Data Standards In The Context Of National Geoinformation Infrastructure

Durgendra Man Kayastha
Project Chief,
Saptakoshi-Sunkoshi Topographical Mapping Project

Abstract

Development of GIS related activities and acquisition of data on spatial variables using remotely sensed data as well as maps have provided new insights in the field of decision-making concerning development planning and administration of natural resources [Kayastha 1999].

With the availability of software tools collection of geographic data for a particular application has become so easy that each institution can generate data on its own. Such an approach would lead to confusion among users, owing to differences in data format, reliability of data and most of all consistency of data could not be maintained due to redundancy [Kayastha 1999].

The paper outlines importance of geoinformation infrastructure and its components, followed by outlining areas where standards can be defined on basic elements of geospatial data such as geo-referencing, data definition, metadata etc. The paper also provided an overview of NTDB.

1. Introduction

Geospatial data in the digital form are becoming an integral part of the general decision making processes relating to land and its environment. Due to diversity of users and availability of software products, creation of geospatial data has become a common practice. Depending on the requirement several forms of database may become available differing in format, content as well as resolution. Variations that may be inherent in different data sets make them less effective. Considering the amount of input required in terms of effort in creating a database, it will be necessary to see that such a product (data) is useful and available to other users as well. If data could be developed on a uniform basis, it would be very useful and would considerably lessen the duplication efforts, which at present is seemingly not common. In the context of national geoinformation infrastructure, such commonality and uniformity may be achieved by following certain standards on geospatial data.

2. Geoinformation Infrastructures

Information infrastructures have become essential element of the development of any country. In order to obtain suitable information to support decision-making concerning land and its environment, geospatial data forms the backbone of all resource data. Availability of appropriate data and access to information resources can be ascertained with the establishment of geoinformation infrastructure. However successful placement of full functioning geoinformation infrastructure is contingent upon the national policy concerning data creation and data sharing.

Traditionally central authority produces paper maps and such maps are available to the user community on demand. The users then can incorporate additional information over the base maps to derive several different products as required. In this way, the basic meaning or the semantics of the data being portrayed in a map is transferred to the subsequent users automatically and is implicitly being maintained to a greater extent. So much so that the projections systems used for the base data generation is also evident. In the digital environment, these processes require special attention and can be considerably improved.

The requirements from the users' point of view have changed in the digital environment. Lack of consistency among data from different sources, scalability, outdated data, projection systems, variations in data semantics, and data resolution are some of the problems that hinder integrated applications. Taking an integrated view on the data model and creating and following certain standards in the core processes of data modelling and implementation

may help in minimizing the efforts in bringing in dissimilar data sets to a common usable format. This calls for development of National Geospatial Data Standards.

2.1 Components of geoinformation infrastructure

A geoinformation infrastructure is an organized collection of several essential components, some of which are:

- Basic spatial data
- Network and transfer mechanism
- Standards
- Metadata
- National Policy and supporting legal backup

2.1.1 Basic Spatial data

Geographic Information Systems operate on geospatial data sets. Geospatial data vary greatly in geographic area, purpose and content. The need for such data nearly always includes a few basic themes such as road network, streams and rivers, elevations etc. These provide a framework for data collection and analysis. Framework data provides a basis on which organizations can accurately register and compile other themes of data. They provide the geospatial foundation with which an organization can perform analysis and to which attribute information can be attached. Availability of such data sets is a key component.

2.1.2 Network and transfer mechanism

Geospatial data residing in several nodes need transferring. Without the mechanism of data transfer the basic philosophy of geoinformation infrastructure will be defeated. It is therefore necessary that transfer mechanism be efficient and transparent as far as possible for which a network can play a vital role. One single institution may not necessarily maintain geospatial data encompassing all data types. The mechanism of transferring geospatial data between different nodes must be so that the user may not have any difficulty in finding out what type of data is available in a particular node. Besides, there will be uniformity in data sets as well, as the data sets will be coherently integrated, because specific data sets will reside only in a particular node specific to that domain.

For instance, Survey Department may maintain basic topographic data sets that are commonly required for all other applications, while the domain specific data sets may be maintained by respective institutions. Such an arrangement however will entail massive data transfers between several nodes; the transfer standards should, therefore, be in place.

2.1.3 Standards

Standards essentially deal with basic specifications to be followed before a particular data element is accepted into a database. Standards therefore provide the user a basis to understand and evaluate the usability of data for intended application, while it provides a basis to the data producer to populate the database with data in a consistent manner.

Considering the practice of converting paper maps into digital geospatial data by virtually any organization, there is a strong need for standards in digital geospatial data to maintain transferability, and compatibility of data sets generated from diverse data sources. Standards in case of paper maps are long in existence normally set by national mapping agencies, however, in case of digital geospatial data; acceptable standards need to be worked out with the participation of possible stakeholders.

2.1.4 Metadata

To find existing maps, files and other data to use with the LIS/GIS or other technology often becomes a frustrating experience. In many a case the existing data is not discovered until it is too late to use for a project, or the data that is discovered is not described in sufficient detail for determining its usefulness. Metadata helps an individual to locate and understand data.

In the analogue domain, where geospatial data is presented in as printed maps, metadata is readily available as part of the map in the form of legend and other marginal information. Metadata is thus, easily transferred between map producers and map users. The same geospatial information, when presented in digital form, metadata becomes equally important a part as the geospatial data itself. Without which the users may not be able to derive the actual information contained in the data. The data producers, therefore, will have to put extra efforts in development and maintenance of metadata and it applies to the subsequent users who may modify the data to suit

their particular needs. Thus metadata serves several important purposes including data browsing, data transfer, and data documentation.

Considering the usability of existing dataset, metadata could be maintained at several levels of complexities. In the basic form, metadata might consist of a simple listing describing basic information about available data, whereas, in other case, detailed information may be included about individual dataset.

2.1.5 National Policy and supporting legislative backup

In most cases digital geospatial data are created from paper prints. It has created few more problems such as duplication of data sets, inconsistency due to semantic differences, incompatibility of data sets due to different georeferencing, resolution and accuracy of data sets etc. Accessibility and copyrights are other problems that are very vital in the success of geoinformation infrastructure.

Many of the cited problems can be solved under a framework of national policy on digital geospatial data handling. Some may be solved through adoption of a national standard such as use of a national coordinate system for georeferencing.

3. Spatial Data Standards

Effective geospatial data standards are necessary for any efforts in collaborative data production. Standards will increase the ability to share geospatial data and preserve its original meaning, to create more complex applications, and to stimulate the business associated with GIS technology and geospatial data modelling.

Basically spatial data standards should cover among others the following key elements:

- Data definition and content specifications
- Positional accuracy /Resolution
- Attribute accuracy
- Data accessibility and exchange
- Georeferencing

A comprehensive treatment about the standards can be obtained from the series of reports and standards released by ISO/TC211. ISO/TC211 is an international body actively engaged in formulation of international standards in the field of geoinformation to support understanding and usage of geoinformation, improve accessibility and to establish geoinformation infrastructures at different levels.

3.1 Data definition and content specifications

Geospatial data, as we all know, are composed of two basic components viz. the location and the attribute. The location implies the position, orientation and extent of a particular feature in the geometric space, while the attribute describes the identity of the feature as to what it represents in the real world.

Information that is implicitly available in a printed map, need to be explicitly represented, in case of digital geospatial data. Depending on the user views and applications intended, a feature might be represented in many ways or not at all represented, which will make meaningful interpretation of geospatial data more difficult to others. In order to make a uniform representation these data need to be well defined in terms of semantics and the geometric representation.

Identification of data elements and the way of their representations may be obtained through the process of data modelling. All data items will then be explicitly defined. The collection gives us data content specifications. It should however include the relationships being modelled.

3.2 Positional accuracy/Resolution

Accuracy implies closeness of an entity to reality both in position as well as attribute value. A particular data set that may be good enough for one application may not be so for other. It is therefore necessary to lay down the accuracy requirement before the data collection itself. In the digital domain most of us have a tendency of overlooking the accuracy component due to the possibility of enhancing the data through several routines. But the end result may be completely different from what may have been expected through the use of a certain data set intended for a different kind of application. For instance a DTM created using 100-metre contours or a district area created from a regional scale map might not be suitable for large-scale applications, but one may be tempted to using such data in the case of

unavailability of high-resolution data or for some other reasons. It applies to other types of data as well. Accuracy therefore plays a vital role in the data manipulation. Similarly in case of raster data one must be very careful about the data resolution. While formulating a data standard it is necessary that the accuracy/resolution level of data elements should also be specified. One should also note that a multi resolution database might be necessary to accommodate multitude of users.

3.3 Attribute accuracy

Attributes are the means to distinguish features from one another for instance distinguishing one location from other or one road from other. Attribute is a fact about some feature. Some of the attributes could be simply names or the land use class but to set the domain value for any attribute some sort of measurements would have to be made. With measurements the uncertainty creeps in depending on the type of measurements. Even a name attribute could introduce errors owing to simple input error such as text or due to uncertainty in toponymy or the transliteration rules; name of a place might be wrongly spelt. In cases like land use class there could be cases when assigning a class value entail confusion. One should therefore take into account all the aspects affecting the definition of attribute type and attribute value domain.

3.4 Data exchange format

The availability of different software products and the diversity in users' views necessitated that data created and maintained in one system should be transparently transferable to any other system. There are variations in operating platforms and the software used which cannot be regulated. The users have their own choices and preferences, yet geospatial data should be available to them to work with. In addition we know that geospatial data pertaining to specific domain may have to be created by respective agencies; such data should be shared also among the data producers as well. This requires a system independent exchange mechanism, whereby data transfer from any system could be made. Several exchange formats have been developed and being used in the world.

In our case too, data exchange format for geospatial data should be developed. It might not be necessary to start from scratch. In fact we can learn from the experiences of others and evolve a reasonable kind of transfer format for our use. In any case a comprehensive effort is still required on the part of data producers.

3.5 Georeferencing

One important component of a geospatial data element, as stated elsewhere, is position. These days it is possible to transform the coordinate system from one system to the other very easily yet considering the fact that all transformations entail introduction of errors, it will be necessary to have a common referencing system to maintain compatibility of data sets. The reference frame for the position therefore should invariably be the national coordinate system. All data developers should stick to that base which will provide a uniform basis for positioning of geospatial entities.

4. Overview of National Topographic Database (NTDB)

In an effort to provide for a uniform basis at least in the topographic data domain, Survey Department has prepared detailed specifications for National Topographic Database. The basic objective of such a database system is to provide a homogenous seamless national data on topographic features present in reality. Following is an overview of the NTDB.

There are in all nine classes by themes in the NTDB viz. Transportation, Building, Topography, Land Cover, Hydrography, Utilities, Administrative Area, Designated Area, and Control Point. Each feature class describes the collection of data by a distinct theme. Within each class, successive subdivisions are made to incorporate feature classification hierarchy down to the feature component. The specification is primarily based on the topographical maps published by the Survey Department. [NTDB 1998]

In addition to feature names numeric codes are used for referencing, as feature names are not always convenient to use for referencing during manipulations. A five-digit code for feature component is developed. The specifications also include attribute catalogue containing a list of possible attributes designated by an attribute type name and an attribute code.

The NTDB at present state may not be fully populated with the help of topographic maps alone. So it is recommended that additional data may be maintained in

the respective line agencies by replicating the basic data. [Kayastha 1999]

Work has already started in the Survey Department towards the implementation of the NTDB. With time, it is expected that the specifications will be modified to accommodate the findings during the implementation process.

5. Conclusion

The utilization of digital geographic information has been growing fast in all sectors of the society. Lack of basic digital geospatial data has resulted into multiple, inhomogeneous and expensive data collection and conversion in order to ensure the needed geospatial data for different projects and activities both in public administration and private sector. This kind of uncontrolled creation of geospatial data for different purposes will lead to serious consequences when trying to combine data from different sources. It is therefore important to start work to stop this frustrating and expensive development. [NTDB 1998]

The NTDB could provide a starting point towards the standardization, however it is very important that a consensus among all stakeholders be reached. This will provide for an environment where serious deliberations could take place toward achieving the goal of geoinformation for all.

Acknowledgement

The author would like to express sincere thanks to Prof. Bengt Rystedt for critical review and providing comments on this paper.

References:

Guptill, S.C and Morrison, J.L (eds.) 1995: Elements of Spatial Data Quality, ICA Publication. ISBN 0-08-042432-5.

Kayastha, D M 1999: National Topographic Database (NTDB): Seminar on "Future Plan and Programs on National Surveying and Mapping", Jan 27-28, 1999, Survey Department.

NTDB 1998: Specifications for Geographic Information Service and National Topographic Database; Survey Department.