources for creating metadata are limited. Production of metadata requires human input, despite recent attempts at automating some parts of production, and therefore it is an expensive undertaking. Therefore, it is preferable that as many different metadata needs as possible are met in as far as possible by a single collection of general-purpose metadata. While this unavoidably leads to compromises between conflicting requirements, fortunately the needs of different uses and of different users of metadata are not mutually exclusive.

The major different uses of geospatial metadata are briefly surveyed and summarized in the next section. In section 2, the focus is on the profound question: What do users need to know about geographic data in order to understand its characteristics and decide whether a dataset is suitable for an intended use? (This section is intended for readers without an extensive familiarity of geographic-data representations.) These characteristics also explain why a specific geospatial metadata standard was developed instead of the adoption of a generic metadata standard such as the Dublin Core. The principle metadata elements of the ISO 19115:2003 are then listed in section 3. Aspects of the usability of metadata services are considered in section 4 since of equal, and one might also say parallel, importance to the content of metadata are the services of providing these metadata to users, and whose services are crucial in building the usability of data. Concluding remarks lead us to ask who should write the metadata for geographic datasets and act as an intermediate between data suppliers and users of data.

DIFFERENT USES OF GEOSPATIAL METADATA

Spatial data infrastructures (Groot and McLaughlin 2000; Williamson et al. 2003) note the role of metadata in getting vast amounts of geographic data resources–that were very expensive to collect and format–into use for the benefit of society, and maintaining the value of these data and their applications over time (Nebert 2001). One major value of metadata is that its presence should reduce production of duplicate data. These aims imply several uses of metadata, presented briefly below.

Inventory of Geographic Data Resources

Metadata can be seen as an inventory of data resources that are advantageously used in the administration of data. This is an important aspect of metadata, not only within organizations producing geographic data (Akervall et al. 1991; Nebert 2001; Taylor 2001) but also in user organizations when the administration and acquisition of geographic data are coordinated (Danko 1997). In the former case, metadata may be embedded in process management systems of data production and not considered as an item on its own. In the latter case, the same may occur but because of the difficulties of querying cross-database metadata collections it has its own complexities and difficulties.

As inventories of data resources, metadata also serve professional users working with geographic information, in that metadata enables the users towards constructing pragmatic knowledge about what kind of data are available and what the data's potential or limiting qualities are. This knowledge is constructed throughout users' careers by various means, e.g., education and work experiences, and through many different information sources such as colleagues, data producers, and metadata services. Thus, metadata sets are only one of these means, and their role varies from time to time. For novice users, metadata can provide an overview that is difficult to obtain in any other way, whereas experts may perhaps benefit primarily from information about changes in data resources, such as new datasets or amended quality of older ones (MADAME 2000).

An example of a metadata service that supports novice, or indeed any other users, in understanding the contents of metadata is the Eurogeographics' metadata service (see Geographical Data Description Directory (GDDD) at <http://www.eurogeographics.org/>), which provides not only metadata but also explanations of enumerated metadata items.

Management of Customer Satisfaction

Customer satisfaction is a mutual interest of users and producers of geographic data, even though public-sector assets dominate the market of geographic data. The process of building customer satisfaction is complex, including cognitive and emotional factors (Gale 1994; Kotler 1997). Proper metadata can prevent unrealistic expectations of the data by users, and consequent dissatisfaction when using geographic data (Ahonen-Rainio 2001). Users of geographic data have frequently expressed their dissatisfaction with the lack of metadata for datasets, and brought up the negative consequences of that lack in various forums, with the fortunate result that the availability and quality of metadata are gaining more attention. For example, Taylor (2001) states the role of metadata in proper use of geographic data; Danko (1997) mentions metadata as a legal documentation that protects an organization if conflicts arise over the use or misuse of data. But generally speaking, metadata is not legally regulated, unlike product descriptions for foods or medicines for example. Imagine it for a moment–a metadata record that presents information about a dataset according to a regulation stating that such information must be accurate!

Needs for the Use of Geographic Data

The most often-mentioned purposes of metadata form a chain: discover geographic data; evaluate their suitability for an intended application; retrieve and transfer the data; and, finally, use the data (Beard 1996; Lillywhite 1991), (Medyckyj-Scott et al. 1996), (Danko 1997; Guptill 1999; Taylor 2001). This chain, each step of which will involve labor, money, or both, arises from the need to use geographic data–that have many uses–for one specific intended application. Any failure in this chain leads to a disgruntled user. Discovery is the most crucial step, then evaluation of each dataset found is the next critical step, which enables users to select appropriate datasets from a wide collection of data in those cases where the users have the good fortune to have more than one dataset from which to select! Metadata must be accessible to users for the evaluation of data, leading to a solid assessment of the fitness of the dataset for the intended use. This must be achieved before further steps are taken prior to the next decision in the chain, retrieving the desired datasets. The last links in the chain–retrieval, transfer, and use–are then performed (if all has gone well) with a collection of datasets from which the undesirable sets have been weeded out.

Data Warehouses

We shall next consider metadata as a condensed representation of underlying data that may be used for decision-making (Günther and Voisard 1998). In this sense, both metadata and data appear to users as means of gathering information for a proper decision, and distinction between the two for this purpose may well be irrelevant. Umwelt-Datenkatalog (UDK), an environmental data catalogue described by Günther and Voisard, is an example of this kind of use of metadata and data. Longley et al. (2001) consider metadata an abstraction of data and, as such, an extreme generalization of geographic data; an approach that also obscures the distinction between data and metadata. In these cases metadata comes close to a data warehouse concept (Widom 1995). Given that metadata provide both information about data quality and also a consistent representation of the underlying data, especially important given that datasets–sometimes even one given dataset!–may be heterogeneous as far as data models, formats, and storage media are concerned (Bédard et al. 2001).

Semantic Interoperability

A recent innovative use of metadata relates to semantic interoperability of geographic information systems (Bishr 1998) and geographic information networks (Green and Bossomaier 2002). Here, the role of metadata relates to computer-interpretable semantics, where, especially in information networks, the roles of metadata and document mark-up can be merged. Online geographic data services (Green and Bossomaier 2002) assume metadata is embedded within data files, although in practice metadata may be generated as a fixed part of the data–rather like a file header–during the data production process, or generated separately and then attached to the data.

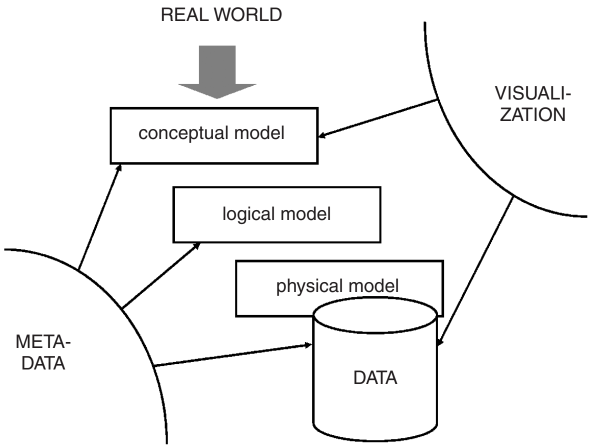
GEOGRAPHIC INFORMATION AS A REPRESENTATION OF THE REAL WORLD

The criteria by which users evaluate the fitness of a dataset for an intended use tend to vary from case to case; for example, in one case the main concern may be the data supply policy, whereas in another case the main attention is on technical constraints. But the most important questions in all cases are how well the geographic data represents the reality and how well that representation meets the view of the intended application.

This geospatial-data representation becomes evident in three forms: conceptual; digital; and visual. The conceptual representation is composed of ideas about the objects or phenomena of the real world, their properties and relationships; all of those aspects of the real world ignored in the conceptual representation are excluded from the data. The digital representation is composed of data representing the real-world instances, following the conceptual model. The visual representation is valid when a portrayal is embedded in or attached to the data, e.g., when the data compose a map (see Figure 1).

Within any one specialized information community, there is, by definition, an agreement on a shared set of concepts (Kuhn 2003), but cross-community concepts can easily differ remarkably. Burrough and Frank (1995) show the differences of views between users who deal with well-defined, crisp and often-static entities, and those who deal

FIGURE 1. The Levels of Real World Representation Embedded in Geographic Data



with complex dynamic systems that may only be approximated; Mark et al. (1999) report on differences in definitions of geographic terms among non-experts. Because of the differences in views, categorizations, and meanings that users have as far as the real world, there is an urgent need for describing the conceptualizations behind datasets, especially when the datasets are available for shared use across disciplinary boundaries.

An ontology is often proposed as a means of describing the semantics of spatial data both for data retrieval and data integration (Kuhn 2002; Rodriquez and Egenhofer 2003). These descriptions, as shared conceptualizations of subject headings (in library-ese) of application domains, are also able to support sharing of geographic data among different information communities. The need to understand the semantics of data is most urgent for integration of databases since correspondence between the components is critical. Despite studies of automated integration by ontological means, human knowledge of the semantics of data remains of supreme importance in final decisions about correspondences, relationships, and meanings of terms (Devogele et al. 1998). Kuhn (2003) proposes development of a more extensive semantic reference system, based on stronger formal foundations than most current ontology languages provide, that would enable transformations among semantic spaces and projections to sub-spaces. The fact remains that extensive ontological domains providing common references for all, or even a majority of, geographic datasets are yet to be developed. For the present, the evaluation of the semantics of geographic datasets is limited to descriptions at the database level or, in some cases, to informal documentations at an application domain level.

Conceptual schemas are the conventional means of documenting the conceptualizations in the context of database design. It is typical of these schemas that logical data models dominate database design as much as does human conceptualization; and the logical data schema is often the only consistent documentation available for a dataset. This data schema defines how concepts and their relationships are to be implemented in a given database management system.

Digital representation of geographic data is dependent on observations and measurements that follow the definitions. The differences between the data and the conceptual schema, as well as between the data and the real world, are indicative of overall data quality and are critical in the evaluation of datasets. On the other hand, we find that when the suitability of data for visual use is evaluated the quality criteria are different; these are discussed below.

Modeling of the Real World for Spatial Data

Geographic data can represent almost any kind of object or phenomenon in the real world, with the only common factor being the direct or indirect association with a location relative to the Earth (following the terminology of the scope of ISO/TC 211); we shall set aside for the purposes of this paper discussions about spatial data of other planetary bodies and of the universe. Despite the richness of the real world and conceptual views of it, the representation of geographic data is in practice limited to two spatial modeling paradigms: object-based modeling; and field-based modeling (see, e.g., Longley et al. 2001).

In object-based modeling, the real world is thought to be composed of discrete objects. The objects, usually three-dimensional and irregular-shaped, are abstracted and reduced in dimensions and generalized in shape into basic geographical objects–points, lines (a set of linked points), polygons (a set of linked lines), and volumes (in a three-dimensional space). Their properties are described by tuples of attributes measured on nominal, ordinal, interval, ratio or cyclic scales. For example, a building is seldom represented as a three-dimensional volume but rather as a generally rectangular polygon or a reference point; and a road is represented as a centerline or a set of connected nodes rather than a surface or a 3-dimensional volume that includes the under-surface structures. The spatial abstraction and the precision of the location coordinates determine the level at which the spatial objects represent the many specifics of the real world.

In field-based modeling, continuous phenomena of the real world are considered, which implies that any point in space can be characterized in terms of a set of attributes. Each attribute is measured on nominal, ordinal, interval, ratio or cyclic scales. It is assumed that most properties vary smoothly so that attribute values at unsampled locations can be determined by interpolation. In practice, continuous fields are represented by sample points (and attached mathematical functions), regular grids, polygon networks or triangular irregular networks (TIN). When the field is represented by a set of points that capture the characteristics and changes of the phenomenon, the number of points as well as their location is critical in determining the level of detail of representation.

Both the grid and the polygon network are actually discrete representations of the continuous model. In a grid, the size of grid-cells defines the level of spatial detail. A polygon network is based on a classification of the phenomenon, such as soil type, forest stand or land use, and–as the polygons are assumed to be homogeneous–the level of detail is determined by the classification system rather than any spatial property. But while the spatial resolution of a polygon network is dependent on the attribute resolution, the spatial resolution of observations that are used for the formulation of the polygons affects on the homogeneity of the polygons (Painho 1995).

These modeling paradigms are appropriate when representing objects and phenomena with crisp boundaries and well-defined attributes. These types of objects and phenomena are typically man-made concrete constructions, such as buildings, roads and utility networks, or abstract objects that exist only in social systems, such as administrative units, parcels and natural protection areas (Frank 2001). These approaches assume a static invariable world and ignore objects that consist of interacting parts or display variation at many different levels of resolution (Burrough and Frank 1995); their limitations are evident when one deals with natural-world phenomena that are vaguely defined or that have indeterminate boundaries (Burrough and Frank 1996). Although many solutions are proposed for imprecise data, these solutions are at the processing level, and the transferable data tend to follow traditional approaches with crisp geometry and well-defined attributes. On the brighter side, uncertainty models may be formulated for and attached to datasets (Horttanainen and Virrantaus 2004). Standard methods are of course needed for describing the interpretation and processing of these data.

Resolution of Representation

The level of detail of representation is an important property of geographic data, one of the primary criteria in the evaluation of datasets. It is typically expressed for spatial data in terms of scale, the ratio of distance on the map to distance on the ground, or resolution-that is, the smallest perceivable unit in length or area (or volume) that may be discerned in the geospatial data. Even though in dataset evaluation scale or resolution is one of the primary criteria, the effects of scale or resolution are manifold, and not all are self-evident.

Level of detail is controlled by abstraction and generalization. Abstraction is typical of mapped data, where a blue wavy line means a river, and a cross symbolizes the location of a church. The more difficult of the two to deal with is generalization, which aims at selection of important features while disregarding unimportant features in respect to the intended use of the data. This separation of the important from the unimportant is a complex mental process involving functions such as ordering, distinction, comparison, combination, recognition of relations, drawing conclusions and abstraction (Brassel and Weibel 1988). The forms of generalization that are appropriate in each case depend on the nature of the phenomena that are represented. Therefore, to understand the effects of generalization of data, it is necessary to know something about the general nature of those phenomena (Müller 1991). Goodchild (2001) emphasizes the spatial dependence in this respect, where differences over a short distance can be ignored because they are likely to be small. What this ‘short distance’ is depends on the phenomena, and should be reflected in the resolution of the data. If the resolution of the representation is relatively high, data can be expected to be noisy; a lower resolution would inhibit perception of significant patterns. This is a contradiction to what happens when we work with exact objects, where the more accurate measurements mean a better representation.

Goodchild (2001) mentions slope (function of point spacing), population density, soil types, and land use classes as examples of phenomena that are scale-dependent in the sense that a length measure is inherent in the variable’s definition. A user who is evaluating data with scale-dependent or multi-scale phenomena must determine whether the data are of sufficient detail to capture significant spatial variations important to the intended application.

Precision and resolution are also important properties for attribute data. The level of detail is expressed by the precision of metric data and resolution of the classification system of nominal and ordinal data. Again, different classification systems help with understanding the nature of variation of different aspects of phenomena.

Conceptual Schemata in Documentation of Semantics

In database design, the users' view of the real world is captured in a formal conceptual schema. It presents one abstraction of the real world among an indefinite number of alternatives. A dataset that is an extract from a database inherits the conceptual schema of the source database or parts of it. The schema follows an established paradigm–such as entity-relationship or object-oriented modeling–and users' concepts must be expressed in terms of this chosen modeling paradigm. A conceptual schema defines the universe of discourse as, "all those [objects or phenomena] of interest that have been, are or ever might be" and for the related integrity constraints, "the information system cannot be held responsible for not meeting those [constraints] described elsewhere" (ISO 1987). The definition covers the object classes (or entity types, depending on the paradigm), their attributes and relationships, as well as the value domains of attributes. In object-oriented modeling, where complex objects and methods of the object classes may be defined, the expression power is higher.

A conceptual schema has a communicative role between users and designers of an information system (Date 1981). The layered model of generic database architecture (presented by ANSI-SPARC in the 1970s) distinguishes the conceptual level from the logical and implementation levels to allow definition of the database by using concepts that are familiar to users, without regard to data structures and database management techniques. In that sense, a conceptual schema is a suitable description of the conceptualization also to users that evaluate the semantics of data extracted from an existing database.

But conceptual schemata have their limitations. Because their role is to serve database design processes, they may leave out concepts or ideas about which the database designers and users have agreed, or that constitute background knowledge about the information system. Complex constraints that govern geographic data may be documented only to a limited extent; the power and mode of expression of conceptual schemata may not be sufficient, or certain expressions, e.g., those in predicate logic, may not be easily readable by ordinary users. Furthermore, despite their layered architecture, conceptual schemata are constrained by the rules of digital representation and data structures in computer technology. For this reason artificial concepts and relationships may need to be included in the schema; if they are not, the concepts may correspond only in part with users' views. As a consequence, the conceptualization of the real world documented in a conceptual schema is difficult to capture, and may not be readily comprehensible to users not involved in the design process.

Conceptual schema languages provide a formal notation that enables unambiguous human communication as well as translation to the logical level. Graphical notations are typically used in human communication; but the expression power of graphical notations is more limited than that of lexical notations and, for an example, constraints may be expressed in a natural language. Lexical notations are not easily readable by human beings and are best used for tasks that will be performed by computer software. In the past, the most common approach for conceptual modeling was entity-relationship modeling (Chen 1976), resulting in ER-schemata, of which there are many slightly varying notations, mainly in graphical form. Currently, UML (Unified Modeling Language) (Rumbaugh et al. 1999) is gaining ground as a standard modeling language, but the standard covers only graphical language; lexical notations are CASE-tool dependent. Conceptual modeling tolerates some inconsistencies, lacking definitions or natural language attachments for example, but when a conceptual schema is translated to a logical schema using a data definition language (such as SQL or XML), the definitions must be complete.

Conceptual schemata are widely available as documentation of databases. The three-layer architecture has also been followed in the design of geographic information systems (GIS) (Laurini and Thompson 1992; Worboys 1995), but an explicit conceptual model is not available for each or for all geographic datasets. For example, map data originating from digitized paper maps may be stored in a layered model (Tomlin 1990); the conceptualization is inherited from the original map and never documented explicitly as a conceptual schema. The main document for conceptualization of this kind of dataset is usually the instructions for fieldwork and compilation of the map representation. Unfortunately, these documents are seldom available to users outside the data-producer organization. On the other hand, many databases holding topographic map data have evolved dramatically since their origins and do indeed hold multi-representation data and follow a rich conceptual schema with tens of object classes. The conceptual-modeling approach assumes an object-based view of the real world, but some spatial phenomena are continuous and, as such, not divisible into discrete objects. Interest in their conceptualization lies in how phenomena are characterized, resulting in the resolution of measures and value domains. In these cases, the data structure at the logical level is the main design interest instead of a conceptual schema with perhaps only a single object class.

Digital Representation and Data Quality

There can be a substantial conflict between human thinking; empirical and constrained by confusion and fuzziness, and the formal models of computers that rely on mathematics (Couclelis 1999). Conceptual models of geographic information are implemented in vector and raster data structures (Laurini and Thompson 1992; Peuquet 1984; Worboys 1995) with extended relational, object-related, and object-oriented data models, in many cases seemingly following the traditional approach of map layers (Tomlin 1990).

In database design, the logical schema defines the data structures that are an important concern for evaluating whether the intended computational analysis is possible with the data as supplied. In the context of spatial data infrastructures, geographic data are transferred from original databases as files, and therefore it is the file structure of the dataset that stands for the logical schema. Topological relationships play an important role in spatial computation. They are either explicitly defined in the schema and expressed in the data, or implicit in the location of objects (as the neighborhood of grid cells).

The implementation level described by the internal schema is principally under the control of modern database management systems. In data transfer, the implementation schema is embedded in the transfer format. If the logical structure of a dataset being evaluated is found to be acceptable, the format should not be critical (Molenaar 1998).

The method of collecting or creating data affects the characteristics of the digital representation of the data (Walford 2002). For data producers, information about the production process indicates the quality of the resulting data, but the same information may be unintelligible to users of external data sources. Instead, these users need to know the quality of the resulting datasets, with quality most often being related to the ability to satisfy customers' needs. This leads to different definitions of quality for different users (Krek 2002) whereas, in the context of metadata, we need objective quality measures that leave the evaluation of quality, and therefore of satisfaction, to each user. The ISO 19113:2002 standard (ISO 2002) defines data quality elements, covering the spatial, thematic and temporal dimensions of geographic data.

Datasets may appear to be of relatively good quality and also error-free when examined individually and in terms of their internal consistency and validity. But the need by many users to integrate geographic datasets from different sources presents a challenge for describing data quality, regardless of whether the integration is based on spatial or on attribute data. In order to evaluate whether the integration is feasible in each case, consistent quality of information is required–a consistency that does not seem to this author necessarily to appear in brochures of every geographic data supplier since each supplier has its own standards and methods of presenting information to potential users.

Cartographic Presentation

When evaluating geographic data it is worthwhile to make a distinction between cartographic and non-cartographic data. Cartographic data are meant for the visual perception by human users, whereas non-cartographic data are for computational purposes. In practice, cartographic data are extensively used also for computational analyses because many geographic datasets still originate from digitized paper maps. As previously noted in this paper, it is important that users understand the effects of cartographic processes, especially the effects of generalization, on data quality. There are few measures that can provide exact information about these effects; thus, very often, evaluation is mainly an intuitive operation.

A map is inherently generalized because of the space in a map available for presentation. Geographic datasets that are created for computational use without concern for visual representation and do not need to be concerned with overcrowding a given geographic-area representation within a dataset; mathematically speaking, points are 0-dimensional and lines 1-dimensional without any width, they do not overload the space, and there is no reason why they would not have the same location.

A hardcopy map is a completely different world; when visualizing data on a map the presentation must be visually legible–and preferably aesthetic in appearance–when graphical signs that represent the data take their space. And the information that the scale conveys for the given hardcopy map is in both the level of detail of data and the scale of intended use.

The scale of a map traditionally determines how much space is available, and therefore how much data, and at what level of detail the data will be presented. But the scale in which users view a digital map is a completely different matter, varying widely depending on factors such as the size of the individual user's monitor for the personal computer.

Cartographic generalization is by nature irregular and characterized by multiple rules and constraints. It depends not only on the object or feature itself but also on their relations to neighboring objects of the same type or other types (Buttenfield 1985; Buttenfield 1989; João 1998). Consequently, an object may be included in the data although a similar object in another location is missing from the data–and in the cartographic sense this may not be an error in the data. Development of automated generalization has produced formal expressions of generalization, but typically these processes form exhaustive collections meant for computers to read. There are no easy or consistent means of describing to map users how the generalization process has affected the representation.

When the suitability of data for visual use is evaluated, visual legibility is the primary concern. As the 'errors' such as displacement resulting from cartographic generalisation are mainly unpredictable and difficult to describe, evaluation of the quality of map data is mainly up to the intuition of each user. Contradiction between requirements of visual and computational use becomes evident here. For example, deletion and combining of objects as a part of cartographic generalisation may enhance the legibility of a map and increase the quality, but if the same dataset is used for computational analysis the good cartographic qualities turn into negative quality factors.

STANDARDIZATION OF GEOSPATIAL METADATA

The international metadata standard for geographic information, ISO 19115 (ISO 2003), defines a schema and an extensive set of metadata elements for describing the characteristics of geographic data.[1] The metadata elements can be applied to datasets, dataset series, and individual geographic features and their properties. Among the metadata elements there are conditional elements for alternative descriptions, e.g., for different types of data (raster/grid; vector data) and for spatial extent. Some of the metadata elements are not meaningful in isolation but only in relation to certain other elements, and these structures are defined in the metadata schema. The standard recognizes varying needs for metadata by defining the majority of elements as optional. In addition, it gives rules for defining extensions to, or a profile of, the standard for communities with special needs.

The standard lists core metadata elements that are required to identify a dataset. The mandatory core elements describe the topic, spatial extent, and reference date of a dataset, forming the minimum information for cataloguing purposes and discovery. A point of contact for further information is included in the mandatory core metadata proposing that users would contact the data supplier if they want to know the details of the data; this will obviously be burdensome if data exchange is to become routine (Beard 1996). As previously noted, the majority of the metadata elements are optional, but it is obvious to librarians that many of them are necessary for many uses of metadata, beyond the fields needed by users to discover data.

The standard defines the metadata elements in the following sections:

- Metadata entity set (an aggregate of the other sections)
- Identification
- Constraints
- Data Quality
- Maintenance
- Spatial Representation
- Reference System
- Content (Feature and Attribute Information)
- Application Schema
- Portrayal Catalogue Reference
- Distribution Information

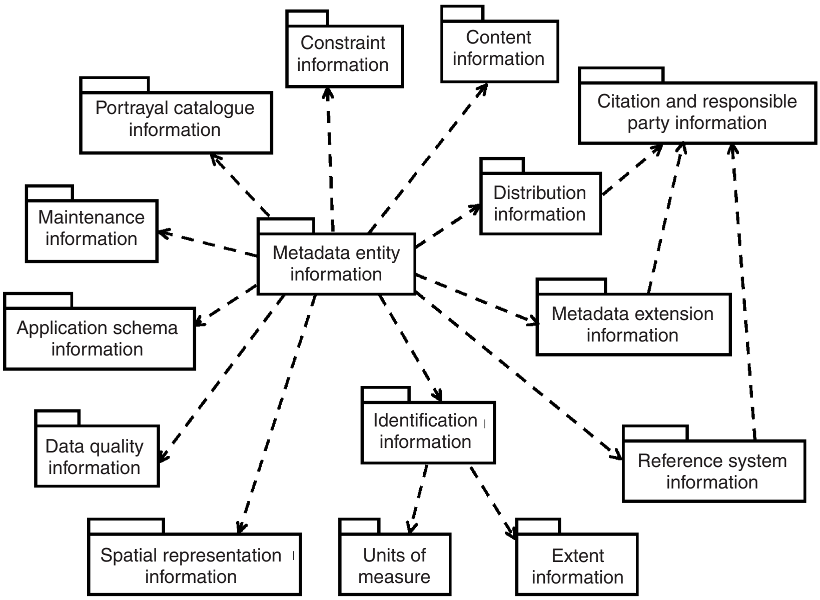
In addition, there are elements for defining possible extensions (see Figure 2).

By defining the metadata elements, the standard also establishes a common set of metadata terminology. It is composed of the names of the elements as well as their enumerated domains. Consistent terminology is a key factor in the comparability of datasets by metadata.

Metadata About Data Contents

The ISO 19115:2003 standard includes the metadata elements *feature catalogue* and *application schema* as the most extensive means of

FIGURE 2. The ISO 19115:2003 Metadata Packages



describing the semantics of datasets (ISO, 2003). An application schema corresponds to the conceptual schema discussed above; it may be provided as an ASCII file, a graphics file, or a software dependent file (with the software). Feature catalogues are derivatives of the application schemata including definitions of objects, attributes, and relationships; metadata may make reference to a feature catalogue or to a description of attributes of grid data.

Metadata elements *keyword* and *topic category* have as their main role the discovery of datasets by users via thematic dimension. They provide information about the content at a rough level. Some metadata elements also provide contextual information that can support understanding of the semantics, especially the elements *abstract*, *purpose*, and *usage* (ISO 2003). The context of the data is an important aspect in understanding the limited information of a conceptual schema (Molenaar 1998). These elements provide informal textual descriptions which implies that their use is up to the authors of the metadata, and therefore elements from metadata created by different authors may not be comparable.

The *abstract* as well as the *purpose* can prove to be important pieces of information in the evaluation of a single dataset; the conventions of using these elements are still to be established. The *usage* element–intended for describing the applications for which the data have been used or are not suitable–is an efficient means for users in communicating about their experiences with the data. In practice, application of this element may face many difficulties in an open universal environment, but in a closed metadata system actual users of data could author it (Ahonen-Rainio 2002).

The level of detail, when not evident from the feature catalogue or application schema, is described by the metadata element *spatial resolution*. It is expressed as a representative fraction referring to "the scale of a comparable hardcopy map or chart" (ISO 2003) or as a ground sample distance. Spatial resolution gives an indication of suitable application scales for data. It can be improved by *browse graphics*, that is, by visual samples of the dataset.

Metadata About Data Structures

Information about the logical level of geographic data is conveyed by metadata for *spatial representation* and *format*. For grid (or raster) data, the metadata elements of spatial representation give information about the geometry and geo-location of the cells and dimensions of the grid. For vector data, the metadata elements of spatial representation give information about the geometric objects and the level of topology. Closely related to spatial representation is information about the *reference system*.[2] Spatial reference is either by geographic identifiers, *ISO 19112:2003 Geographic information–Spatial referencing by geographic identifiers*, or by coordinates. In the first case, the reference system is labeled in the metadata. In the latter case, the reference system is labeled or described according to the *ISO 19111:2003 Geographic information–Spatial referencing by coordinates*. The metadata elements relating to *format* provide "description of the computer language constructs that specifies the representation of data objects in a record file, message, storage device or transmission channel" (ISO 2003).

Metadata About Quality

The ISO 19115 standard defines five main quality components: *completeness*; *thematic accuracy*; *temporal accuracy*; *positional accuracy*; and *logical consistency*. Each component has subdivisions; for example, completeness appears as commission or omission, and logical consistency can be expressed as conceptual, domain, format and topological consistencies. The standard does not predefine measures for quality components, but does define the structure of several elements that allow description of a measure, its evaluation method and procedure, and the result. Commonly used measures are, for example, a classification correctness matrix for thematic accuracy and RMSE (root-mean-squared error) for positional accuracy of point coordinates.

In addition to these quality components, ISO 19115 defines other metadata elements relevant in describing data quality. A data-production oriented element–*lineage*–provides information about source data and processing steps, including maintenance, of a dataset. This information is aimed at users who can judge the quality of data by the lineage description. Whether the data are up-to-date is a crucial matter for very nearly all applications. This may be described by temporal validity as part of temporal accuracy, together with *maintenance frequency*.

Metadata About Visual Representation

A bibliographic reference to a *portrayal catalogue* to display the data can be given as a part of the metadata. One basic way of portrayal is the metadata element *browse graphics* that "provides an illustration of the dataset" (ISO 2003). The standard does not specify the content of the illustration, so there are various ways in which the graphics can add information to metadata, with a conventional option being a display of sample data. When the dataset is a digital map the layout of the sample is inherent, and only the location and size of the sample have to be decided; when the dataset is not a map, the layout must be designed. The display "scale" also has importance in the interpretation of the graphics. The metadata for the browse graphics include the graphics file name and format and a text description of the illustration. Several browse graphics can be provided for a dataset.

In all cases, the rules and conventions of visualization of geographic information (Kraak and Ormeling 2002) should be followed.

Other Metadata Elements

In addition to the above mentioned metadata elements, user evaluation of suitability of datasets requires availability as well as user restrictions of the data. Availability of the data may be described in respect to geographic and *temporal extent*. Though these are important search criteria, in the evaluation their detailed description can prove to be even more so. Metadata elements for *geographic extent* are defined for bounding polygon, geographic bounding boxes, and geographical description by identifiers. Each of these has its benefits: the bounding polygon is an exact method; the bounding box is mathematically simple (though, in the worst case, it conveys very rough information); and the identifiers are a familiar means for users. The standard also defines metadata elements for vertical extent. In cases where users are not interested in the whole dataset, a valid piece of information is the tiling or geographic areas for which the data are available; this is to be included in the metadata of distribution.

Metadata about *constraints* define use limitations such as access, copyright, and security constraints. Metadata about *distribution* includes elements for digital transfer options, distributor, and standard order process. These may already have value in the evaluation stage before the actual distribution because, if not in the technical aspects, then at least the price of data can become a critical factor in deciding on a dataset.

Current Supply of Geospatial Metadata

Geospatial metadata sets and collections are developing in number, and several initiatives have been announced to convert existing metadata collections to conform to the ISO 19115 standard or its profiles. Despite present intentions, most of the current geospatial metadata have been created according to heterogeneous specifications and for data that have been produced at some time in the past. Therefore, such metadata tend to be devoid of detailed information, and several metadata elements are lacking values; and even worse, the accuracy and reliability of these metadata may not be known. In the long term, when we may hope that the creation of metadata will be integrated into the various phases of data production (Beard 1996), the extent and quality of metadata should be considerably enhanced. Software tools, such as ArcCatalog from ESRI, have been developed to help with the creation of metadata.

Most geospatial metadata that are used for cataloguing purposes are currently collected at the dataset level (Taylor 2001). The problem with metadata at the dataset level is the presumption that a dataset is homogeneous, which is not always the case. The quality of data may be different for different object classes, or it may vary from region to region, or between the periods during which data items have been collected. Especially large nationwide datasets often are heterogeneous in many respects; for example in Finland where the population density is very different in the southern and northern parts of the country the accuracy and currency of data tend to reflect the different intensity of activities in those regions. The wider the internal variation of the dataset, the larger the uncertainty of metadata describing the whole dataset. The ISO 19115 metadata standard provides a mechanism to reduce this uncertainty by defining subsets of the dataset as *scopes* and describing quality elements for each scope differently (ISO 2003).

Although the importance of geospatial metadata is generally acknowledged, there has until recently been much criticism on the usefulness and usability of metadata and related services as they have been implemented. Descriptions of data are considered to be dominated by data producers' views (Frank 1998), and doubts have been expressed about the readability of the metadata (Foresman et al. 1996) and the usefulness of even very detailed descriptions (Timpf et al. 1996).

USABLE SERVICES FOR GEOSPATIAL METADATA

Although standardization of metadata content in ISO 19115:2003 has emphasized both discovery of datasets and evaluation of their suitability for use (ISO 2003), the mainstream development of metadata services has targeted the use of metadata for discovery of data by users. Numerous geospatial metadata services, usually called catalog or clearinghouses, have been implemented at the local, regional and national levels, as well as across national boundaries (see, for example, links to various metadata services at <http://www.gsdi.org>). This development has been most extensive in the context of digital geolibraries (Goodchild 1998), especially in the United States where a well-known example is the Alexandria Digital Library (<http://www.alexandria.ucsb.edu>) (Smith 1996; ADEPT 2002).

Search functions based on indexes in cataloging have long traditions in library systems and provide a sound basis for data-discovery services for users. The fly in the ointment is that traditional text-based cataloging systems do not match the cognitive models employed by users of geographic data (Frank 1994), which is a combination of spatial, thematic and temporal criteria by which users may search for geographic data (Beard and Sharma 1997). Search systems using these latter criteria have been developed in digital geolibraries.

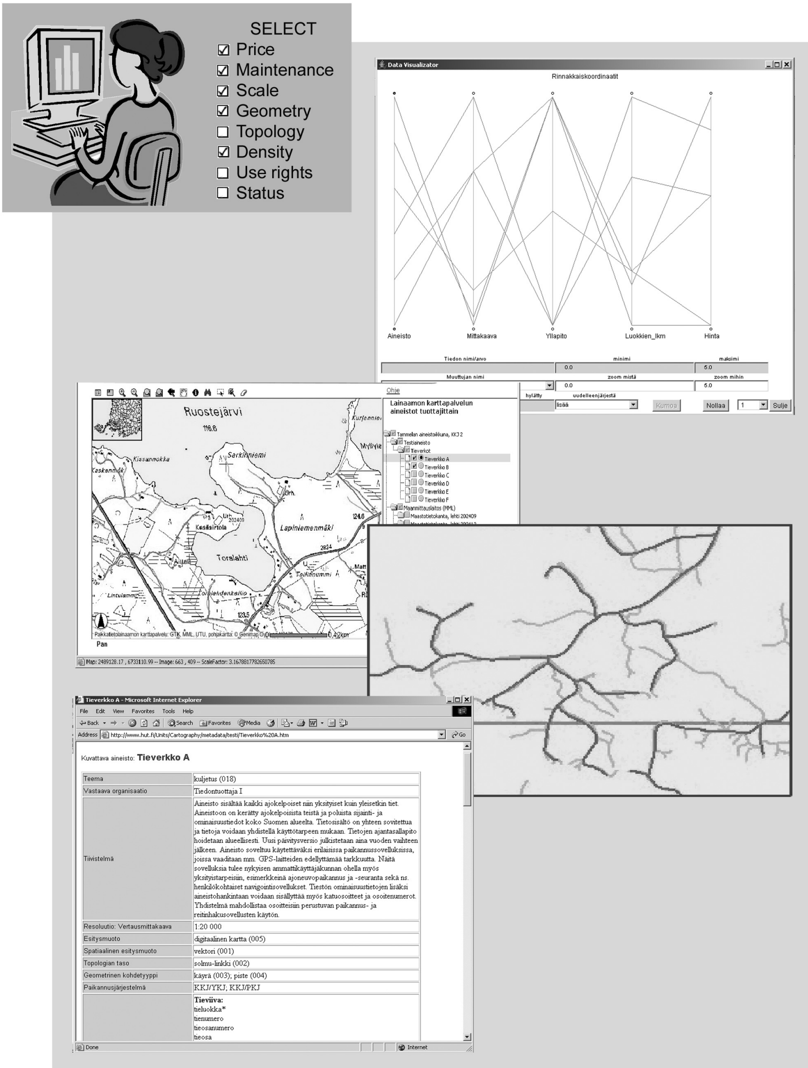
Because of the inefficiency of textual forms of metadata for dealing with georeferenced information, visual interfaces have been developed for searching in geolibraries (Ancona et al. 2002; Jung 1999). Beard and Sharma (1997) propose a multidimensional ranking schema based on thematic, spatial, and temporal values in a search context, with visualization of the ranking by bar glyphs. Göbel and Jasnoch (2001) discuss various information visualization approaches, such as box plots, glyphs and tile bars, and give examples with metadata that include dates and time periods, location/spatial reference, and categories of sample data. They also give estimates of user intuitiveness of these approaches. Albertoni et al. (2003) propose statistical graphics and multivariate visualization methods for visual data mining of geospatial metadata.

Despite these developments, current geospatial metadata services in the context of spatial data infrastructures still provide users with very limited evaluation tools. After searching, browsing of metadata is limited to one user-selected dataset at a time. Users cannot select metadata elements nor control the way in which metadata are presented, and a typical presentation is a text document or hypertext. The presentation does not support comparison of datasets even though it is a crucial function in evaluation. In most cases, users are evaluating datasets by several criteria but the criteria, or their relative importance, may not be predefined or fixed. Or, users may need an overview of potential alternatives before fixing the criteria; or they may have to make compromises between contradicting criteria. An exploratory visualization environment for geospatial metadata is proposed for evaluation purposes (Ahonen-Rainio and Kraak 2005); it would combine maps showing the location and spatial extent of data resources, an interactive parallel coordinate plot that displays metadata of several datasets together, sample maps illustrating datasets, and textual metadata (see Figure 3).

What kind of insights can users gain from these visualizations? Users' level of knowledge on various dimensions of geographic information–e.g., domain knowledge, geographic information science, skills with GIS tools, etc.–affect very strongly what users can interpret and understand from metadata (Ahonen-Rainio 2002). Even professional users of geographic information act in different roles, such as application users, system developers, and data administrators, each with different knowledge and needs for metadata.

User studies are scarce in the context of geospatial metadata as well as visual interfaces for geolibraries. Users' acceptance of visualization methods affects how much benefit users may derive from visual-

FIGURE 3. A Visualization Environment for Geospatial Metadata That Combines a Parallel Coordinate Plot, Sample Maps, and Textual Metadata Is Proposed (Ahonen-Rainio and Kraak 2005)



Reprinted with permission.

ization, or whether these methods are utilized at all. In the context of geovisualization, usability is thus considered a crucial design aspect (Slocum et al. 2001), and an increasing number of user studies are being documented (Andrienko et al. 2002; Furhman and MacEachren 2001; Haklay and Tobon 2003). Users of metadata typically view metadata occasionally, and study it even less often; therefore they cannot be assumed to be motivated to put effort into learning to interpret complex visualization methods.

The majority of exploratory visualization tools are developed for researchers and other users who are strongly involved in clarifying the characteristics and nature of data. Though users of metadata share the same need for insight, they consider metadata rather a supportive means than the main interest of their work. Visualization methods applied to metadata thus have to be intuitive to the occasional user. One usability test in an early design stage of metadata visualization indicates that there are differences in users' acceptability of some commonly used multivariate visualization methods, as well as in the usefulness of different browse graphics (Ahonen-Rainio 2003). Another test proposes that evaluation of datasets by metadata is an iterative process, and users need to gain insight of the characteristics of datasets available before they can specify in detail the criteria by which they select the most suitable dataset for an intended use (Ahonen-Rainio and Kraak 2005).

CONCLUSIONS

During the past two decades ideas about geospatial metadata have developed from simple descriptions of file structures and encoding to rich information that enable users to evaluate the suitability of data for different uses, as well as for computers to interpret the compatibility of different datasets. Despite advanced ideas and general acknowledgement of the importance of metadata for spatial data infrastructures in modern societies, availability of geospatial metadata is still limited and the resources needed for creating metadata are debated. The ISO 19115 standard for geospatial metadata is a solid foundation for development and cooperation and a welcome arrival upon the international-metadata standards scene (ISO 2003).

The understanding of geographic data resources available for shared use is a complex matter. In addition to technical aspects of geographic data that demand deep subject and computer knowledge, the issue of data semantics challenges us in our formulation of what constitutes metadata content for geospatial data. Metadata records are easily considered objective, but catalogers know that terminology and phrases used in describing data introduce subjective shadings, and that no two catalogers ever catalog even a paper map exactly the same way. Taking into account that metadata as a concept is about communication, it follows that expressions and interpretations are the key to success. The way in which metadata records are presented to users, and how users can interact with them, are a crucial part of the interpretation. And finally, visual environments are an important step in the development of metadata services, especially when geographic data are in question.

NOTES

1. A closely related standard, ISO 19131 *Geographic information-Data product specifications*, is under preparation and planned to be published in 2006. An extension to the metadata standard ISO 19115-2 *Geographic information-Metadata-Part 2: Extensions for imagery and gridded data* is planned to be published in 2007.

2. In addition to spatial reference systems, ISO 19115:2003 defines metadata elements for a temporal reference system.

REFERENCES

ADEPT 2002. Annual progress report of the Alexandria Digital Earth Prototype project. (ADEPT).

Ahonen-Rainio, P. 2001. Description of the Content of Geographic Datasets. In *The 8th Scandinavian Research Conference on Geographical Information Science*, Ås, Norway. T. Bjørke J. T., H. eds., pp. 125-139.

Ahonen-Rainio, P. 2002. Contextual Approach to Metadata of Geographic information. *Finnish Journal of Surveying Sciences* 20(1-2): 64-79.

Ahonen-Rainio, P. 2003. Concept Testing of Some Visualization Methods for Geographic Metadata. In *Proc. 9th Scandinavian Research Conference on Geographical Information Sciences*, Espoo, pp. 103-107.

Ahonen-Rainio, P. and M.-J. Kraak. 2005. Towards multivariable visualization of metadata of geographic information. In *Exploring Geovisualization*. M.J. Kraak, ed. Amsterdam: Elsevier.

Akervall, L., K. Degerstedt and B. Rystedt. 1991. Spatial metadata systems at the National Land Survey of Sweden. In *Metadata in Geosciences. Group D Publications*. D. Medyckyj-Scott, I. Newman, C. Ruggles, D. Walker, eds. Leicester, pp. 153-170.

Albertoni, R., A. Bertone, M. De Martino, U. Demšar and H. Hauska. 2003. "Visual and Automatic Data Mining for Exploration of Geographical Metadata." 6th AGILE Conference, Lyon, France, pp. 479-488.

Ancona, D., M. Freeston, T. Smith and S. Fabrikant. 2002. Visual Explorations for the Alexandria Digital Earth Prototype. In *Visual Interfaces to Digital Libraries. Lec-*

ture Notes in Computer Sciences Vol. 2539. C. Chen, ed. Heidelberg, Germany: Springer-Verlag.

Andrienko, N. et al. 2002. Testing the Usability of Interactive Maps in CommonGIS. *Cartography and Geographic Information Science* 29(4): 325-342.

ANZLIC. 1995. *Core Metadata Elements: Guidelines Draft 7*. Sydney: Australia and New Zealand Land Information Council: Working Group on Metadata.

Beard, K. 1996. A Structure for Organizing Metadata Collections. *3rd International Conference on Integrating GIS and Environmental Modeling*. Santa Fe, NM: NCGIA.

Beard, K. and Sharma, V., 1997. Multidimensional ranking for data in digital spatial libraries. *International Journal on Digital Libraries* 1(2): 153-160.

Bédard, Y., T. Merrett and H. Jiawei. 2001. Fundamentals of spatial data warehousing for geographic knowledge discovery. In *Geographic Data Mining and Knowledge Discovery*. J. Han, ed. New York: Taylor & Francis, pp. 53-73.

Bishr, Y. 1998. Overcoming the semantic and other barriers to GIS interoperability. *International Journal of Geographic Information Science* 12(4): 299-314.

Brassel, K.E. and R. Weibel. 1988. A review and conceptual model of automated map generalization. *International Journal of Geographic Information Science* 2(3): 229-244.

Burrough, P.A. and A.U. Frank. 1995. Concepts and paradigms in spatial information: are current geographical information systems truly generic? *International Journal of Geographic Information Science* 9(2): 101-116.

Burrough, P.A. and A.U.Frank. 1996. *Geographic Objects with Indeterminate Boundaries. GISDATA*. New York: Taylor & Francis.

Buttenfield, B.P. 1985. Treatment of the cartographic line. *Cartographica* 22(2): 1-26.

Buttenfield, B.P. 1989. Scale-dependence and self-similarity in cartographic lines. *Cartographica* 26(1): 79-100.

CEN 1998. *ENV 12657 Geographic Information-Data Description-Metadata*. Brussels: Comité Européen de Normalisation.

CGSB 1994. *The Canadian Directory of Information Describing Digital Geo-referenced Data Sets*. Ottawa: Canadian General Standards Board.

Chen, P. 1976. The entity-relationship model: Toward a unified view of data. *ACM Transactions on Database Systems* 1(1): 9-35.

Couclelis, H. 1999. Space, time, geography. In *Geographical Information Systems*. M. Goodchild, D.J. Maguire and D.W. Rhind, eds. New York: John Wiley & Sons, pp. 29-38.

Danko, D. 1997. Perspectives in the Development of ISO Metadata Standards. *Earth Observation (EO)/GEO World Wide Web Workshop 97, Washington DC*.

Date, C. J. 1981. *An Introduction to Database Systems*. Reading, Mass.: Addison-Wesley.

Devogele, T., C. Parent and S. Spaccapietra.1998. On spatial database integration. *International Journal of Geographic Information Science* 12(4): 335-352.

DoE 1987. *Handling of Geographic Information: Report of the Committee of Enquiry*. chaired by Lord Chorley. London: Her Majesty's Stationary Office (Department of the Environment).

FGDC 1994. *Content Standard for Geospatial Metadata (revised 1998)*. Washington, DC: Federal Geographic Data Committee (FGDC).

Foresman, T.W., H.V. Wiggins and D.L. Porter, D.L. 1996. "Metadata Myth: Misunderstanding the Implications of Federal Metadata Standards" First IEEE Metadata Conference, Maryland.

Frank, A.U. 1998. Metamodels for Data Quality Descriptions. In *Data Quality in Geographic Information: From Error to Uncertainty*. R. Jeansoulin and M. Goodchild, eds. Paris: Hermès, pp. 15-29.

Frank, A.U. 2001. Socio-Economic Units: Their Life and Motion. In *Life and Motion of Socio-Economic Units. Gisdata 8*. A.U. Frank, J. Raper and J.-P. Cheyland, eds. London: Taylor & Francis, pp. 21-34.

Frank, S. 1994. *Cataloging Paradigms for Spatial Metadata. Unpublished Ph.D. Thesis*. Orono, Maine: University of Maine, Dept. of Surveying Engineering.

Furhman, S. and A. MacEachren. 2001. Navigation in desktop Geovirtual Environments: Usability Assessment. *International Cartographic Conference, Beijing*, pp. 2444-2453.

Gale, B.T. 1994. *Managing Customer Value: Creating quality and service that customers can see*. New York: The Free Press.

Göbel, S. and U. Jasnoch. 2001. Visualization techniques in metadata information systems for geospatial data. *Advances in Environmental Research* 5(4): 415-424.

Goodchild, M. 1998. The Geolibrary. In *Innovations in GIS 5*. S. Carver, ed. London: Taylor and Francis, 59-68.

Goodchild, M.F. 2001. Models of Scale and Scales of Modelling. In *Modelling Scale in Geographical Information Science*. P.M. Atkinson, ed. New York: John Wiley & Sons, pp. 3-10.

Goodchild, M.F. and J. Zhou. 2003. Finding Geographic Information: Collection-Level Metadata. *GeoInformatica* 7(2): 95-112.

Green, D. and T. Bossomaier. 2002. *Online GIS and Spatial Metadata*. London: Taylor & Francis.

Groot, D. and J. McLaughlin, eds. 2000. Geospatial data infrastructure: Concepts, cases, and good practice. *Spatial information systems*. Oxford: Oxford University Press.

Günther, O. and A. Voisard. 1998. Metadata in Geographic and Environmental Data Management. In *Managing Multimedia Data: Using Metadata to Intergrate and Apply Digital Data*. A. Sheth and W. Klas, eds. New York: McGraw-Hill, pp. 57-87.

Guptill, S.C. 1999. Metadata and data catalogues. In *Geographical Information Systems, Volume 2, Management Issues and Applications*. P. Longley, M.F. Goodchild, D.J. Maguire and D. Rhind, eds. Cambridge: John Wiley & Sons, pp. 677-692.

Haklay, M. and C. Tobon. 2003. Usability evaluation and PPGIS: towards a user-centred design approach. *International Journal of Geographic Information Science* 17(6): 577-592.

Horttanainen, P. and K. Virrantaus. 2004. Uncertainty Evaluation of Military Terrain Analysis Results by Simulation and Visualization. *12th International Conference on Geoinformatics. University of Gävle, Gävle, Sweden*, pp. 473-480.

ISO 1987. ISO/TR 9007:1987. *(E) Information processing systems-Concepts and terminology for the conceptual schema and the information base*. International Organization for Standardization.

ISO 2002. ISO 19113:2002. *(E) Geographic information-Quality elements*. International Organization for Standardization.

ISO 2003. ISO 19115:2003. *(E) Geographic information-Metadata*. International Organization for Standardization.

João, E. M. 1998. *Causes and consequences of map generalisation*. London: Taylor & Francis.

Jung, V. 1999. MetaViz: Visual Interaction with Geospatial Digital Libraries. TR-99-017, *International Computer Science Institute*, Berkeley, California.

Kotler, P. 1997. *Marketing Management: Analysis, Planning, Implementation and Control*. New Jersey: Prentice Hall International.

Kraak, M.J. and F.J. Ormeling. 2002. *Cartography: Visualization of geospatial data*. London: Prentice Hall.

Krek, A. 2002. *An agent-based model for quantifying the economic value of geographic information. Doctor of Technical Sciences Thesis*. Vienna: Vienna University of Technology.

Kuhn, W. 2002. Modeling the Semantics of Geographic Categories through Conceptual Integration. In *GIScience 2002. Lecture Notes in Computer Sciences 2478*. D.M. Mark, ed. Berlin: Springer-Verlag, pp. 108-118.

Kuhn, W. 2003. Guest Editorial: Semantic reference systems. *International Journal of Geographic Information Science* 17(5): 405-409.

Laurini, R. and D. Thompson. 1992. *Fundamentals of Spatial Information Systems*. Academic Press.

Lillywhite, J. 1991. Identifying available spatial metadata: the problem. In *Metadata in Geosciences. Group D Publications*. D. Medyckyj-Scott, I. Newman, C. Ruggles, D. Walker, eds. Leicester, pp. 3-11.

Longley, P.A., M.F. Goodchild, D.J. Maguire and D.W. Rhind. 2001. *Geographic Information Systems and Science*. New York: John Wiley & Sons.

MADAME 2000. *Comparative Evaluation of On-line Metadata Services and User Feedback, Deliverable D2 of INFO2000 Project: PUB1108-MADAME Methods for Access to Data and Metadata in Europe*.

Mark, D.M., B. Smith and B. Tversky. 1999. Ontology and Geographic Objects: An Empirical Study of Cognitive Categorization. In *Spatial Information Theory: A Theoretical Basis for GIS. Lecture Notes in Computer Science*. D.M. Mark, ed. Berlin: Springer-Verlag, pp. 283-298.

Medyckyj-Scott, D., M. Cuthbertson and I. Newman. 1996. Discovering Environmental data: metadata bases, network information resource tools and the GENIE system. *International Journal of Geographical Information Systems* 10(1): 65-84.

Medyckyj-Scott, D., I., N., C., R. and D., W., 1991. *Metadata in Geosciences*. Leicestershire: Group D Publications.

Moellering, H. 1986. Developing digital cartographic data standards for the United States. In *Auto Carto London*. M. Blakemore, ed. London: Auto Carto London, pp. 312-322.

Molenaar, M. 1998. *An introduction to the theory of spatial object modelling for GIS*. London: Taylor & Francis.

Müller, J.C. 1991. Generalization of spatial databases. In *Geographical Information Systems Principles and Applications*. D.W. Rhind, ed. Harlow, Longman/New York: John Wiley & Sons, pp. 457-475.

Nebert, D. ed. 2001. Developing Spatial Infrastructures: The SDI Cookbook. OED, 2001. *Oxford English Dictionary Online*. www.oed.com (accessed 10.9.2004).

Painho, M. 1995. The effect of generalization on attribute accuracy in natural resource maps In *GIS and Generalization-Methodology and Practice*. R. Weibel, ed. London: Taylor & Francis.

Peuquet, D. 1984. A conceptual framework and comparison of spatial data models. *Cartographica* 21(4): 66-113.

Rodriquez, M.A. and M.J. Egenhofer. 2003. Determining semantic similarity among entity classes from different ontologies. *IEEE Transactions on Knowledge and Data Engineering* 15(2): 442-456.

Rumbaugh, J., I. Jacobson and G. Booch. 1999. *The Unified Modeling Language Reference Manual*. Boston: Addison-Wesley.

Slocum, T.A. et al. 2001. Cognitive and usability issues in geovisualization. *Cartography and Geographic Information Systems* 28(1): 61-75.

Smith, T. 1996. A Digital Library for Geographically Referenced Materials. *IEEE Computer* 29(5): 54-60.

Taylor, M. 2001. Metadata-Describing geospatial data. In *Developing Spatial Infrastructures: The SDI Cookbook*. D. Nebert, ed.

Timpf, S., M. Raubal and W. Kuhn. 1996. "Experiences with Metadata." *7th International Symposium on Spatial Data Handling, SDH'96, Delft, The Netherlands*, pp. 12B.31-43.

Tomlin, D. 1990. *Geographic Information Systems and Cartographic Modelling*. New York: Prentice-Hall.

Vahala, M. 1986. Data standardization of the integrated LIS in Finland. In *Auto Carto London: Hardware, Data Capture and Management Techniques*. M. Blakemore, ed. London: Auto Carto London, pp. 323-332.

Walford, N. 2002. *Geographical Data: Characteristics and Sources*. Chichester: John Wiley & Sons.

Widom, J. 1995. "Research Problems in Data Warehousing." *4th International Conference on Information and Knowledge Management*.

Williamson, I., A. Rajabifard and M.E. Feeney. 2003. *Developing Spatial Data Infrastructures: From concept to reality*. London and New York: Taylor & Francis.

Worboys, M. F. 1995. *GIS: A Computing Perspective*. London: Taylor & Francis, 376 pp.

Received: May 10, 2004
Reviewed: June 4, 2004
Accepted: July 30, 2005