EXPLAINING THE SCIENTIFIC AND TECHNICAL CHARACTERISTICS FOR ASSESSING SPATIAL DATABASE TRANSFER STANDARDS

Prof. Harold Moellering
Dept. of Geography
1036 Derby Hall
Ohio State University
Columbus, Ohio
U.S.A. 43210
Internet: GEOHAL+@OSU.EDU

ABSTRACT

For more than a decade various national and international organizations have been designing, developing, implementing and testing various spatial database transfer standards. However, if a developer, implementer, or user is interested in examining a particular standard in more detail, or one is interested in comparing two or more standards, no really effective material has been available to help make that assessment or comparison.

Since 1991 the ICA Commission on Standards for the Transfer of Spatial Data has recognized this problem and has been working to develop a comprehensive set of scientific and technical characteristics with which such assessments and comparisons could be carried out. The technical characteristics have been organized into 13 broad categories, ordered beginning with rather general characteristics to more than 85 middle level subcharacteristics, and then to more than 220 third level subcharacteristics. Together, these characteristics and subcharacteristics form an authoritative set of scientific material that can be used to assess and compare any national or international spatial database transfer standard.

1 INTRODUCTION

For more than a decade various national and international organizations have been designing, developing, implementing and testing various spatial database transfer standards. However, if a developer, implementer, or user is interested in examining a particular standard in more detail, or one is interested in comparing two or more standards, no really effective material has been available to help make that assessment or comparison. Worse yet, there has not been a set of scientific materials available to assist with the international comparison of such spatial database transfer standards. To complicate matters even further, different national and international standards utilize different definitions for concepts and terminology imbedded in the standard.

In 1991 the ICA Commission on Standards for the Transfer of Spatial Data published its first book that examined the work going on by 18 national and international organizations to develop spatial data transfer standards (Moellering, 1991). The publication of the book coincided with the ICA Congress held in Bournemouth, England that year. At that time most time most standards were in the process of being developed and tested, and few had been formally approved as a national or international standard. However, it was clear that the various standards that were under development were being developed with varying definitions for the concepts on which the standards were based, and with different terminology with which the standards were being defined. This is in addition to the fact that most of the standards were being developed in the native language of the country in question, and hence were different.

Since that time the ICA Commission has recognized this problem and has been working to develop a comprehensive set of scientific and technical characteristics with which such assessments and comparisons could be carried out. This paper explores the Commission's effort to develop the set of scientific and technical characteristics such that any national or international spatial data transfer standard could be characterized and assessed. The result is a set of thirteen major characteristics of the database transfer process, about 85 subcharacteristics and about 220 detailed characteristics. The following discussion begins with the concepts that underlie the spatial database transfer process, then describes the goals of the Commission to develop these characteristics, and finally examines the characteristics themselves.

2 GENERAL DESCRIPTION OF CHARACTERISTICS DEVELOPMENT EFFORT

This work began in the conceptual setting of the spatial data transfer process. It then moved to defining the goals of the work, and later beginning the effort to develop the general terminology and then define the transfer characteristics themselves.

2.1 Concepts Involved in the Transfer Process

At the base of the spatial data transfer process are several fundamental concepts that underlie the process. They are: real and virtual maps, deep and surface cartographic structure, Nyerges Data Levels, and syntax/semantics. The concept of real and virtual maps was originally developed by Moellering (1980, 1983, 1984, 1987) and provides an initial starting point to understand transfer standards. Read Maps have the characteristics of being both directly viewable as a cartographic object and having a permanent tangible (hard copy) reality. Standard products such as sheet maps, globes, 3-D relief models, and digitally produced maps are examples. However, there are three more classes of maps, called Virtual Maps, that lack one or more of these two characteristics. A CRT image is an example of a Virtual Map Type 1 that is directly viewable as a cartographic image, but does not have the permanent tangible reality. It is still a map and can easily be transformed into a Real Map by sending the image to a hard copy unit. A Virtual Map Type 3 lacks both of the characteristics above. What kind of a map can that be? It is a data file of spatial (cartographic) data and information that is in a digital form. All digital spatial databases are in the Virtual 3 form. The database transfer standards that are of interest to the ICA Commission are all Virtual Map Type 3 databases.

The second crucial concept required to understand the spatial database transfer process is that of deep and surface cartographic structure as articulated by Nyerges (1980a, 1991). Conventional cartographic maps and images are excellent examples of surface structure representations. However, it is well known that such graphic representations are based on numerical data which are not in a graphic form. That data and other relationships which may not necessarily be cartographically representable form the basis of deep structure. It turns out that the spatial (cartographic) database transfer process is involved in deep structure transfers of spatial data. What are being transferred are deep structure spatial (cartographic) databases and not graphic pictures.

The third important concept is that of Nyerges Data Levels (1980a, 1980b). Nyerges has shown that there are six defined levels of spatial and cartographic data: Data Reality (information from the real world), Information Structure (a formal model of information), Canonical Structure (a data model), Data Structure (what most think of as data structure), Storage Structure (the file structure), and Machine Encoding (the way in which bits and bytes are encoded in the architecture of the system hardware).

The fourth crucial concept to spatial database transfer processes is that of syntax/semantics. Syntax is the encoding of spatial information into a form that can be transferred digitally. Hence, the syntactic structure of the digital spatial data sending system to be transferred is encoded into the form of the transfer format or metafile transfer mechanism so the file of spatial data can be transported to the receiving system and decoded into its data structure. Semantics deals with the meaning of the information that is contained in the data. In order for the transfer to be really successful, the meaning of the information embedded in the data from the sending system must be compatible with the meaning of information in the receiving system. One example of the need for compatible semantics are recent trials to transfer databases of information of wetlands between several systems. Such tests ran into complications because the definitions of what constitutes a wetland differed between the systems and hence the transfers did not work out. The transfer syntax was successful because the data was encoded and physically transferred successfully, but the overall transfer failed because the definitions of the wetlands information embedded in the data in the transfer file was different from the definitions in the receiving systems, and hence not compatible. The lesson here is that the concepts of both syntax and semantics are important to properly understand the transfer process.

Briefly stated, a spatial database transfer process involves Type 3 Virtual Maps (data bases) in a deep structure form in a Nyerges Data Level 4 (Data Structure) which is encoded in a workable syntax that can be transferred successfully. In order for the process to be successful, the semantic definitions in the sending and receiving system(s) must be compatible. The flexibility of the transfer process ranges from that of a fixed format, such as forcing data into a production format, to a very flexible general transfer mechanism that can transform the data structure from the sending system into that of the receiving system.

2.2 Goals of the Commission Effort

This effort by the ICA Commission to develop these characteristics began in earnest at the 1991 ICA Congress in Bournemouth, England. The first task was to define the goals for the work in the 1995-99 ICA cycle. The following goals were discussed and developed by the Commission:

- support education regarding transfer standards;
- standardize the terminology and structure used for describing transfer standards;
- facilitate the description and interpretation of transfer standards for users;
- facilitate the comparison of alternative standards;
- provide for 'value-free' comparisons of selected standards;
- enable users to apply their own value judgements to the comparisons;
- enable short listing of potential standards by users for particular applications; and
- assist standards implementers to identify international commonalities between the various national standards.

These goals of the Commission for the time period provide the framework for an individual to more clearly understand the important components of the transfer process, and to develop a set of characteristics with which any standard can be assessed and understood.

2.3 Development of the General Terminology and the Characteristics

During the discussion of the goals necessary to successfully carry out this effort, it became that clearly evident that establishing a general terminology with a common set of definitions so that Commission members could understand what each member was saying. This complicated by the fact that standards from individual countries are written in the language of that country, which because of the highly technical nature of the task are extremely difficult for a non-native speaker of that language to understand. This situation is further complicated by the fact that different standards have different definitions for similar concepts, and hence that particular standard is defined with a different conceptual structure. Hence, the several standards written english use the same concepts with different definitions in different ways.

The Commission members had to establish a common set of terms with an agreed set of definitions just so the Commission members could discuss the overall topic at hand. This initial task included defining a general standards vocabulary to facilitate these general discussions by the members of the Commission, and later a specific set of terms that are particular to the characteristics and listed by sections in it. With these definitions the Commission has only tried to facilitate and clarify its own work and clearly convey the meaning of its work on its Technical Report and planned book to the reader. These definitions have not been intended to be the official world definitions. Rather, that task will fall to the newly formed Technical Committee 211, (TC 211) Geographic Information/Geomatics, of the International Standards Organization. Those official world definitions will probably be developed by the ISO TC211 over the next few years.

Developing the set of characteristics that can be used to assess any spatial database transfer standard has been the major task for the Commission for the last four years. This specific effort has taken two full years to complete and has resulted in a very carefully developed and refined set of assessment characteristics. There are 13 major characteristics that involve all the major aspects of the transfer process. This includes basic characteristics of various aspects of the transfer process, kinds of objects defined, update and query capabilities, quality information, and metadata, among others. Within these 13 fundamental sets of characteristics there are about 85 secondary characteristics. These secondary characteristics provide many additional facets of the of each of the 13 major characteristics is a way to expand the understanding of the standard being examined, and to provide a means for comparing and contrasting various standards with each other. Beyond the secondary level, there are about 220 detailed characteristics that further amplify the technical aspects of the 13 basic characteristics. These detailed characteristics provide a wealth of information on the details of the design and facilities of a standard. It explores such things as detailed yes-list/no style questions for may of the characteristics at this level where such an approach is appropriate. In other places it requests lists such as the Table of Contents for the standard, or a list of the feature/object definitions in the standard. The result is a large volume of detailed information about the standard being assessed that can be used by anyone who is interested in the scientific and technical characteristics of it.

3 DISCUSSION OF THE 13 MAJOR CLASSES OF CHARACTERISTICS

With the above discussion about the organization and structure of the characteristics at the three levels, one can then turn to a discussion of the substance and content of the 13 fundamental classes of characteristics:

- Administrative Information
- Transfer Context
- Transfer Specification Method
- Conceptual Data Model/Schema
- Transfer Process
- Transfer Elements
- Update Information
- Query Information
- Quality Information
- Feature Information
- Attribute Information
- Relationship Information
- Metadata Information.

An initial module of information about who conducted the assessment of the standard is included as section zero.

3.0 Assessment Information

The assessment information provides information on who prepared the assessment and when the assessment was prepared. A brief 200 to 300 word summary of the standard is included to introduce the reader to the scope and content of the standard in a general way. In many cases the general objective of the standard are also reviewed and described.

3.1 Administrative Information

Initially this section specifies the official name of the standard, version, acronyms and the like. The institution(s) responsible for developing, testing, conformance, maintenance, and distribution of the standard are listed. Then a list of anticipated end users is provided, along with a list of application areas and products to be served. Significant information is requested about the development and managerial history of the standard along with its current status in terms of official recognition and where it is in that process. Information on the existence of documentation, software tools and training materials is also requested. It also provides a point of contact for further information on the standard.

3.2 Transfer Context

The transfer context is an examination of the broader scope of the standard in terms of spatial referencing, conceptual data model, conceptual data schema, etc, as well as a question on the kinds of languages supported, such as data description, query language, general exchange mechanism, etc. Then the examination turns to the ways in which the standard allows implementation, such as telecommunication transmission, magnetic media, optical media, etc. Following is a major subsection dealing with the major kinds of data types that are supported, such geometric and topological modes, semantic aspect, dimensional aspect, etc. Then it focuses on the kinds of design approaches supported, such as hierarchical, network, relational, object based, object oriented, etc, as well as whether the standard is designed to support update and query transactions. Sender and receiver relationships are then examined, as well as whether the standard incorporates other standards in itself, or conversely, whether it is incorporated into other standards.

3.3 Transfer Specification Method

The method by which the transfer is specified, and whether an informal (natural) and/or formal specification language is used is a rather interesting question. The structure of the standard document itself is examined in terms of whether it is composed of one or more free standing components, whether it includes descriptive information, subset or profile implementations, encoding rules and appendices, and whether it includes standard use technical terms and/or the standard includes a comprehensive set of definitions of technical terms itself. A full table of contents is requested, and many of the standards have provided several pages of a detailed table of contents. This gives the reader a very detailed insight into the scope and organization of the standard. Some standards provide more than ten pages of detailed table of contents pages.

3.4 Conceptual Data Schema

One can now begin to probe the more detailed technical characteristics of the standard. The first query examines whether the standard contains conceptual data models or schema. A subsequent question inquires whether the data models or schemas define things such as abstractions, semantics, structure, implementation, or other things. Conceptual model/schema diagrams are then requested. This section provides the first real detailed insight into the technical structure of the standard. This sets the stage for a deeper probing of further detail in the following sections.

3.5 Transfer Process

The structure and organization of an implementation of the transfer standard is discussed in this section. The process used to accomplish the data transfer is of great interest. The first section asks for the specifics of the implementation while the second section on self description refers to the degree to which information contained within the transfer file digitally describes the data and data structures being transferred in the file. The third requests a diagram of the transfer file of each implementation used in the standard. Currently no standard is fully self describing, but most standards provide at least some facility for transferring digital information that describes some of the data and data structures being transferred by the standard. This section requests specifics on the sort of header information being transferred by the standard.

3.6 Transfer Elements

These transfer elements are the most fundamental level of data elements being transferred by the standard being assessed. A list of spatial primitives is requested. Most standards contain a fairly rich set of 0-, 1-, and 2-D primitives and some contain a few direct 3-D primitives. For most, 3-D primitives will be developed in the future. In most cases 3-D or higher dimensional data can be transferred by decomposing it into the lower level dimensional primitives that are defined in the standard. The second section requests details on the aggregate spatial data types that are supported, some vector based, and some tessellation based. Most standards support these sorts of aggregate objects, but the number of such objects defined and supported varies widely. A similar query on nonspatial primitives is also asked which includes feature, attribute and relationship primitives. A further question probes the kinds of data structures supported, such as topological, object-oriented, hierarchical, relational, image, or hybrid structures. These are important clues to the depth and richness of the kinds of spatial data the standard can process.

Further questions then probe graphic elements, which many standards support to some degree, but generally not extensively because most standards are designed to be Virtual 3 deep structure transfer standards. Spatial referencing via various kinds of coordinate systems supported such as polar, cartesian, geographic, or geocentric is examined. Most standards support many of these sorts of coordinates, but not all do for reasons of simplicity. Height information is also examined to see how it can be referenced. Most standards can reference height as a 3-D triplet of coordinates. A more fundamental question deals with the kinds of datums that are supported. Most standards have a fairly extensive set of datums that are supported, most list 20 and some list more than 50. Finally, a query is made as to the number of projections supported. Some standards again have a rather extensive list, but others lean towards efficiency and economy. There is still a debate as to whether a large number of datums and projections is a real advantage, or something of a disadvantage.

3.7 Update Information

Many standards are able to implement update information sets such that
information can added, deleted, changed, or replaced, and whether it can be carried out for the entire dataset, or only a part of a dataset. This can be a very useful part of a transfer standard if the systems are designed to use it in that way. If so, one needs to know the assumptions and underlie the standard in terms of the data model, database, and previous transfers. Authorization or access information is also of interest here as well. It is clear that more standards development effort must be devoted towards this capability.

3.8 Query Information

Some of the standards support query transactions on data sets. This can be a rather useful function in some settings. Of interest is whether the standard supports query transactions for one, or several data sets at the same time, and whether the same can be done for data descriptions. Functionality is also of interest as to whether such operations can be conducted in logical, spatial or other modes, and whether these operations can be performed by spatial, nonspatial or other kinds of primitives. The resulting output is or interest as to whether it could be a whole dataset, part of a dataset, or some other form. It is clear that more standards development effort must be devoted towards this capability.

3.9 Quality Information

This section on data quality information is one that most people would easily recognize as being part of a transfer process. The first item concerns what kinds of quality elements can be described, lineage, positional accuracy, attribute accuracy, logical consistency, completeness, currency, and other items, and whether these elements are described with the transfer standard itself. Attention then turns to just how this information is structured in terms of codes, structured text, unstructured text, referenced to external standards, or by some other method. The final item deals with whether the quality information can be specified in terms of spatial primitives, spatial objects, composite objects, whole files, specified geographic areas, or by other means.

This information provides the scientific and technical basis for making decisive comparisons. The Commission is currently finishing a book that utilize these characteristics to conduct assessments of each of the more than 20 national and international database transfer standards throughout the world (Moellering and Hogan, 1995).

5 References

Moellering, H., 1980, "Strategies of Real-time Cartography", Cartographic Journal, 17(1) pp. 12 - 15.

_______________, 1983"Designing Interactive Cartographic Systems Using the Concepts of Real and Virtual Maps," in Wellar (Ed.), Vol. II Proceedings of the Sixth International Symposium on Automated Cartography, pp. 53-64.

_______________, 1984, "Real Maps, Virtual Maps, and Interactive Cartography", in Gaile and Wilmott (eds.), Spatial Statistics and Models, Dordrecht: Reidel Publishing Co., pp. 109 - 131.

_______________, 1987, "Understanding Modern Cartography Using the Concepts of Real and Virtual Maps", Proc. XIII International Cartographic Conference, Morelia, Mexico, forthcoming.

_______________, 1991, Spatial Database Transfer Standards: Current International Status, editor, London: Elsevier Applied Science, 235 pp. + index.

Moellering, H. & K. Wortman, Eds., 1994, Technical Characteristics for Assessing Standards for the Transfer of Spatial Data and Brief International Descriptions, ICA Technical Report, ICA Commission on Standards for the Transfer of Spatial Data, 98 pp. + vi.

Moellering, H., and R. Hogan, Eds., 1995, Spatial Database Transfer Standards: Characteristics for Assessing Standards and Full Descriptions of the National and International Standards in the World, ICA Commission on Standards for the Transfer of Spatial Data, in press.

Nyerges, T., 1980a, Modeling the Structure of Cartographic Information for Query Processing, unpublished Ph. D. dissertation, Dept. of Geography, Ohio State University, 203 pp.

____________, 1980b, "Representing Spatial Properties in Cartographic Data Bases", Proc. ACSM Spring Meetings, St. Louis, pp. 29 - 41.

____________, 1991, "Analytical Map Use", Cartography and Geographic Information Systems, 18(1), pp. 11 - 22.