mmm.fi

Improving operational efficiency with geographic information

Practical guidelines for the harmonisation of core geographic datasets

12a2006

Improving operational efficiency with geographic information

Practical guidelines for the harmonisation of core geographic datasets

Ministry of Agriculture and Forestry 12a/2006

Publication:

Improving operational efficiency with geographic information
Practical guidelines for the harmonisation of core geographic datasets

Author:
Ministry of Agriculture and Forestry

Photos:
ESA-J.HUART (Galileo) TEKES, Maanmittauslaitos, Merenkulkulaitos, MMM kuva-arkisto

ISBN 952-453-301-4
ISSN 1238-2531

Graphic Design: Z Design Oy

Layout: Vammalan Kirjapaino Oy

Printing Press: Vammalan Kirjapaino Oy, 2006

Contents

SECTION I Basis

1. Introduction

These guidelines are intended for the enhancement of public administration and private sector productivity and operational planning using core geographic datasets. Compatible core geographic datasets are vital in planning locations for social, health care and civic services. Security services for the society are more dependent on the use of geographic information, thus emphasising the quality and interoperability of geographic information used. An example of such security services are emergency response centres, whose operations require the use of reliable and up-to-date geographic information. At the European level, environmental impacts have risen to play a key role in operational planning and building. An example of this is the currently ongoing preparations of the INSPIRE geographic information directive, whose primary objective is to make all geographic information that has or could have an impact on the environment fully available to public administrations.

The national geographic information strategy has set targets for dataset interoperability, commonality and harmonisation. In this context harmonisation refers to all measures which are taken to improve the commonality and interoperability of datasets and information systems and the integration of operational processes.

These guidelines address the requirements set for the harmonisation of geographic information. These requirements apply to modelling methods, data transfer, data content, process management and quality. Interoperability assessment levels are examined from a global, European, national, regional and local point of view.

The guidelines can be used in the planning and decision-making concerning the development of organisational productivity for public administrations, the planning of operational processes, and evaluation of possibilities for the utilisation of geographic information. **The operations of core geographic information producers should be based on networked production and maintenance processes as well as facilitate the use of common geographic information through information service networks.** The guidelines are divided into three main sections. In the first section the benefits of harmonisation for society and users of geographic information are outlined. In the second section technical specifications are defined. The third section deals with implementation.

The guidelines were drafted by the Finnish Council for Geographic Information Harmonisation Division task force T1. The task force chairman was Chief engineer, D.Sc. (Tech.) Antti Jakobsson, of the National Land Survey of Finland, and the task force members were M.Sc. Antti Saarikoski of the National Land Survey of Finland, Specialist researcher, Lic.Sc. (Tech.). Lassi Lehto of the Finnish Geodetic Institute, and Land-use engineer, M.Sc. Matti Holopainen of the Association of Finnish Local and Regional Authorities.

1.1 National geographic information strategy objectives for harmonisation

The national geographic information strategy has set the following objectives for the interoperability and harmonisation of geographic datasets:

> "A prerequisite for the effective use of geographic information at all levels of society is that the datasets are comprehensive and mutually interoperable, possess integrity in logical and technical terms and, above all, are readily available. To achieve interoperability common standards and recommendations need to be applied to the maintenance and management of geographic datasets in widespread use."

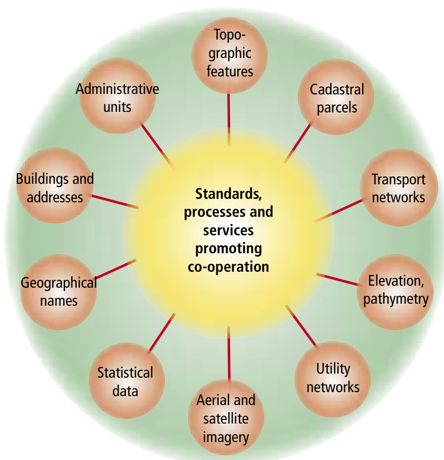

Figure 1.1 View of the core geographic datasets as a system[1]

[1] National Geographic Information Strategy

”Core geographic datasets are widely used both individually and combined on the local, national and international level. For these reasons the datasets need to be well managed and generally available, and form a harmonised, integral whole, providing nationwide coverage and facilitating shared use.”

Harmonisation involves the following steps:

“Step H1: The efforts to harmonise datasets with the intention of enabling efficient data sharing shall first be directed to core datasets (i.e. basic geographic information) which are most widely used, and to the development of the maintenance. **The permanent co-operative body shall ensure that basic geographic data are defined and that the requirements to be fulfilled by these are determined by the end of 2005.** The modelling methods used for core geographic datasets shall be consistent with the common standards and **the data shall be sufficiently interoperable in terms of quality and content.**”

“Step H3: Administrative organisations maintaining core datasets shall introduce the individual identification of objects in cases where there is a continuous transfer of changed data from one organisation or application to another. This shall also apply to generalised datasets.”

“Step H4: Administrative bodies maintaining core datasets shall ensure that quality monitoring during production and quality assurance of finished material are carried out and that an auditing capability is maintained. The results on quality shall also be included in the metadata and product data.”

2. Targeted outcome

The outcome targeted by the geographic information strategy:
"Public authorities make extensive use of databases and services maintained by other public authorities, as the data information networks employed by the public administration are effective, a transparent and fair pricing policy is applied and great care is paid by the respective agencies to ensuring the quality and availability of the resources and services."

2.1 Benefits of harmonisation

2.1.1 Integration and improvement of production processes

In accordance with the targeted outcome of the geographic information strategy public authorities make extensive use of databases and services maintained by other public authorities. This principle contains the fact that **in an efficient society it is not sensible to compile the same or similar geographic information at the behest of several different organisations using several data acquisition methods and processes**. If at all possible, the information should be compiled once, so that it can be used by all governmental authorities requiring it and, based on the information, services required by citizens can be created. This requires that governmental authorities provide integrated production processes. Data specifications, data quality requirements and data acquisition processes must be agreed upon by all concerned parties to meet all their needs. It is also possible that organisations will be forced to make compromises regarding data structure, data quality and data perfection.

Data acquisition should be organised in such a way that eliminates overlapping data acquisition. Data acquired once in accordance with uniform specifications and the elimination of overlapping data acquisition might require governmental authorities to establish a mutual agreement concerning the data acquisition carried out by each organisation, even if the organisation in question does not necessarily require all the data for its own purposes. Even if a comprehensive agreement on the division of data acquisition tasks was not made, all organisation needs would have to be taken into consideration in modelling, so that another organisation could link its own attribute data to the acquired data. Parties to such an agreement would also have to forge agreements concerning the regional orientation and timing of data acquisition so that the data is available at right point in time. Integrated data acquisition processes require that all involved parties have a uniform perception of data structure, quality and process operation as well as the opportunity to audit process operation, conduct quality evaluations or otherwise ensure the reliability of the end data acquisition result. Quality evaluation should be done independently, e.g. by a third party. Organisations should also agree on a complaint protocol for substandard quality as well as a dispute resolution protocol.

Generally, the most effective approach is to have the data producer maintain data in the database and make data available for use by other actors through the information service network. Geographic information production and maintenance processes are labour intensive. The effective organisation of data acquisition and data maintenance through integrated production processes **will result in considerable savings labour costs**.

2.1.2 Making core geographic datasets available for society and commercial applications

The body of core geographic datasets is comprised of **datasets from several public authorities and data producers**. The greatest benefit is obtained from the datasets when they can be combined with one another. Through harmonisation core geographic datasets can be combined in such a way that the data can be utilised. Using new standards, it is possible to produce services which advantage several datasets without the user necessarily knowing or needing to know that he is using multiple datasets. **In Finland 200[2] million euros is used every year for core geographic information, but investment in R&D and post-processing varies.** For example, geographic datasets produced by municipalities are often only used for their own or local needs, when the datasets could be used in the production of national datasets and national services. One of the obstacles to this is the fact that producer specifications do not correspond to one another.

[2] There is no researched data available for geographic information production costs.

2.1.3 Increasing productivity and decreasing costs

The productivity of the public sector is a key factor in national competitiveness. Core features of the Finnish social model are comprehensive and high-quality social, health care, educational and other civic services. A functional infrastructure and a covered bond system, which is based on a reliable real estate system are considered givens. Public services are guaranteed comprehensively for each and every citizen. Public services are primarily financed with tax revenues, which explain the high taxes in Finland. Society's main challenge in ensuring international competitiveness is reducing taxes. Improving the efficiency of public administration without compromising on quality has become a key objective.

Harmonisation can significantly increase productivity in the geographic information sector. Core geographic dataset production processes should be integrated so that they transcend producer organisation boundaries. If specifications are compatible, more accurate datasets can be used in the production of generalised datasets. The greatest benefits can be achieved through the elimination of overlapping data acquisition, thus resulting in labour cost savings in state agencies.

2.1.4 Legal basis

The pending INSPIRE directive indicates a need to enact legislation in Finland for the common use of geographic information. The roles of basic geographic information producers should be strengthened and defined. In Finland there are already numerous existing laws specifying the common use of geographic information. These include legislation on the road and street network information system (991/2003) and Act on the Land Information System and Related Information Service (453/2002) pursuant to the establishment of Digiroad (the National Road Data System).

2.2 Scenarios

The purpose of the following scenarios is to illustrate the benefits achieved during the different phases of harmonisation using a few basic geographic datasets. The examples are simplified, but are indeed based on real situations.

2.2.1 Common building data

Datasets: municipal building data, Population Information System building data, National Land Information Service, Topographic Data Base.

The objective of harmonising building data is to get municipalities, as the producers of the original building data, to model building features whilst taking into consideration the various use and data generalisation needs of the end user. The user would be able to select the parts of data and data models produced by municipalities that it needs for its own purposes. The goal is that the data would not need to be acquired, saved, remeasured or calculated for generalisation.

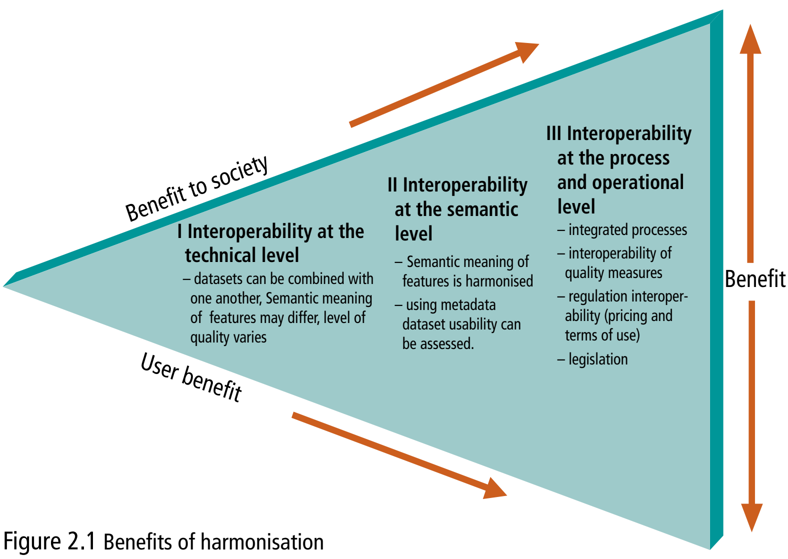

Figure 2.1 Benefits of harmonisation

Scenario challenges for harmonisation:

- The challenge is to agree on semantic content with different actors and the development of integrated processes
- Costs in achieving interoperability

Scenario benefits for producers and users:

- Producers save on data acquisition and maintenance costs
- Improvement in quality

Description of main scenario features:
During the building permit process, data on the existence of the building is generated through municipal agencies. In some municipalities building position data is plotted on a base map, but in sparsely populated areas the municipality might only have centerpoint co-ordinates.

Figure 2.2 New building completed approx. 2–3 years after receiving a building permit

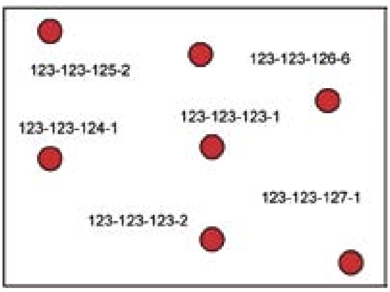

Figure 2.4 Population Information System building position data (simulated example)

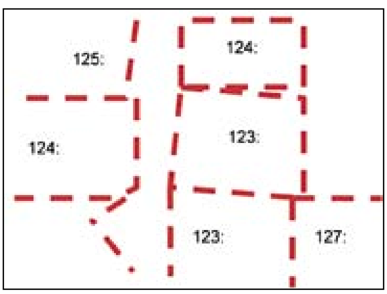

Figure 2.6 Land Information Service data

During the building permit process, data on the building centerpoint position is acquired for the national Population Information System. The centerpoint is listed in co-ordinates on the form, from which it is entered into the system manually. This process currently does not have a map user interface, which provides opportunity for error. In some cases the centerpoint might also differ from the centerpoint calculated on the base map.

The National Land Survey of Finland enters new buildings in the Topographic Data Base after a delay of a few years (1-10 years). In Figure 2.4 our example building has not yet been entered, as the regular interval for updates is 5-10 years.

The National Land Information Service does not contain building data. However, the building identifier is formed from the land identifier.

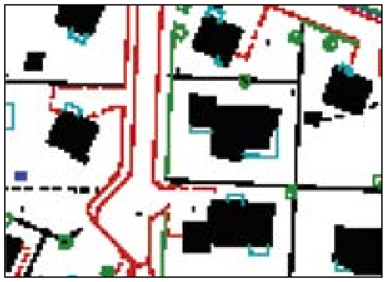

Figure 2.3 Building position data on the base map

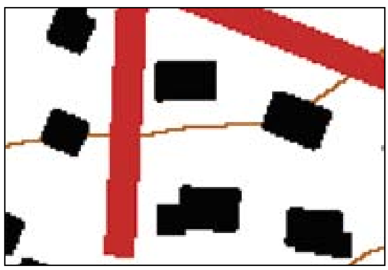

Figure 2.5 Topographic Data Base building position data

Subphase I: Realisation of technical interoperability

Building data can be combined from different registers. Figure 2.7 shows a possible situation, in which current register data is combined

The benefit of technical interoperability is that potential register quality problems can be detected better. (Figure 2.7?). Some of the quality problems can be solved by automatic routines, thus improving dataset quality. The problem is that connections between features are not recognised. The building identifier portion of the land identifier may differ from the actual land identifier. The building centerpoint might fall within the wrong plot. The building centerpoint might differ from the base map centerpoint. The Topographic Data Base building data might differ from the base map building data. The Topographic Data Base building might fall within the wrong plot.

Subphase II: Semantic interoperability

In this phase the link between the Population Information System building centerpoint, the base map and Topographic Data Base building is determined. The link can be made, for example, using an ID code, thus allowing individual buildings to be identified accurately. The benefit of integrating processes cannot be gained at this point.

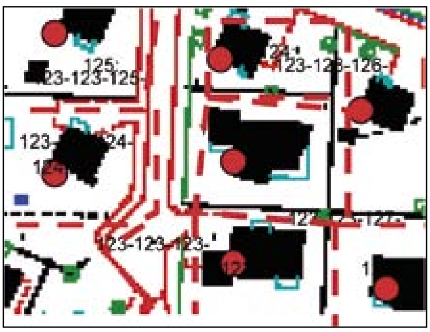

Figure 2.7 Subphase I possible situation (simulated)

Subphase III: Process and operational interoperability

Common features are exploited in the production of core geographic datasets. This requires the specification of quality requirements and regulations (such as pricing and terms of use). Society and users benefit from the elimination of overlapping data acquisition.

2.3.2 Common address data

Datasets: municipal building monitoring system, guide maps, Population Information System, Topographic Data Base, Digiroad, Genimap, Teleatlas, Navtech

Current status: The municipality decides on street names and building addresses. Municipalities submit building centerpoint co-ordinate address data to the Population Register Centre. Municipalities also produce guide maps. The National Land Survey of Finland acquires building point datasets from the Population Register Centre and records the road names and addresses in the Topographic Data Base as attribute data. The Finnish Road Administration's Digiroad obtains address data from the Topographic Data Base and these data are possibly supplemented by municipal data. Data is entered in user systems from the Topographic Data Base and Digiroad, and private data producers also have their own address datasets.

Problems: When a municipality decides on a street name, the data is not saved in the national system. When a municipality assigns a building (co-ordinate point) address data, the data is not linked to the street or road. Transferring data from a municipality to user systems (service chain) is complicated and time-consuming. Dataset consistency and reliability suffer due to the multiphase operating chain and several different operating cultures.

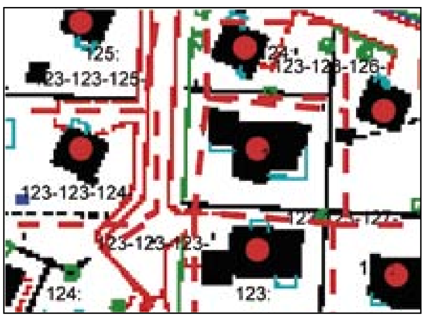

Figure 2.8 Compatible processes and operations

Scenario description: Through an information network, the municipal building monitoring system retrieves road systems and buildings in a sparsely populated area from the National Land Survey's Topographic Data Base service interface. Datasets can be displayed on top of one another along with the base map. The municipality records the new road name on the road line. If the road line (e.g. a newly constructed street) or an existing building is missing, the municipality can insert a comment, based on which the National Land Survey can conduct the necessary survey. The municipality records the new building's centerpoint co-ordinates whilst simultaneously linking the co-ordinate point to the road line. The building centerpoint co-ordinates and links to road lines can be taken from the municipal building monitoring system and, via the interface, installed in national systems. Topographic Data Base data can still be moved through interfaces to user systems. Topographic Data Base datasets can be used to produce an up-to-date address map immediately.

Scenario challenges for harmonisation: The municipal building monitoring system, National Land Survey Topographic Data Base and Population Register Centre building data system should be developed so that data can be moved via interfaces. The data being moved should be modelled in such a way that the systems will know how to use it. Agreement on data acquisition processes, data transfer protocols and copyrights must be reached. Data quality criteria and approaches should be agreed upon to ensure quality.

Scenario benefits for producers and harmonisation: Data reliability, consistency, positional accuracy and up-to-dateness are improved. Data can be quickly installed in user systems, such as for emergency rescue service use. Data is acquired once and maintained on a single system. Unnecessary processes are eliminated, which results in cost savings.

2.3.3 Common plan data

Plan datasets: regional and provincial zoning, master and detail plans, site layout plans, building prohibitions and restrictions for other measures, nature reserve areas, building protection data, approved general road plans and detail road plans.

Scenario description: The objective of harmonising plan data is to have the various original producers of plan data model plan data taking into consideration the different end user applications and data generalisation needs. The user should select the parts of plan data and data models which it needs for its own purposes. The objective is that plan data will not require interpretation, acquisition, saving or calculated generalisation. Detail plan positional data are available from the plan data system based on their drafting co-ordinates in order to confirm that the plan and operation being carried out are absolutely compatible.

Scenario challenges for harmonisation:
The challenge is the modelling of existing plan data: the data model structure and extensive datasets. The objective of harmonising plan data is to have the original plan data producers model the plan features taking into consideration the various use and data generalisation needs of the end user. The user would be able to select the parts of the plan data and data models that it requires for its own purposes.

Scenario benefits for producers and harmonisation:
Correct, precise plan data plays a crucial role in the determination of value, both in real estate trade and in other commercial activities. Accuracy of plan data interpretation: In community building it is vital that the plan position and range of influence are correct and precisely interpreted, so that interoperability with processes realising the plans, such as land parcelling and building (general areas, underground cables and structures, buildings) can be confirmed.

Usage example: Land within the detail plan area is parcelled according to municipal detail plans and site layout plan positional data, for both National Land Survey and municipal land parcelling. Land parcels are precisely and clearly plotted on the plan. The position of buildings is planned and marked in the field using the precise distance from parcel boundaries.

2.3.4 Common watercourse data

Datasets: Water Framework Directive reporting information system, Topographic Data Base

Current status: The Finnish Environment Institute is creating a database as required by the EU Water Framework Directive. The dataset base is taken from the National Land Survey's Topographic Data Base, but because the data modelling and integrity are different, the Environment Institute will, in practice, produce a new version.

Problems: The Environment Institute has provided the National Land Survey with a report on inconsistencies in the Topographic Data Base water course network which the National Land Survey has checked on the topography and corrected the Topography Data Base. The Environment Institute has made similar corrections required by the Water Framework Directive to the database. The National Land Survey constantly acquires and maintains watercourse data, but changes are not forwarded to the Environment Institute. Data has a life of its own and overlapping operations are performed in agencies.

Scenario description: When drafting a new data model for the Topographic Data Base, specifications required by the Water Framework Directive are taken into consideration. Data is acquired and maintained one time for all of Finland. The Environment Institute receives updated data through an information network XML interface in real time and focuses on the utilisation of data instead of its editing. Organisations agree on the transfer of data, data quality requirements, process audits and complaint protocols.

Scenario challenges for harmonisation: In Topographic Data Base data modelling attention should be given to Water Framework Directive specifications. The National Land Survey should add data to the XML interface and the Environment Institute should develop its information system so that it can access datasets through the interface.

Scenario benefits for producers and harmonisation: Data acquisition and maintenance are planned and performed one time only freeing resources for the utilisation of data.

2.3.5 Common land parcel data

Datasets: Ministry of Agriculture and Forestry Land Parcel Register, Topographic Data Base

Current status: Ministry of Agriculture and Forestry Land Parcel Register maintains data on land parcels. Every year in accordance with EU requirements, a designated percentage of agricultural subsidy applications shall be audited, during which the positional data of land parcels is precisely measured using GPS and attribute data is updated. The National Land Survey Topographic Data Base contains data on land used for agriculture. There are no joint operations in the maintenance of registers.

Scenario description: In connection with the annual EU audit, the precise measured positional data would also update Topographic Data Base positional data. Data would be questioned in an application-application context using an information network. The Topographic Data Base could also be given data on the cultivation status (under cultivation, fallow, pasture).

Scenario challenges for harmonisation: Ministry of Agriculture and Forestry Land Parcel Register should add the precise positional data measured in the audit to a standardised interface and the Topographic Data Base should be able to retrieve data from the interface.

Scenario benefits for producers and harmonisation: Data acquisition and maintenance are planned and performed one time only freeing resources for the utilisation of data.

2.4 Targeted outcome

The harmonisation of geographic information involves co-operation between producers of geographic information in the

a) **reduction of production costs and**
b) **enhancing of fitness for use.**

Because core geographic datasets are primarily data produced by society, co-operative efforts can reduce overlapping data acquisition and increase the quality (fitness for use) of core geographic information. The goal is that core geographic dataset modelling methods be based on common standards as well as that their semantic content and quality are compatible to the extent required. The targets set for core geographic dataset content specification, modelling, production processes, quality evaluation, data transfer and metadata are based on Finnish Public Administration recommendations (Julkisen hallinnon suositukset, JHS). Producer organisations have specified how the same datasets describing real-world features agree with one another.

2.5 Phases in the realisation of target outcome

I Tools for the combination and comparison of datasets (technical interoperability)

In the first phase of interoperability datasets can be combined easily through technical interfaces (WMS, WFS). Technical interoperability is achieved. Problems: the geometric interoperability of datasets is not achieved, the semantic content is not harmonised. The quality of geographic datasets is unknown. Interoperability with local and national datasets is not achieved.

II Harmonisation of core geographic dataset semantic content and quality

In the second phase basic geographic information producers specify the semantic content for basic data content to the extent that, based on user needs, it is necessary. In such cases the same real-world features are described compatibly so that a more precise feature can easily lead to a generalised feature. The connection between local and national datasets is known. The usability of datasets can be determined based on geographic dataset quality and metadata.

III Interoperability at the process and operational level

The positional accuracy of core geographic information allows for the easy combination of datasets. Datasets sharing similar accuracy can be used together: e.g. Topographic Data Base and boundary data (from the National Land Information Service). Datasets can be combined between municipalities, national actors and neighbouring countries. Production processes are integrated in such a way that the same core geographic information feature is compatible in all datasets.

2.6 Requirements for harmonisation

The following requirements can be set for the harmonisation of datasets:

- Requirements for modelling and data transfer
- Requirements for data content
- Requirements relating to quality
- Requirements relating to legislation and regulations
- Requirements relating to assessment levels

The sections below describe in greater detail the content of harmonisation requirements.

3. Interoperability assessment levels

Interoperability can be assessed based on dataset scope, positional accuracy and generalisation at the global, European, national, regional and local level. The objective is that data at the most precise levels can be maximally utilised at other levels. In terms of core geographic information this comprehends the scalability of common geographic information, i.e. the same feature can assume various instances at different levels. Technical solutions can be based on, for example, multiple representation databases or unique identification applications.

Figure 3.1 Illustrates the importance of assessment levels.

3.1 Interoperability globally

The Global Spatial Data Infrastructure Association (GSDI)[3] promotes international co-operation in the development of local, national and international geographic information infrastructure. It has established guidelines for the development of geographic information infrastructure - the Spatial Data Infrastructure Cookbook[4]. Its most recent version was released in 2004.

The development of geographic information standards is central to the establishment of interoperability requirements. The standardisation of geographic information was begun at the beginning of the 1990s, first at the behest of the European Committee for Standardisation (CEN) and then in the mid 1990s by the International Standards Organization (ISO).

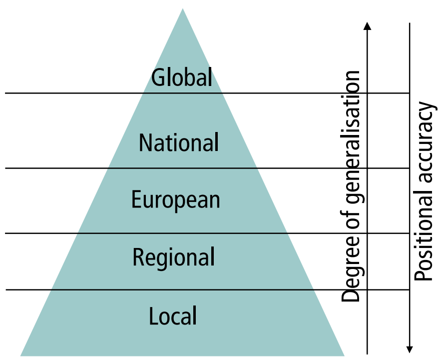

Figure 3.1 Interoperability assessment levels

[3] www.gsdi.org

[4] http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf

Key standards for harmonisation are related to content standardisation, whose ISO 19100 Series is addressed below. Figure 3.2 summarises the importance of these standards in terms of information content. The ISO 19100 Series provides guidelines on the drafting of concept models, data classification as well as metadata and quality control, but does not make any specifications regarding geographic information content. Key interoperability specifications are those published by the Open Geospatial Consortium[5] (OGC), which will be examined below.

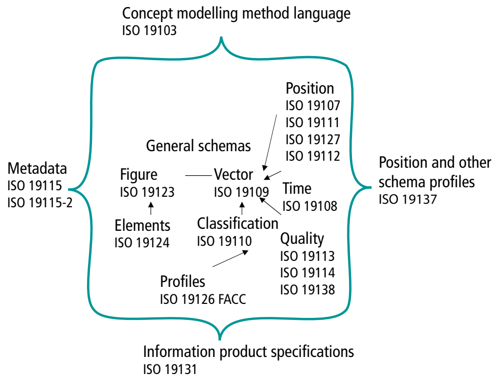

Figure 3.2 Classification of ISO 19100 standards specifying semantic content[6]

Standards specify a general frame of reference for interoperability. Content related standards include DIGEST, which was developed primarily for military applications, and particularly the FACC standard, which was developed for the classification of topographic data. The FACC profile is being made into the ISO 19126 standard. Other content related standards include nautical chart content standard IHO S-57[7] and GDF[8], which was developed for road system classification. In order to increase the application of standards ISO TC 211 established a task force for geographic information producers[9]. The purpose of the task force is, for example, to compile materials on the implementations carried out by various countries.

[5] http://www.opengeospatial.org/

[6] Source: Olaf Østenssen and (redacted) Jakobsson

[7] http://www.fma.fi/palvelut/merikartat/karttam.php?page=kartat_enc_3_iho_standardi

[8] http://www.ertico.com/en/links/links/gdf_-_geographic_data_files.htm

9 http://www.isotc211fgdp.info/

Global vector-based datasets include VMAP 0 (1:1,000,000), which was primarily produced by the National Geospatial-Intelligence Agency (NGA) as well as VMAP 1 (1:250,000) and VMAP 2 (1:50,000), which were produced in co-operation NATO member states. Only VMAP 0 covers the entire world. The dataset is based on a military standard, which is compatible with the DIGEST standard. The problem with the datasets is their up-to-dateness.

A joint effort between NASA, NGA and the German and Italian space agencies was the 2000 Shuttle Radar Topography Mission, in which specially modified radar system flew onboard the Space Shuttle Endeavour and mapped the entire earth's elevation[10] to an accuracy of 30 metres (encompassing an area between 60° north latitude and 54° south latitude).

Established from an initiative proposed by the Geographical Survey Institute of Japan, the International Steering Committee for Global Mapping (ICSGM) co-ordinated the creation of Global Map, a 1:1,000,000 dataset. The ICSGM was founded in 1996 and its dataset is based on the datasets produced by survey institutes. The objective is to achieve worldwide coverage by 2007 (coverage is currently 86%). EuroGeographics has produced a dataset for Europe, which will be presented in greater detail below.

Figure 3.3 Visualisation of the elevation model (Courtesy NASA/JPL-Caltech)

3.2 Europe

3.2.1 European situation

Europe-wide datasets include the EuroGeographics EuroGlobal Map[11] (1:1,000,000) and Seamless and Consistent Administrative Boundaries of Europe (SABE) dataset.

The goal is also to produce the 1:250,000 EuroRegionalMap dataset, in which a majority of Europe will be covered by 2007. EuroGeographics datasets are based on the original datasets of National Mapping and Cadastral Agencies (NMCAs). EuroGeographics has completed a report on Reference Information[12] produced by NMCAs for the EuroSpec programme. The objective of EuroSpec is to harmonise the specifications of geographic datasets produced by NMCAs and establish a decentralised information network for the distribution of datasets. EuroGeographics has also published a metadata service (EuroMapFinder), which allows the user to examine discovery level metadata.

Figure 3.4 EuroGlobalMap coverage

[10] http://www2.jpl.nasa.gov/srtm/index.html

[11] http://www.eurogeographics.org/eng/04_products_globalmap.asp

[12] http://www.eurogeographics.org/eng/documents/ref_data_ver1 01_part_A.pdf

Covering nearly all of Europe, the Corine Land Cover 2000 inventory (Corine2000) is based on the visual interpretation of satellite images in most European countries, but in Finland the SLICES land-use classification has also been used in the production of datasets (see Annex 1). The positional accuracy of the dataset is approx. 150 m and reliability ranges between 70-90%.

Figure 3.4 Corine 2000 inventory coverage [13]

The European Committee for Standardisation (CEN) has published guidelines (draft) on the drafting of geographic information infrastructure. The model is based on INSPIRE data architecture, the United States' Geospatial Interoperability Reference Model (GIRM)[14], the OpenGIS Reference Model and ISO 19100 standards. Because ISO 19100 standards will primarily be adopted as the European standards, any national standards not in line with them can no longer be drafted for the areas in question.

3.2.2 INSPIRE requirements

According to the basic idea of the INSPIRE-directive, European-level geographic information infrastructure will be based on the national geographic information infrastructures of individual countries. Consequently, an information network-based service infrastructure which ensures that availability of national geographic information will have to be realised in the near future also in Finland. According to the INSPIRE directive these types of infrastructure components include: geographic information metadata; geographic datasets and geographic information services, information network services and technologies; user rights agreements; co-ordination; and follow-up. The implementing rules for all these components will be drafted during 2006 and 2007.

The INSPIRE process focuses particularly on geographic information which is, in one way or another, related to the environment. The text of the directive lists in its annexes a number of datasets, which are divided into three different categories based on the degree of harmonisation urgency and requirements. The lists cover, in practice, all major geographic datasets.

The INSPIRE code of conduct will set requirements for metadata content and the function of the metadata service for its distribution. The directive text itself specifies a few basic components, which must be included in the metadata. These include: information on dataset interoperability with the INSPIRE implementing rules; information on dataset user right restrictions and fees; quality data; information on the dataset compiler, administrator and distributor; and information on the availability of datasets for public use. At the national level, these types of metadata should be available within two years (datasets in annexes I and II), or within five years of the coming into force of the code of conduct (datasets in annex III).

The harmonisation of geographic information is naturally a key objective in the establishment of a Union-wide geographic information infrastructure. An implementing rule will be drafted specifically for this subject. The directive text discusses the harmonisation of datasets and their related services on one hand, and interoperability on the other. The directive places a large number of conditions on the harmonisation of datasets. It is particularly stated that this must not incur unreasonable costs at the national level. The set goal is to bring newly compiled datasets and their related services into accord with the harmonisation code of conduct within two years of the coming into force of the code, and for other datasets, within seven years.

The objective of the directive, is the extensive interoperability of datasets. This is considered possible using special transformation services (see web services below). The directive proposal on harmonisation content states that it should comprehend the definition, classification and geo-referencing of geographic information for datasets specified in the annexes. According to the directive proposal, these types of details include: the unique identification of features; relationships between features; key feature attribute data required for envi-

[13] http://dataservice.eea.eu.int/dataservice/metadetails.asp?id=6 67

[14] http://gai.fgdc.gov/girm/v1.1/

ronmental policy; dataset temporal dimension data; update information.

According to the directive text, the aim of harmonisation is to prevent conflicts between data referring to the same place and data referring to the same geographic feature in different scales. The harmonisation implementing rule should be drafted in such a way that the information taken from various geographic datasets in accordance with it are comparable. Specific reference is made to the uniform processing of transnational geographic features.

In regard to geographic information infrastructure web services the directive text lists the following key service types: directory (metatieto) service, Web Map service, geographic dataset service, transformation service, other geographic information service. The directory service should be supported by the following search criteria: keyword, feature class, quality, interoperability with code of conduct; position; availability; responsible organisation. The purpose of the transformation service is to ensure that other services function compatibly with the harmonisation implementing rule. In regard to services other than directory services at the national level it is possible to restrict access to them based on the following criteria: statutory confidentiality; national security; due process of law; statutory commercial interest; copyright; statutory protection of the individual; environmental protection. Directory services and web map services should be free of charge to the public. The implementing rule for web services should be specify the technical details for different service types and the minimum performance requirements for services.

3.3 National projects

There are national harmonisation projects running in, for example, Germany, Great Britain and Denmark. In Germany the AAA project[15] combines three key register reference models – Official Geodetic Control Point System (AFIS), Real Estate Cadastre (ALKIS) and Topographic Survey (ATKIS) data models based on the ISO 19100 standard. Modelling is done using a UML with metadata (quality and historical data) and portrayal technology. In Great Britain the Ordnance Survey MasterMap has three levels of interoperability[16]. In Denmark technical maps produced by municipalities and the National Survey and Cadastre - Denmark TOP10DK are combined to form a single product - TopTK. The final version of the joint model is slated for release in 2006 and integration will be completed by 2009- 2010. The project is linked to Denmark's currently ongoing municipal reform, in which the present 271 municipalities and 13 regions will change to 5 major regions and 98 municipalities. Annex 1 contains an examination of Finnish harmonisation projects, which possess some of the characteristics described in these guidelines. At the national level it is important that links between core geographic datasets are defined and common feature types are based on the same source.

3.4 Regional scope

Finland is currently running municipal co-operation projects, an example of which is the joint production and administration of geographic information by municipalities in the Lahti and Turku regions. Lahti and its surrounding municipalities (5 municipalities, 160,000 residents) will implement the ETRS89 co-ordinate system (and ETRS-GK26 map projection) and a common geographic information system during 2006. The Helsinki Metropolitan Area Council has compiled statistical data in regional municipalities (SeutuCD) and produced a public transport route service, among others.

3.5 Local datasets used as the source for national datasets

The goal for interoperability should be that, for core environment modelling concepts, interoperability is complete from the lowest level, i.e. the needs of the data producer and its customers, all the way to the national level. Generalisation at the national level would then be primarily based on the selection of features. In executing an environment model using INSPIRE process concepts, it is possible to achieve a reasonable degree of interoperability also at the European and global level. A requirement for achieving interoperability is that the selection/generalisation needs are taken into consideration based on the needs of the data producer and its users in specification and modelling.

[15] http://www.advonline. de/extdeu/broker.jsp?uMen=cc6706fe-9792-9101-e1f3- 351ec0023010

[16] Vanessa Lawrence, presentation on the geographic information market 2005

SECTION II Technical requirements

4. Definition of harmonisation

In these guidelines harmonisation refers to the achievement of interoperability at three different levels:

- **geometric (positional) interoperability**: features are geometrically fit together. Geometrical interoperability can involve correlating positions of different features horizontally or vertically (positional accuracy), or their topological relationships.
- **semantic interoperability:** Interoperability of feature classifications and their meaning. Features with related meanings (such as building and address, road and address). Different degree of feature generalisation. Metadata interoperability. Conceptual modelling
- **technical interoperability:** data transfer modelling interoperability, data transfer interoperability. Interoperability can be achieved through the specification of semantic interoperability; geometric interoperability can be achieved by applying quality management and quality assessment methods, so that the specified positional accuracies are achieved. Feature links between different datasets can be defined for common features using, for example, identifiers. Technical interoperability is ensured by adhering to common modelling and data transfer principles.

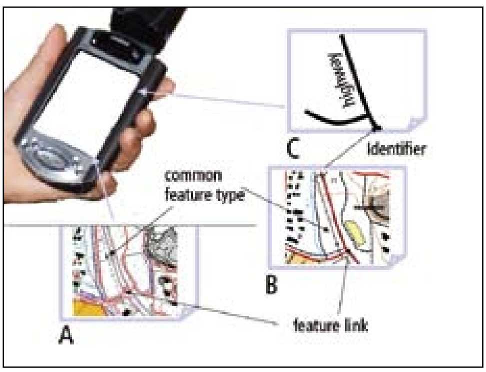

Figure 4.1 Interoperability between datasets

Data producer A's dataset has a **common feature type** (AB) with data producer B. Description of the feature type is identical for both data producers and the feature is maintained using a specified process. The specified maintenance process can be common or either of the data producers maintains the common feature. Producers datasets may contain different types of features, but they should be **interoperable** in the specified manner. Producer A and B have feature types Ay and By, which have a **feature link.** The feature link can mean that producer B uses producer A's feature types as a source, but adds, for example, attributes to it. Post-processor C takes features from producer B and generalises them for the map product. The post-processor retains the original feature **identifier data** for maintenance. The user compiles a dataset from post-processor dataset features and producer A datasets for their own use. The user can define the link between feature types because both feature types are based on data producer datasets, whose correlation is known.

5. Harmonisation areas

5.1 Requirements for semantic descriptions

The following ISO geographic information standards apply to semantic modelling:

ISO 19110 – Methodology for feature cataloguing

ISO 19131 - Data product specifications

ISO 19115 – Metadata

ISO 19113 – Quality principles

JHS-documentation:

JHS-data product specification[17]
Several JHSs for Core geographic datasets[18]

In the modelling of geographic dataset semantic content five different levels must be taken into consideration:

1) definition of features and their meaning; definition of feature topological relationships; definition of attribute data
2) definition of production process; definition of data maintenance method; definition of quality criteria (scope, positional accuracy, correctness of attribute data)
3) definition of different user requirements for data; description of intended applications; product specification; definition of data distribution requirements; description of metadata
4) individual identification of objects
5) harmonisation of the definition of similar or same features between different producers

5.1.1 Definition of features

Core geographic dataset feature and data models should be defined in accordance with the set guidelines in the JHS- data product specification. Core geographic information data product specifications are published as JHS recommendations. Features should be modelled as real-world features, taking into consideration the required degree of generalisation and intended application. Real-world features should not be generalised (combined) or distributed (refined) for a noncorrelative level. The feature model is described in catalogue form (ISO 19110 Methodology for feature cataloguing), in which each feature's data type, feature dependencies on other features as well as the principles for each feature's instance and storing of attribute data instances (required, recommended, optional). Based on the data type, the dataset can be designated as a 1D, 2D, 2.5D, 3D or 4D dataset. For example, this means the portrayal of aerial features as polygons in 2D models or as centre of gravity points in 1D models.

5.1.2 Definition of processes

In modelling semantic content attention must be given to both the data production process and data maintenance process. Data cannot be modelled to be more detailed than the production processes (data acquisition methods) allow, e.g. buildings cannot be modelled as 3D items if the data acquisition does not meet 3-dimensional requirements. Quality criteria (ISO 19113 Quality principles) should be designated for each feature type; the feature instances (incl. attribute data) should fulfil these criteria. Allowable statistical limits must also be specified for quality deficiencies. Quality should be audited using documented quality verification methods. Data production and maintenance should be designed so that the dataset correlates to the specified semantic content. When defining content, it must be ensured that data production and maintenance are not necessarily separate processes, but parts of the service process, in which data is processed in addition to other operations.

5.1.3 Definition of user requirements

When modelling semantic content and defining data acquisition processes various user requirements must be taken into consideration. Particularly in the specification of core geographic datasets, attention must be given to the data needs of different organisations in society. Different user groups should be included in the development process for the definition of data types, definition of topological relationships, and definition of quality criteria. It must be especially taken into consideration that feature data types and their meaning are understood in the same way in different organisations. The various intended data applications should be described, if different users have different needs for the dataset. Various user organisations should establish

[17] Recommended for drafting; see Section III
[18] Recommended for drafting from each core geographic dataset

common criteria for existing datasets and, if necessary, data models should be combined so that data can be compiled and maintenance performed during a single session. The combination of data models does not mean that datasets are completely combined. On the basis of user needs, combination involves the co-ordination of data modelling in such a way that the common use of different datasets would be possible. The objective is to combine the definitions of similar features when it is possible based on user needs. Dataset metadata should be described (ISO 19115 Metadata) so that the different user groups receive the necessary information based on the metadata description.

5.1.4 Unique identification of features

A unique identifier system should be defined for core geographic datasets. The unique identifier specifically individualises the each feature instance. If any changes are made to the feature instance, the identifier changes. With the identifier, the user can receive update information concerning the feature from the data producer, for example, via a network, without requiring the user to update the entire dataset. The unique identifier also allows the user to link its own data (e.g. attribute data) to a feature. The management of multilevel databases, such as for generalised small-scale features, requires the use of unique identifiers.

5.1.5 Harmonising the definition of similar or same features between different data producers

In connection with the definition of features the producer of core geographic information must examine other producers' datasets, which contain common real-world features (e.g. a building) and the combination of whose data models would be of benefit to the user and society. The combination of data models does not mean that datasets would be completely combined. On the basis of user needs, combination involves the co-ordination of data modelling so that the common use of datasets would be possible. Examples of datasets with common real-world features are:

- National Land Survey Topographic Data Base and municipal base map datasets
- Population Register Centre building data, National Land Survey Land Information Service and Topographic Data Base, municipal building data
- Ministry of Agriculture and Forestry Land Parcel Register and Topographic Data Base parcels
- SLICES land-use classification and datasets used in classification (Topographic Data Base, 1:100,000 map database, Finnish Environment Institute datasets)
- Geological Survey of Finland (GSF) soil datasets and National Land Survey Topographic Data Base
- Digiroad and the datasets used in its creation (Topographic Data Base roads, municipal datasets, Finnish Road Administration datasets)

5.2 Requirements for modelling and data transfer

ISO standards related to geographic information modelling

ISO 19103 - Conceptual schema language

ISO 19107 - Spatial schema

ISO 19108 - Temporal schema

ISO 19109 - Rules for application schema

ISO 19118 – Encoding

ISO 19136 – Geography Markup Language (GML)

JHS documentation:

JHS Geographic information modelling for data transfer[19]
JHS Geopraphic Information content services[20]

5.2.1 Conceptual schema language

A conceptual, implementation-independent data model (conceptual schema) on basic geographic information **should be drafted** using UML (Unified Modelling Language). When drafting a conceptual schema, the modelling protocol specified in ISO 19100 standard series **should be observed.** The basis for the modelling guidelines is formed by standard ISO 19103 Conceptual schema language. The standards list a number of key basic data types, which should be applied in modelling.

The ISO standard series also specifies a generic model to be used as a basis when modelling vector-form geo-

[19] In preparation
[20] In preparation

graphic information. This model is known as the General Feature Model (GFM), which is specified in the standard 19109 Rules for Application Schema.
Key principles in GFM-compliant data models are:

- modelling geographic information as real-world objects - **features**
- in accordance with object-oriented modelling, features are classified into feature type categories
- the properties of these feature categories are defined in the corresponding **feature classes**
- if required, feature classes comprise object-oriented modelling-compliant **inheritance hierarchies**
- in addition to the inheritance relationship, other **relationships** may prevail between features (and, in turn, feature classes), such as composition relationships and other freely definable and designatable relationships important for applications
- features (and, in turn, feature classes) are comprised of a number of properties
- properties are assigned a **data type**, which must be either 19103 compliant, specified in other ISO 19000 series standards, or locally defined in application-specific models
- also, the specified **position** of a feature is regarded as one of its properties and the corresponding data type is one of the geometric types defined in the 19107 Spatial Schema standard
- features can have several **properties related to position**

More detailed information on the drafting of conceptual models is presented in the JHS Geographic information modelling for data transfer.

5.2.2 Data transfer schema

The dataset based on the conceptual model can be transferred from one system to another using a standardised data transfer mechanism. A concrete data transfer schema using Geography Markup Language (GML) **should be** derived from the conceptual model for use in implementation. Key ISO 19100 series standards for the data transfer schema are 19118 Encoding and 19136 Geography Markup Language (GML).

The Encoding standard provides a general frame of reference for the building of data transfer mechanisms. A key principle in this frame of reference is the role that the application schema plays during data transfer as a common language used between the systems transferring data. The basic idea behind the Encoding standard is that the systems' internal data structure is preserved. As a result, data transfer requires a mechanism, which converts the internal data structure of the system for the data being transferred into a structure compatible with the common application schema. Correspondingly, the receiving system should convert the incoming data into a form that is compatible with its internal data model before the data can be entered into the system. The application schema (and its corresponding data transfer schema) should, in terms of the two systems, actually be a virtual data model which is only realised during data transfers. Transformation processes become vital to data transfer.

ISO standard 19136 specifies an XML-based method for encoding feature-based, vector-form geographic information. The standard also provides detailed rules on how a corresponding GML-compatible data transfer schema can be derived from an application schema described in UML form. In accordance with the basic principles of GML this data transfer schema is defined in XML schema document form. It is, however, important to note that, because data modelling in accordance with basic GFM principles is implementation-independent, the corresponding data transfer schema does not necessarily have to be a GML schema. Consequently, various implementations can be derived from the conceptual schema for data transfer or other practical applications. More detailed information of the GML is presented in the JHS Geographic information content services.

5.2.3 Data service

In order to ensure the smooth transfer of data, it is important that information systems can be accessed through a standardised, application service interface. ISO has begun to refine the Web Feature Service (WFS) interface standard, which was originally developed by the Open Geosptial Consortium (OGC). In the same ISO process the WFS standard is accompanied by its affiliate standard, Filter Encoding (FE), which defines the geographic information query language (corresponds to the "where" clause of the SQL sentence). By default the interface distributes geographic information in GML form. If the ISO work results in completion of the official ISO 19100 series standard for geographic information service query interfaces, this is also the obvious choice for a geographic information interface standard in Finland.

The WFS interface consists of three query types: GetCapabilities, DescribeFeatureType and GetFeature. GetCapabilites is the service metadata query; DescribeFeatureType query returns the schema of the desired feature class, specified in XML Schema form; and GetFeature is used to request the desired set of features from the service. The feature query is given using a filter, as specified in the FE standard.

More detailed guidelines regarding the application of WFS and FE standards in Finland are presented in JHS document "Geographic information service"

5.3. Quality requirements

ISO standards and technical specifications related to the quality of geographic information:

ISO 19113 – Quality principles

ISO 19114 – Quality evaluation procedures

ISO 19115 – Metadata

ISO 19138 – Data quality measures

ISO standards and technical specifications related to quality control:

SFS-EN ISO 9000
Quality Management systems. Basis and terminology

SFS-EN ISO 9001
Quality Management systems. Requirements

SFS-EN ISO 9004
Quality Management systems.
Guidelines for improving performance

SFS-EN ISO 19011
Quality Management and/or Environmental Management system auditing guidelines

ISO/TR 10013:fi
Quality Management system documentation guidelines

Public Administration Recommendations:

JHS 152 Process descriptions. General structure, representations and concepts
JHS Description and assessment of geographic information quality[21]

Core geographic datasets should include:
1. Common data quality measures
2. Defined quality requirements based on user needs and using common indicators
3. Described and measurable quality management process
4. Datasets are tested by an independent party based on generally approved methods
5. Auditability of geographic information and production processes
6. Quality information in metadata

5.3.1 Common data quality measures (ISO TS 19138)

Common data quality measures should be developed for core geographic datasets to describe their quality using key quality factors, which are specified in the JHS recommendation on the description and assessment of geographic information quality based on the 19113 standard. Key quality factors are completeness, logical consistency, positional accuracy, thematic accuracy and temporal accuracy. One example of common data quality measures is Root Mean-Square Error (RMSE). Completeness: allowable number of errors; Logical consistency: passing of integrity tests, etc. Common data quality measures are used to ensure comparability between different datasets.

5.3.2 Defined quality requirements based on customer needs using common data quality measures

The quality requirements of geographic information should meet user needs. Some data producers have already defined quality requirements, e.g. the National Land Survey Topographic Data Base Quality Model. Users must know the quality of geographic information, so that they can assess, for example, the combinability of datasets. Quality requirements should be defined according to user needs and changes in these needs must be monitored constantly. Geographic information quality requirements can be "goal-oriented" or empirical. In the empirical approach the current quality level of geo-

[21] In preparation

graphic information is determined and the targets are set based on this. The definition of quality requirements are vital to basic geographic information, because quality data presented in metadata are usually quality requirements. It is then possible to test whether the quality requirements were achieved in the quality evaluation process. Quality requirements can also be defined as binding quality promises.

5.3.3 Described and measurable quality management process

Core geographic dataset producers should describe how they control quality during production. Geographic information production processes are often very long in duration, and can take several years, after which only a follow-up quality evaluation can be performed and the production output may be lost unless the test is approved. The JHS recommendation on process description can be advantaged in describing the geographic production process.

5.3.4 Datasets are tested by an independent party based on generally approved methods

Applicable random sampling test methods are being developed for core geographic datasets. Quality evaluation procedures must be developed for both vector and raster datasets.

5.3.5 Auditability of geographic information and production processes

The processes of core geographic information producers are audited. The producer must be prepared for the auditing of their geographic information production process. Audits are conducted to ensure that the production process functions according to plan and that the production quality adheres to set quality models. Audits are conducted at the behest of the orderer. The ISO 19011 standard is observed in auditing.

5.3.6 Quality results in metadata

Quality results must be integrated in metadata and included in geographic information.

5.3.7 Trackability of geographic information

For some core geographic information, the sources on which the datasets are based must be shown.

5.4 Development of legislation and regulations

In the harmonisation of core geographic information attention must be given to the fact that the compilation of many datasets is based on legislation. The success of harmonisation projects then also depends on the amendment of legislation where necessary. The common use of geographic information also requires agreements on dataset copyrights and terms of use.

SECTION III Implementation

6. Measures for the harmonisation of core datasets

6.1 Finnish Council for Geographic Information measures

The Finnish Council for Geographic Information plays a central role in the realisation of the National Geographic Information Strategy in deciding on further measures to be taken by divisions operating under its mandate. The Council has **approved the harmonisation definitions specified in this document as the principles for harmonisation of core geographic information in the realisation of the National Geographic Information Strategy.**

The Council shall initiate **the harmonisation of geographic information protocol in administrative bodies, which involves the internalisation of the result management between ministries and administrative bodies.** In result management it may be required that, based on the achievement of the abovementioned harmonisation levels, core geographic dataset producers from administrative bodies draft a plan. Furthermore, the Council should initiate joint projects that support harmonisation, which will be discussed in greater detail below. **In addition to the Finnish Council for Geographic Information, harmonisation will also require the presence of a strong support organisation. The Harmonisation Division can serve as a co-operative forum for various administrative bodies, but clearly there is a need for an actor, which can oversee practical co-ordination.** Core geographic information can be assigned national certification, which can be used by ministries to ensure the realisation of targeted harmonisation outcomes. Certification can be granted by the Council based on an audit. The audit process determines how an administrative body's core geographic dataset fulfils the harmonisation requirements specified in these guidelines. The audit results in procedural recommendations and possible certification, e.g. Topographic Data Base building data fulfils harmonisation requirement level I.

6.2 Administrative body measures

6.2.1 Harmonisation preparations

The harmonisation of geographic information takes time and requires co-operation between administrative bodies. It involves the redefinition of data and quality models, the reorganisation of data acquisition and the specification and realisation of information system-based data service interfaces. The impact of harmonisation can be felt for a relatively long time, because datasets are usually extensive, containing a large number of feature classes. Harmonisation generally leads to new data specifications and even requires new data acquisition. The cost savings over the long term can, however, be substantial. Harmonisation demands a considerable investment in the training of organisation personnel. The adoption of international standards; the practical application of data specifications (UML, XML); the definition of quality criteria and quality measures; the elimination of overlapping data acquisition between organisations; and the drafting and enactment of national JHS recommendations require considerable human resources. Harmonisation can also lead to the amendment of legislation, which requires preparation. Administrative bodies should consider **joint support organisations,** such as fora, which can be used to share experience and provide training.

6.2.2 Feature-specific harmonisation of core geographic datasets

Administrative bodies should perform the feature-specific harmonisation of core geographic datasets. This begins with the identification of datasets containing or processing the same or similar data. In practice, the data specifications for datasets should describe standards (UML, XML, JHS recommendations) in the required uniform manner. In connection with the description of data specifications, dataset features with the same features and whose data specifications require harmonisation are identified. At the behest of administrative bodies, these features are assigned a common data specification, which meets the needs of all organisations and users requiring data. The objective to compile and maintain features in one process and save them in the database in a manner which allows for the exchange of data as an interface service in an information network. In feature-specific specification decisions must be made concerning the organisation of the data acquisition process, identification of individualised features and possibility for the exchange of change data. Examples of harmonised datasets include:

- topographic data (municipal – National Land Survey)
- building (Population Register Centre – National Land Survey - municipalities)
- watercourses and conservation areas (Finnish Environment Institute – National Land Survey)
- land parcels (Ministry of Agriculture and Forestry Land Parcel Register – National Land Survey).

6.2.3 Implementation of unique identifiers of features

Administrative bodies should implement a feature-specific identifier for geographic datasets, in which change data is exchanged between organisations or applications. Examples of these kinds of datasets include:

- land division (National Land Survey – municipalities - National Land Survey)
- road system data (Finnish Road Administration (Digiroad) – municipalities – National Land Survey – user application)
- buildings (Population Register Centre – National Land Survey - municipalities)
- topographic data (municipal – National Land Survey - user applications).

Other datasets should also be prepared for the unique identification of features, particularly if the intention is to use one dataset to control products being made at different levels of generalisation (e.g. using the Topographic Data Base in the production of small-scale maps). The technical execution of the unique feature identifier implementation must be examined. Technical protocols have not yet been standardised, but in several European countries (such as Great Britain (OS) and Germany (ATKIS) there are examples of functional execution solutions. One alternative is the Universal Unique Identifier – UUID [ISO 19118, ISO 11578], which has received mixed reviews in the geographic information sector. For example, UUID is being used in the production of common European road system data. A JHS recommendation should be drafted on the use of individualised feature identifiers in core geographic datasets.

6.2.4 Definition of feature quality requirements

Administrative bodies should draft quality models on core geographic datasets. A project, in which a sample quality model whose structure can be reproduced by organisations is defined, should be initiated in order to complete the work. A sample quality model is based on international standards and quality management on JHS recommendations. The quality model specifies the dataset quality level and quality criteria, which are used to assess quality. In the quality model it is possible to separately define internal quality (process quality requirement) and external quality (quality promise to the user). Quality criteria are specified for such things as geographic information feature-specific positional accuracy, feature-specific coverage, and the correctness of attribute data. Quality criteria are set in relation to feature data specifications, not in relation to absolute truth (feature in topography). The quality model or an annex to it lists the quality control procedures used in dataset data acquisition; quality control procedures for datasets developed by organisations; and protocols for quality tests conducted by third parties. Quality tests should be based on standardised statistical methods.

6.2.5 Process standardisation

In connection with the data specification of common features, administrative bodies should also identify overlapping data acquisition processes and agree upon a joint effort to ensure that overlapping operations can be eliminated. A common data acquisition process requires uniform data specification, uniform quality requirements and agreements on the regional assignation and timing of data acquisition. A common data acquisition process especially requires that data can be smoothly transferred from the data acquirer to data user. In practice data transfer will be based on a standardised (WMS, WFS, WSDL, SOAP) message handling system via information networks. Examples of functional data services are the National Land Survey's service interfaces and their applications (Karttapaikka (Mapsite), UKTJ (New Land Information Service) browser service). When combining processes, administrative bodies should define the protocols for process auditing, complaints processing, and dispute resolution.

6.2.6 Definition of geographic information product specifications

Administrative bodies should define the product specifications for core geographic dataset products. A project making preparations for a sample product specification should be in order to complete the work. This should be based on ISO 19100 standards (such as ISO 19131, ISO 19110, ISO 19115, ISO 19113, ISO 19114). The product specification describes a feature model, i.e. the meanings for features and their attribute data (in relation to the real world), relationships and operations between features, feature classification, reference system, feature quality data, geographic data production and maintenance process, feature portrayal and performance specifications.

Annex 1: Projects promoting commonality in Finland and Europe

UKTJ (www.ktj.fi) Common national cadastre and data service

The New Land Property Information Service was launched on 1 June 2005. The cadastre data and property position data of municipalities maintaining cadastres and the National Land Survey have been combined into a single national cadastre, for which there is a cadastre service. The objective of UKTJ is to improve both municipal and National Land Survey operations by eliminating the need for double maintenance of cadastre data in municipal and National Land Survey registers. Another objective is to make possible the use of correct, up-to-date and comprehensive positional cadastre data using a single service, which facilitates the utilisation of data in both public administration and commercial service production in the private sector. Correspondingly, municipal and National Land Survey data service revenues are more proportionate to data production expenditures.

Digiroad (www.digiroad.fi) National road and street information system

The basis of the Digiroad information system is the "Act on the national road and street information system" as well as two statutes related to it (Information Types (997/2003) and Payment (1082/2003) statutes) The Finnish Road Administration is responsible for the information system. The National Land Survey, Finnish Road Administration and municipalities produce datasets for the system. The Finnish Road Administration and National Land Survey have agreed on data specifications and methods in which geometric maintenance is performed. Digiroad's road geometry is similar to the road geometry in the Topographic Data Base, but data structures and attribute data in the Data Base are different. Road system change data (new features, deleted features and changed features) are transferred from the Topographic Data System to the Digiroad database in XML form 6-10 times every year. In the Finnish Road Administration the Administration's own register data and attribute data produced by municipalities are added to geometric data. Digiroad products are also based on XML messages. The Digiroad project is an example of a co-operative scheme, **in which a common data specification is agreed upon among different organisations.** It is also agreed that the data acquisition process shall be carried out at the behest of different organisations (National Land Survey, municipalities) and the administrator of the information system. Data is transferred between organisations using standardised message handling systems. **Operations are also based on existing legislation**.

Slices (www.slices.nls.fi) Common use project for the creation of a regional land-use database

Slices is a geographic information common use project, whose co-operative parties are the Ministry of Agriculture and Forestry, Ministry of Environment, National Land Survey of Finland, Finnish Forest Research Institute, Environment Institute of Finland and the Population Register Centre. Slices is a raster form (25 m pixel size) dataset, which describes regional land use. The dataset covers the entire national area. It was completed in 2000 and updated in 2005. The sources used are: Topographic Data Base; Land Parcel Register; National Forest Inventory (VMI); Water System Database; Conservation Area Database; Building and Housing Register (RHR); High-voltage lines (Finngrid); and Corine Land Cover datasets. Vector-form source datasets are first converted into raster form and all datasets are added to the Uniform Co-ordinate System. Derivative raster datasets are processed in three different phases (preprocessing, processing and post-processing) per emergency service sheet (80 km x 80 km). During processing, datasets are filtered and combined with one another appropriately. Processing is done at a pixel size of 19 m. The end product describes the minimum size of areas larger than 0.25 ha in raster form (25 m pixel size). The Slices project has processed a large number of very differently modelled datasets, which are heterogeneous in quality and resolution.

Harmonisation of basic geographic information

The harmonisation of information produced by municipalities has been done in several different phases. 1980- In the 1980s an effort was made to harmonise municipal **control point register datasets** in the Association of Finnish Local and Regional Authorities KATKO task force "Blue Book", containing recommendations on the Control Point Register data content, which list point-like features to be saved in municipal control point registers (feature name, explanation, encoding, number of co-ordinates, level, height, both). The guidelines represents the de facto standard and is closely observed in both municipalities and by information system providers. The reasons behind its success would seem to be concise data content, the right timing and the low cost of implementing the guidelines. **Municipal topographic data classification** (now known as Geographic information classification) has made an effort to expand the conceptual specifications of geographic information produced by municipalities in the transition from the control point register to geographic information databases, in which environments are modelled as entities and geometric concepts have been expanded to include points as well as lines and areas in attribute data. Preparations for classification were begun in the early 1990s. In original topographic data classification an effort was made to cover municipal base map dataset concepts, whose attribute data included those which typically appeared on base maps and site plan ordnance survey maps. The classification of municipal geographic information was published as a database, as an application and a publication. Due to the extent of the data classification its implementation is typically associated with information system renovations. Otherwise, the hurdle for implementation of classification has been cost as well as the fact that no concrete benefits of implementation have been seen. Existing data transfer specifications (such as the "Fingis" format or other information system-specific data transfer methods, e.g. DXF, etc.) has been sufficient. The minimal need for transferring attribute data as well as the lack of possibilities for processing in information systems used by municipalities has resulted in a situation where transfer formats intended for geometric transfer have proven satisfactory and there has been no need to upgrade to a more advanced data transfer format.

EuroSpec www.eurogeographics.org Improving the interoperability of national ordnance survey datasets and dataset harmonisation project

The project is an example of the co-operation of European survey organisations in harmonisation. The EuroSpec programme is a EuroGeographics development project, whose aim is to create common specifications for basic datasets produced by European ordnance surveys. One of the most important objectives for the near future is the creation of a common specification for the already produced EuroGlobalMap 1:1000 000, EuroRegionalMap 1:250 000 and Seamless and Consistent Administrative Boundaries of Europe (SABE) dataset. Key datasets have been assigned targets for road system, water system, elevations, names systems and cadastre datasets. Currently ongoing projects related to the Eurospec programme are **EuroRoads** (creation of a common road system specification), **RISE** (creation of a 1:250, 000 dataset specification in accordance with the Water Framework Directive), **EuroMapFinder** (creation of common metadata directory), **EuroGeoNames** (names system specification), **EuroBoundaries** (harmonisation of boundary datasets between countries). The objective of the project is to work in co-operation with the European Union INSPIRE project in the formulation of specifications. Services based on common specifications are expected to be in place by 2008.

GiModig gimodig.fgi.fi Research project on the development of a common data model for mobile topographic applications

The GiMoDig project drafted a common data model as a GML schema, based on the topographic geo-databases of four national mapping agencies (Finland, Sweden, Denmark and Germany). The model is comprised of seventeen feature classes: Administrative Boundary, Water, Watercourse, Lake/Pond, Marsh/Swamp, Park, Building, Contour Line, Cropland, Named Location, Built-Up Area, Railway, Road, Trail/Footpath, Airport/Airfield, Forest, Grassland. There was correspondence between national data models and the GiMoDig schema, with some exceptions. As a result, based on the datasets of the four different national mapping agencies, a reasonably homogeneous and usable map could be produced for the two test areas, which were located in border areas. The project also developed a functional, six-level service architecture-based prototype service, which implemented the OGC service interfaces as follows: At the Data Service level and Integration Service level: Web Feature Service (WFS); At the Portal Service level: Web Map Service (WMS) and OpenLS PresentationService. The service's Processing Service level interface was extended from the WFS standard. The service performed on-line database queries to the national WFS services and made the necessary data model and co-ordinate system transformations in real-time. Further information: http://gimodig.fgi.fi

Topographic Data Base and municipal datasets (www.fgi.fi/….) Research project on the harmonisation of datasets

This was a joint research project funded by the Ministry of Agriculture and Forestry, Finnish Geodetic Institute and National Land Survey. The reference group included funding agencies, the Association of Finnish Local and Regional Authorities and representatives for municipalities in the test areas. The project was launched in April 2003 and ended at the turn of the year in 2004. The research project examined the interoperability of National Land Survey Topographic Data Base and municipal base maps (as well as the use of municipal datasets in built-up areas in the production of topographic data. The research project conducted a survey aimed at municipalities of dataset coverage, maintenance and pricing. There were five test areas, two of which were located in Helsinki and the remainder in Espoo, Jyväskylä and Mikkeli. These test areas were used to form a simulated topographic database based on municipal datasets, and feature model differences and quality were examined. A connection was modelled between the National Land Survey Topographic Data Base and municipal geographic information classification. A user needs survey was conducted by interviewing major users and post-processors selected from the National Land Survey user register. Based on the results, the Topographic Data Base can be at least partially derived from the base map dataset produced by the municipal population centre. Approximately 90% of Finland's entire population lives in population centres and major portion of the country's infrastructure is found there. Municipal datasets are currently for up-to-date cadastres and buildings, but only 37% of municipal representatives believe that the datasets as a whole are very up-to-date. User needs seem to be focused especially on buildings and road systems. The National Land Survey datasets should cover the entire country, be similar in quality and description, and in the same format. The realisation of common use requires the development of legislation, guidelines and technical requirements.

Appendix 2: Concepts

The source used for terminology is the geoinformatics dictionary[22] unless otherwise specified.

geometry
data describing the form of a geographic feature

geometric object
a spatial object which describes the geometry

feature
an object which corresponds to individualised abstract or concrete things or phenomena in the real world

feature type
an object class which defines features

universe of discourse
view of the real or hypothetical world that includes everything of interest[23]

feature substitution
feature succession in which one feature instance is replaced by another feature instance of the same or different feature type[24]

feature relationship
conceptual connection between two or more features

conceptual model
model that defines concepts of a universe of discourse[25]

conceptual schema
formal description of a conceptual model[26]

model
simplified portrayal of a system, process, device or concept A model can be a mathematical or visual portrayal

object
entity an individualised entity in the information system which has spatial and behavioural attributes

object class
objects with the same surveyable attributes and same functions

[22] http://www.tsk.fi//fi/info/GeoinformatiikanSanasto.pdf
[23] ISO TC 211 Terminology (2006-08-05)
[24] ISO TC 211 Terminology (2006-08-05)
[25] ISO TC 211 Terminology (2006-08-05)
[26] ISO TC 211 Terminology (2006-08-05)

attribute; property
a characteristic or descriptive property

geographic feature
representation of real world phenomenon associated with a location relative to the Earth[27]

inheritance
method in which the new object class being formed, or subclass, assumes the attributes and functions of the existing class, or superclass

spatial object
object used to describe the attributes related to the position of the geographic feature

spatial attribute
attribute describing the geometry or topology of a geographic feature

schema
formal description of a model[28]

application schema
conceptual schema for data required by one or more applications[29]

thematic attribute
an attribute describing a geographic feature according to a selected theme

data model
a model which describes data and relationships between pieces of data

topology
data describing the positional relationships of geographic features or parts of them, which remains constant in continuous transformation

topological object
a spatial object which describes the feature topology

interoperability
capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units[30]

[27] ISO TC 211 Terminology (2006-08-05)
[28] ISO TC 211 Terminology (2006-08-05)
[29] ISO TC 211 Terminology (2006-08-05)
[30] ISO TC 211 Terminology (2006-08-05)

Publications of Ministry of Agriculture and Forestry 2006

Viimeisimmät julkaisut:

1/2006	Maaseutuelinkeinoneuvontajärjestöjen valtionapua saavan toiminnan arviointi ISBN 952-453-249-2
2/2006	Etelä- ja Länsi-Suomen maaseudun kehittämisohjelma vuodelle 2006 ISBN 952-453-250-6
2a/2006	Program för utveckling av landsbygden i södra och västra Finland för år 2006 ISBN 952-453-257-3
3/2006	Tavoitteena hyvinvoiva nauta (2003) ISBN 952-453-242-5
4/2006	Alueelliset metsäohjelmat 2006–2010 Yhteenveto metsäkeskusten metsäohjelmista ISBN 952-453-272-7
5/2006	Eläinlääkärit 2006 ISBN 952-453-274-3
6/2006	EU Competences in Forestry Policy ISBN 952-453-277-8
7/2006	Suomen maaseututyypit 2006 ISBN 952-453-283-2
8/2006	Suomen Leader+ -ohjelman väliarviointi 2005 ISBN 952-453-286-7
9/2006	Tekemällä oppii – EU-osarahoitteiset maaseudun kehittämishankkeet nuorten elinolojen vahvistajina ISBN 952-453-287-5
10/2006	Kansallinen viljastrategia 2006–2015 ISBN 952-453-288-3
11/2006	Metsäsektorin tulevaisuuskatsaus – Metsäneuvoston linjaukset metsäsektorin painopisteiksi ja tavoitteiksi ISBN 952-453-295-6
12/2006	Toiminnan tehostaminen paikkatiedoilla ISBN 952-453-300-6

ISBN 952-453-301-4
ISSN 1238-2531

www.mmm.fi
P.O.Box 30, 00023 GOVERNMENT