

DATA STANDARDIZATION OF THE INTEGRATED LIS IN FINLAND

Matti Vahala
National Board of Survey in Finland
PL 84
SF-00521 Helsinki 52
FINLAND

ABSTRACT

The main idea of establishing an integrated national land information system is to use the location of objects as a link in combining various sets of land related information. There is no intention to establish any centralized system. The integrated LIS will consist of individual land information systems, and all the individual administrative units will still be responsible for the information under their authority. The integration of data exchange will be achieved by developing classification standards, both logical and physical representation of information, as well as the data transfer involved in combining various sets of land related information. Standardization has reached the stage where standards for geometric elements and classification of the spatial accuracy of objects are ready to be documented.

THE ORIGIN OF DECENTRALIZED INFORMATION SYSTEMS

In Finland, as in other countries, the development of data technology has resulted in many existing registers being transferred to computers. Registers consisting of spatial data have also been the object of similar developments. In registers consisting of spatial data, identification of entities is made with the aid of identifiers based on the rule of thumb. The location is given by map sheet number or, in the best cases, using a reference point with coordinates.

Large national systems now in the operative stage are information systems for buildings and population. Reference points expressed in the form of coordinates are used in these systems. The number of information systems that are in the development or implementation stage has increased. The most important of these at the national level are:

- real estate information system
- forest information system
- environmental information system
- information system for water and waterways
- digital terrain models
- digital map systems

Many of the above systems are designed to serve one application or administrative unit.

The end users collect their data from several systems. Since there is no common reference system, the integration of digital spatial data is difficult or even impossible. Problems in the integration of spatial data also means that the redundancy of data is common. This leads to the problems of updating. The problems are severe, and in different cases adjustments have been made in order to find solutions to the integration problem.

SEARCH FOR SOLUTIONS TO INTEGRATED LAND INFORMATION SYSTEMS

System development in spatial data-processing has been carried out at the National Board of Survey for over ten years. The main goal at the beginning of this activity was to stimulate the automation of graphic map production. As the work proceeded, however, we discovered new ways of utilizing information systems in spatial data-processing. It also became evident that location could be used as a link between different data sets. Development activity has been directed at solving the problems associated with the management of spatial data at the level needed in information system applications.

When no suitable solution was found to the integration of information systems, the advisory board for mapping activities proposed the establishment of a group including representatives from different administration units. The group's task was to determine the need for and possibilities of creating integrated land information systems. Work was started in 1983 and completed in 1984. The work included:

- drawing up a definition of integrated LIS
- processing information about the need qualification of integrated LIS
- investigating the types of classification used in existing registers and the need for reclassification
- reviewing the methods used in the determination of location, and user's accuracy criteria
- assessing the demand for registers used in integration
- making proposals for the development of the integrated LIS and how the task should be organized

Based on these recommendations, the Ministry of Agriculture and Forestry set up a project, the LIS-project, at the beginning of 1985. The task of the project was to develop standards for classifying and representing the data, as well as to make recommendations about how to remove obstacles. The standards needed to operate in a batch mode environment will be ready by 1987. Those needed in on-line operation, such as query language, will be ready two years later.

DESCRIPTION OF THE SOLUTION FOR INTEGRATED LIS

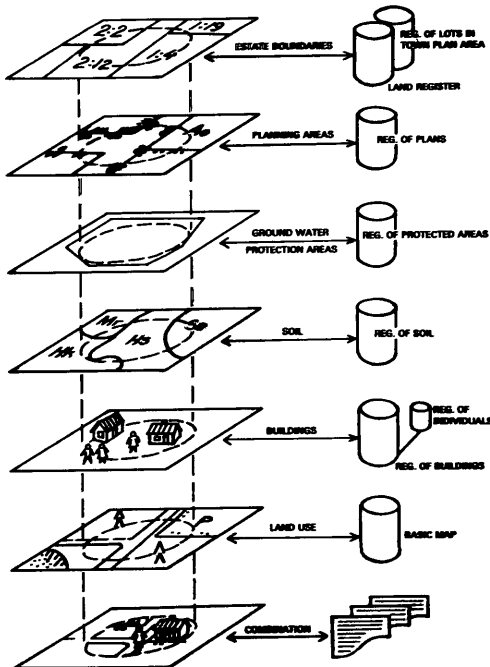
The problem of integration between systems can be solved in the following manner:

- integration is based on decentralized registers
- the link between objects is location
- data coming within the sphere of integration includes all systems able to present their spatial data in standard form; these include systems outside normal map presentation, e.g. population
- representation of data is standardized in the data exchange

The solution has the following advantages:

- the systems are independent of each other
 - . data contents and identifiers
 - . updating can be made without effecting other registers
 - . the technical solutions used in of different systems can vary, and they can be restructured without having repercussions in other systems
- the number of systems is unlimited; new systems can be integrated without altering others, the only condition being the ability to represent spatial data in standard form
- data redundancy can be minimized

Data producers are natural data controllers. Therefore a register can be and in fact also are decentralized administratively as well as regionally. The picture below illustrates the solution.



THE LIS PROJECT AND STANDARDIZING

The task of the LIS project is to develop standards for data classification and representation, as well as to make proposals about how to eliminate restrictions on the scope of activities. The work is done in groups, and the aims are given below.

Geometric entity types

All objects in integrated LIS data transfer have to be described as geometric units. This is a necessity when the location of objects is used as the link. As the definition of location also has an effect on other standards and their structure, this standard must be available to others in good time. The proposed standard will be described in detail in the next chapter.

Location and location accuracy

A common coordinate system, reference system, is needed in integration. This is the system where all data is represented when integrating data from different systems.

When location is used as a link, the accuracy of location will determine the success of linking. Measurements of position are made using different techniques and hence information about uncertainties of location is needed when data sets have to be integrated.

Semantic classification

In order to depict and interpret data, information is needed about classification. The classification gives data character. The integration is useless if the contents of the data are not understood. When standardizing existing classifications, any differences between them must be cleared up.

The work has been continued for over a year and it has become obvious that it is unrealistic to try to develop a national semantic classification which could be used in all registers. Many registers and applications have their own semantic classifications for historical and practical reasons. These will be retained and solutions must be found from other sources. Semantic classification in different applications is also varying.

A classification directory is based on one standardized framework. Other classifications are determined by defining the relations between entities in the framework classification and the classification in question.

Data transfer

The first aim is to develop a standard in order to be able to transfer data in a batch environment. Data will be transferred primarily on magnetic tapes and by remote access. This level does not satisfy the needs of on-line data service.

Query language

An effective integration of LIS requires a capacity for on-line queries. This means that a common language is

needed. The language must also be user friendly. A preliminary investigation is to be made in other groups in order to be sure that the solutions will also work in on-line. The standardization work will not start before 1987 and will continue for about two years.

Description of existing land information systems

Users need information about existing registers and methods must be available for describing the physical and logical structure of the systems. In order to serve users there must be meta data about land information systems; the classification used and how the information can be utilized.

When development of the query language became a part of the LIS project's activity, an idea was born. The classification framework linked to the classifications used in different land information systems can act as part of the user interface in an integrated systems network. A data dictionary can be joined to the same user interface. With the aid of the interface users are informed of registers of interest and what kind of semantic classification has been applied. The user interface is a necessity in on-line processing, but it can also act as an effective tool already before that stage is reached.

Juridicial and administrative problems

The solution to integration of land information systems cannot be a mere technical one if functionality is to be achieved. The problems concerning licences, price of data and other obstacles have to be taken into consideration. Coordination of the system must be organized in order to avoid redundancy or lack of information.

DEMONSTRATION OF THE DRAFT STANDARD FOR GEOMETRIC ENTITY TYPES

Data transfer of digital spatial data gives a clear starting point. The aim is to transfer data not data structures. Information systems updating spatial data have different technical solutions. The standards for integrated land information systems define only the logical and physical implementation of data transfer.

The proposal defines the geometric entity types used in integration. In data transfer all data is based on geometric entity types. Attributive data can be joined to geometrically defined entities.

Depending on the structure of the selected integration model, the geometric entity types are the basis for whole solution. All data in data exchange have to be defined geometrically and represented as stated in the standards.

Geometric representation can be done using different methods: vector geometry, pixel geometry or operations with entities.

Representation by vector geometry

Entity types in vector geometry are in 1-3 dimensional space.

Entity types in one dimension are
Point, which is a real number

Subinterval, which is a bounded, closed subinterval of reals.

Entity types in two dimensions are
Point which is given by two coordinates

Standardwise **curves** are continuous piecewise interpolated functions. They are given by control points and interpolation methods in accordance with the standard.

Interpolation methods available are:

- 1) linear splines
- 2) Hermite's two point interpolation
- 3) interpolating cubic B-splines
- 4) circle method
- 5) average tangent method

The order of control points induces the natural order of curve which is orientation of the curve at the same time.

The boundary of a **region** is defined as a disjoint union of closed, standardlike curves.

Entity types in three dimensions

Point is given by three coordinates

Curve is defined as curves in two dimensions. Interpolation methods are:

- 1) linear splines
- 2) Hermite's two point interpolation method
- 3) interpolating B-splines

Surfaces can be defined on the plane in subtriangles as R^3 valued maps which are either interpolated linear maps or have been obtained from a rectangular control point lattice by means of base spline functions. Standardwise base functions are:

- 1) interpolating splines
- 2) base functions for Hermite's two point interpolation method
- 3) B-spline base functions

Solid body is a bounded, connected, closed subset of three dimensional space. The faces are determined with a finite number of surfaces compatible with the standard. There can be "bubbles", as in Swiss cheese, in a solid body. The border surfaces are oriented in the data representation such that their edges are listed clockwise, as viewed from outside the block.

Pixel geometry representation

Entity types in pixel geometry are observation lattice and observation element. The system of coordinates is rectangular. The objects can be attached using observation vector. A constant length observation-vector, which contains the same attributes at constant places is attached to every observation element.

Use of operations in representation

New geometric entities can be formed with the aid of existing geometric entities using standardized operations. Operations are mainly intended for forming the sets and entities in three dimensions. The arguments of operations are identifiers and the new entity caused by an operation has always an identifier.

Unary operations are linear functions of R^N .

Binary operations are:

- 1) Boolean set operations
- 2) Geometric, regular, Boolean set operations
- 3) Cartesian product
- 4) Vector product
- 5) Join with entities in 2 dim.
- 6) Rotation (a line or area in a plane rotates around a straight line in space)
- 7) Sweep (a line or an area in 2 dim. moves along a line in 2 or 3 dim.)

Propable standard

Proposing a standard for geometric entity types is problematic because it permits versatile representations of data and therefore it is like an antistandard. This is the dominant factor in many interpolation methods especially. The standard to be confirmed may include only some of the interpolation methods in the first stage, but later on others will be added to the standard if needed in practice.

STANDARD FOR LOCATION AND LOCATION ACCURACY

National mapping activity in Finland uses the national grid coordinate system (kkj) for horizontal coordinates. Some towns and municipalities have their own coordinate systems for historical reasons. Data acquisition is based in these areas on local systems.

Vertical coordinates are given according to the N60 system in national mapping work. Towns use their own vertical coordinate systems which are much more numerous than those in the plane. The reason for this is that the ground is rising continuously in Finland. National systems like N60 must be replaced with new systems in the future.

In order to be able to integrate different systems the location of the objects must be given in the same coordinate system. Since integration presupposes a common reference system (kkj + N60) a transformation must be made from local systems. When such transformations are done,

the accuracy of data in the common reference systems varies from that in the local systems.

We have no standard for estimating the accuracy or uncertainty of position. The existing technical instructions for mapping describe methods to be used when mapping is done to a given scale. These instructions do not tell about the accuracy reached, but only how to perform the measurement.

The usage of location is essential in the integration of spatial data. Since the accuracy of objects in different systems can vary over a wide range, and determination of location is done by many different methods, integration is only successful if the mutual accuracy of the different objects is known. This enables us to compute the available coordinates using the right methods.

Data used in integration must be known in location and its accuracy, both given in a common reference system. The term **uncertainty** gives a better idea than accuracy in this connection. Uncertainty is defined to be 2.5 times the standard error. As far as horizontal uncertainty is concerned, uncertainty is computed from the mean square error of point determination in position.

For technical reasons, the measurement of elevation has been made a ten-fold more accurate than horizontal coordinates. Therefore in classification, accuracy of elevation is given using its own class. Because of land uplift, national elevation systems must be renewed every few decades. For the same reason there are more local elevation coordinate systems than horizontal ones in different towns.

The standard of accuracy

The accuracy of a point is formed for both horizontal and vertical coordinates of two components:

- the uncertainty in the reference system
- the uncertainty in the local coordinate system

The reference systems are: kkj for horizontal, and N60 for vertical coordinates. The group of local systems includes only systems which are based on geodetic fix points of the third order (this classification of geodetic points is based on instructions for large scale mapping). Transformation coefficients are determined for these systems, as well as uncertainty which is caused by transformation. This is the maximum error of the position determined in relation to the fix points of the local system.

Principles of classification The accuracy class gives the uncertainty to the nearest fix points. The uncertainty of points in a reference system are as follows:

$$m_s^2 = m_c^2 + m_m^2 + m_i^2 \quad (1)$$

m_s uncertainty in the system (lemma: national system)
 m_c " of control points
 m_m " of measuring method
 m_i " of identification

When items are not determined using national systems (kkj, N60), they must be transformed to conform to these systems. The uncertainty in national systems is thus:

$$m_s^2 = m_l^2 + m_t^2 \quad (2)$$

m_s uncertainty in national system
 m_l uncertainty of local system
 m_t uncertainty of transformation

Horizontal accuracy classes

identification	error of position	uncertainty
2-2	8 mm	20 mm
5-2	20 mm	50 mm
1-1	40 mm	100 mm
2-1	80 mm	200 mm
5-1	200 mm	500 mm
1+0	400 mm	1 m
2+0	800 mm	2 m
5+0	2 m	5 m
1+1	4 m	10 m
.....

Vertical accuracy classes

identification	error of position	uncertainty
2-3	0.8 mm	2 mm
5-3	2 mm	5 mm
1-2	4 mm	10 mm
2-2	8 mm	20 mm
5-2	20 mm	50 mm
1-1	40 mm	100 mm
.....

Application of the classification for other entity types

The classification demonstrated here is suitable for points as such. The accuracy of a curve is usually given by one class common to all points along the curve. When the object is defined as a curve, an arbitrary point along the curve will naturally have the same accuracy class as the curve. The class of the curve gives a "zone" around the curve.

The points along the curve can have an individual meaning, too. In this case accuracy can be given point by point, and the curve itself can have no accuracy class. The uncertainty between known points is now determined by computing the weighted mean from the uncertainties of two adjacent points on the curve.

The uncertainty of **region** is restored to uncertainty of edge (curve). In the classification sense, the class of a region can be considered to give the accuracy to all curves surrounding the region in question in a similar way as the accuracy class for a curve gives the same point accuracy.

The accuracy class can be obtained for large quantities of data by forming sets.

The accuracy class for **surface** and **solid body** are returned points and surfaces as a region is returned to curves and points. Derived quantities e.g. length and area are not classified. The accuracy of these items can be derived, however, from location accuracy.

Data in pixel form can be classified by giving the accuracy of the corner points of the lattice. In addition, an individual accuracy class of the object can be added to the observation element.

CONCLUSION

A solution for integrated LIS is described, and the first proposals for standards are now ready. The approach based on decentralized registers and the use of location as the main link between objects is chosen without an external example. It will be interesting to see the outcome when the first trials are tested with real data and in production applications.

The number of standards usually indicates the maturity of a particular sector. The solution for integrated LIS will now have the first standards in Finland, and it is obvious how this really shows the stage of maturity of this sector.