EUROPEAN REFERENCE DATASETS FOR EUROPEAN SPATIAL DATA INFRASTRUCTURE – STATE OF THE ART AND DEVELOPMENT OF COMMON SPECIFICATIONS

Mr Antti Jakobsson

National Land Survey of Finland
PO BOX 84
FIN-00521 Helsinki
Finland
Tel: +358 20541 5177
Fax: +358 20541 5070
E-mail: antti.jakobsson@maanmittauslaitos.fi

Abstract

This paper gives an overview of the reference data sets in Europe. It is based on the report on Reference Data Sets and Feature Types in Europe compiled by Eurogeographics Expert Group on Quality. The report covers 33 European countries and represent datasets from 82 organisations producing 236 datasets.
Results cover topographic, cadastral, address and administrative data sets, digital elevation models, and orthophotogrammetry. Furthermore the paper explains the availability and characteristics of common feature types including roads, administrative boundaries and areas, addresses, buildings, lakes and ponds, watercourses, coastlines and shorelines, names, contour lines, benchmarks, cadastral parcels and built-up areas.
The state of the art report will form the basis for development of common specifications among NMCAs in Europe. The Eurogeographics programme EuroSpec will facilitate the proposed directive for European Spatial Data Infrastructure (INSPIRE). The programme includes specification work, distributed data production and maintenance and distributed services. At the moment the policy papers set the target for the next decade. NMCAs will take the challenge of bringing common spatial reference datasets to the customers and citizens in Europe. Significant investments at the national level will be required to the data management, specifications and content of national datasets.

INTRODUCTION

Eurogeographics is an association representing nearly all European National Mapping and Cadastral Agencies (NMCAs). Currently it has members from 46 organizations from 41 countries. The vision of the organization is to achieve interoperability of European mapping and other GI data. This paper will present the results of the study of state of the art of reference datasets in Europe, and it is part of the Eurogeographics programme of developing common specifications (EuroSpec). The Expert Group on Quality carried out this study in 2003 and 2004. The first draft results were presented at the Eurospec workshop in Paris on March 5th 2003, and the second version was presented at the second workshop in Paris on July 3-4th 2003. The report on the reference datasets (1) was published in 2005.

The reference data report presents the situation as regards many reference themes described in the proposed directive (2) on establishing an infrastructure for spatial information in the Community (INSPIRE) at the most accurate level with national coverage. Reference themes of highest priority in the INSPIRE directive (Annex I) are co-ordinate reference systems, geographical grid systems, geographical names, administrative units, transport networks, hydrography and protected sites. The report covers geographical names, administrative units, transport networks and hydrography. From the second highest priority we covered elevation, identifiers of properties (addresses), cadastral parcels, and orthoimagery. From the third category we covered buildings and built-up areas.

METHODS

This paper is based on the report on Reference Data Sets and Feature Types in Europe (1). The study was based on a questionnaire sent to Eurogeographics members and all National Mapping Agencies that we had contact details for. It was sent out on January 17th 2003 and updated in 2004. The questionnaire was divided into three main part in a word document: Organization, databases and reference themes. We used INSPIRE definitions (3) for the reference data themes. The FACC data dictionary (4) together with EuroRegional map specification (5), SABE Administrative hierarchy (6) and GiMoDig global schema (7) were used for the definition of feature types. ISO 19110 Feature cataloguing methodology (8), ISO 19115 Metadata (9) and ISO 19131 Data Product Specification (10) were also applied to create a general structure for the questionnaire.

We divided Europe into regions, and a regional co-ordinator was responsible for assisting NMCAs to fill in the questionnaire. The questionnaire was sent to 42 organizations covering 37 countries in Europe and 1 country outside Europe (Armenia). Answers covered 33 countries and datasets from 82 organizations. In total, 89% of the countries replied, and this questionnaire represents 73% of the European countries (45 countries in Europe), and nearly all EU countries including the new members that joined in 2004 (except Luxemburg).

REFERENCE DATASETS

In total, 236 databases were identified. The following chapters will describe the characteristics of the topographic and cadastral datasets. The status of digital elevation models (DEMs) and orthoimagery are also covered.

Object orientation is typically the goal for vector data sets. We did not specify precisely what object-based means in the questionnaire. We did, however, give an example stating that 'object-based' means e.g. a 'real-world' geodatabase, whereas 'point and line' refers to a cartographic database. Nevertheless NMCAs did seem to understand the options given. Most data sets are not yet object-based, but conversion is planned for most cadastral and topographic databases that still have a point-line structure.

In the questionnaire we set out to determine the progress made by NMCAs in achieving interoperability with a question about edge matching covering technical aspects, content and updating policy. Only a few countries had engaged in co-operation with other countries, and for the most part there were no agreements in place.

Some standards were mentioned. They were, however, mainly national standards or product specifications. It can be concluded that the harmonisation process has not yet been initiated at the national level among nations, and international standards in geographic information are not yet generally applied.

Topographic datasets

All countries had topographic data in some format, butconcluding from the names of the data sets, some countries did not appear to have a topographic database. A topographic database was inferred to exist when a database comprised of several reference data themes. Based on this assumption, a topographic database existed in 23 (out of 33) countries. The majority of the countries had already converted the databases so as to be object-based. The database was object-based in 10 countries and had a point-line structure in 5 countries. In 7 countries were plans to introduce a an object-based database, 1 country did not report the structure used. 11 countries had separate digitised themes. The coverage of the topographic databases is shown in Figure 1. Coverage in France is based on a 1:50 000 database, but there is also a 1:10 000 database with 60% coverage. In Slovenia coverage is based on 1:25 000 database but they also have a 1:5000 database with 20% coverage. In Switzerland the coverage is based on 1:25 000, but they also have a 1:1000 database with 60% coverage.

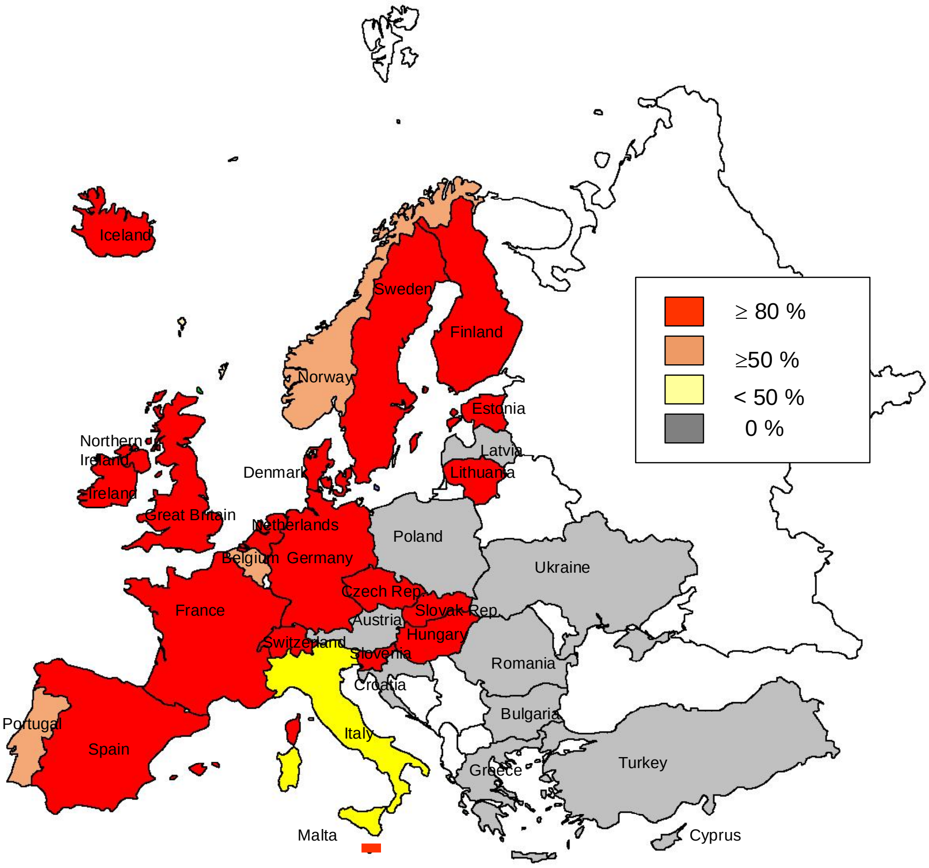


Figure 1. Coverage of topographic datasets

Cadastral datasets

Cadastral databases were identified in 23 countries, 2 countries had a cadastral map and 8 countries had no cadastral database whatsoever. The structure was object-based in 13 countries, point-line in 3 countries and other in 1 country. No information was given in the case of 1 country and plans to convert to object-based within two years existed in 5 countries. Figure 2 illustrates the coverage of the cadastral databases. Cadastral parcels are among the most frequently updated feature types in the countries (continous updating in 68% of the countries).

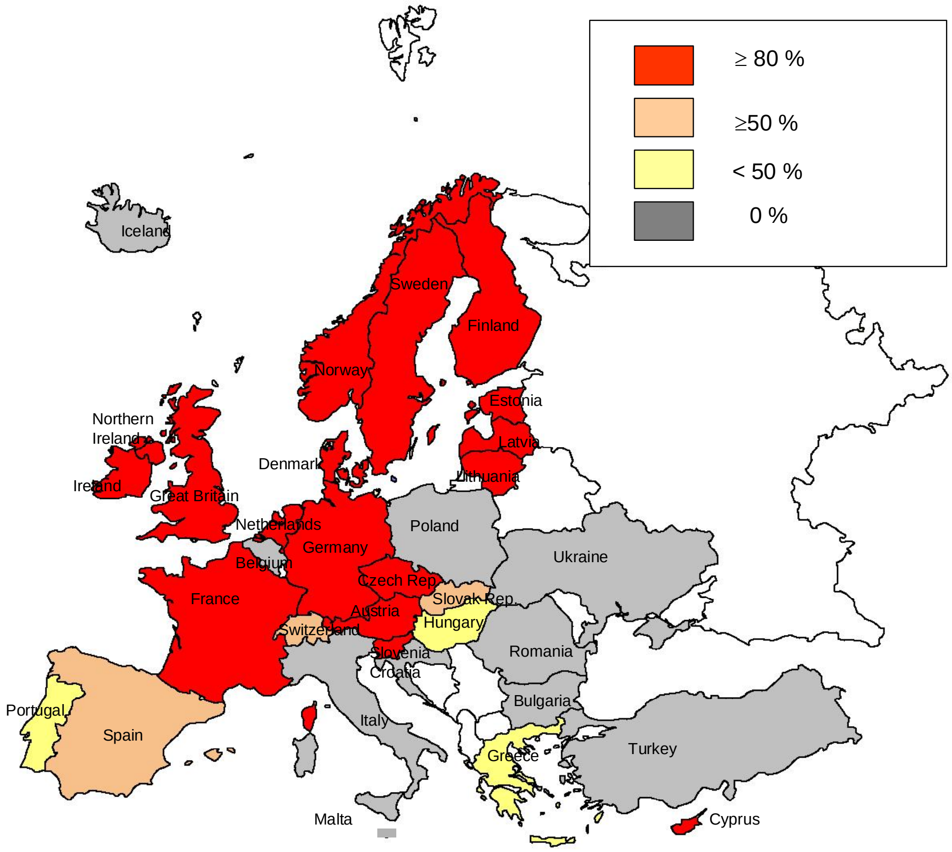


Figure 2. Coverage of cadastral datasets

Digital elevation models

Digital elevation model (DEM) existed in 26 countries (78%). Figure 3 illustrates the coverage of DEMs. The typical resolution was between 20-50 meters (16 countries).

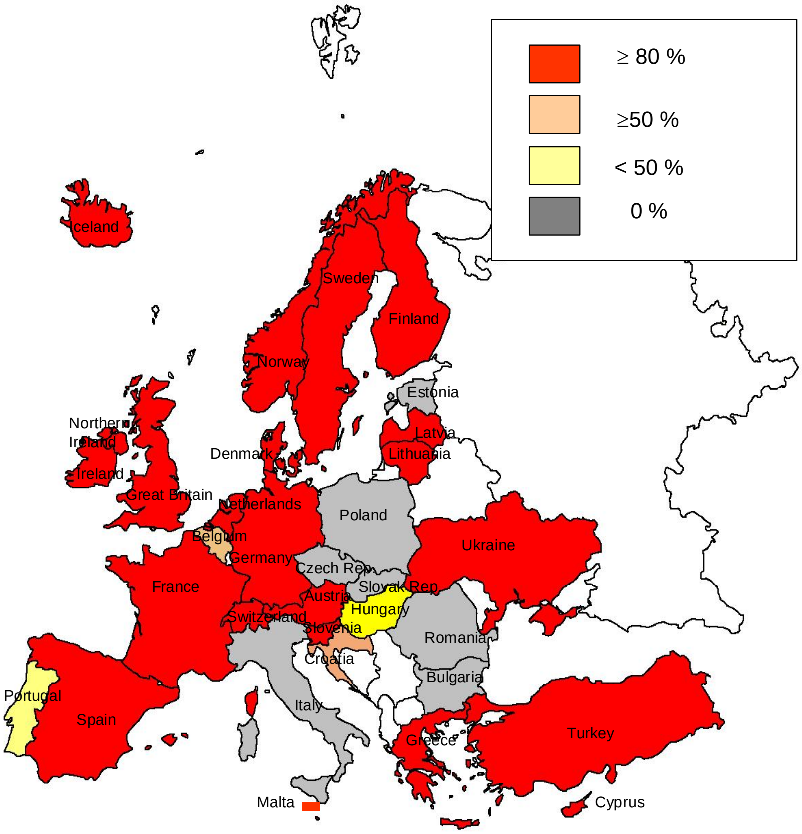


Figure 3. Coverage of DEMs

Orthoimagery

Digital orthoimagery was produced in 27 countries (82%). Figure 4 illustrates the coverage of orthoimagery. Most countries had 100 % coverage, and those that reported less than 100 % typically had an accuracy greater than or equal to 0.5 m. Resolution was typically either 0.5 m or 1 m, but the trend seems to move towards more accurate orthoimagery. The source of the orthoimagery was aerial photographs. Two countries mentioned that some of the imagery was available in colour.

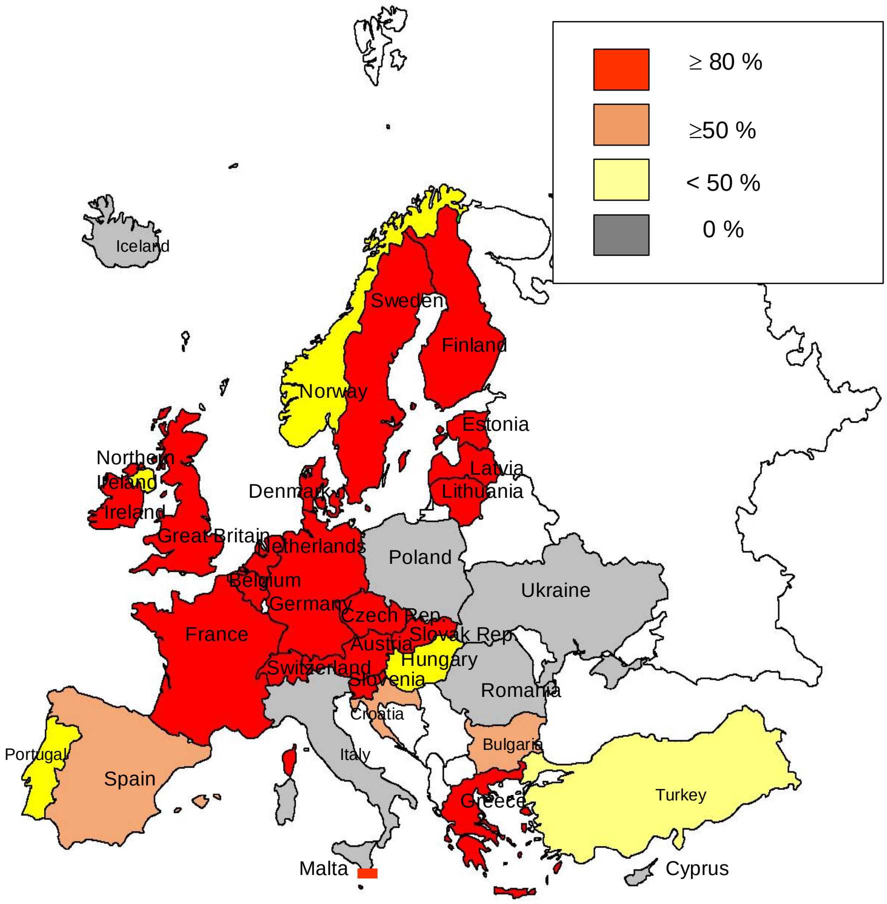
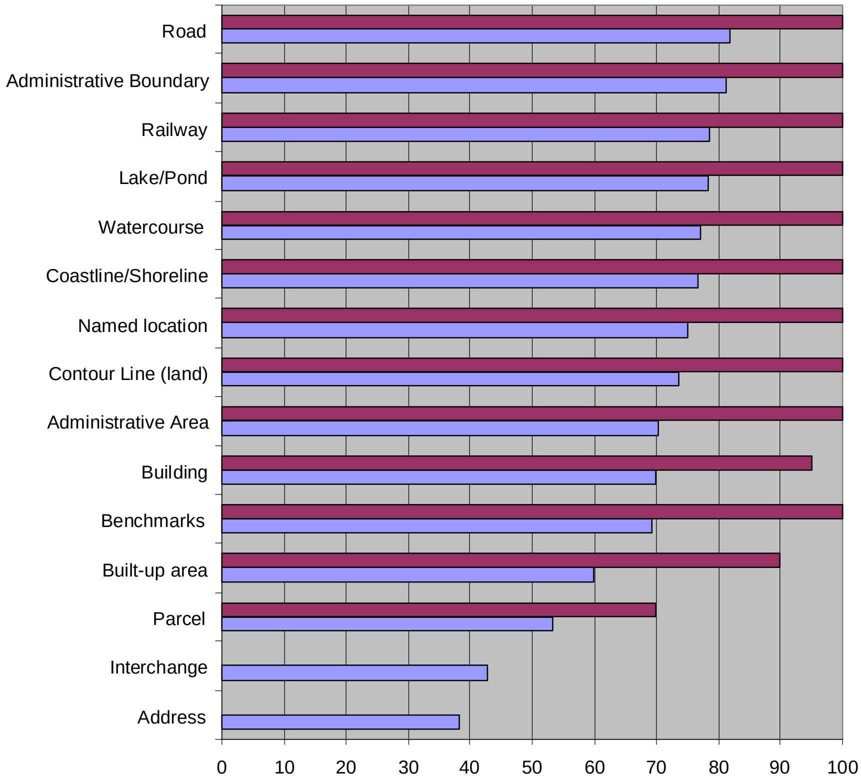


Figure 4. Coverage of orthoimagery

AVAILABILITY AND COVERAGE OF FEATURES

For each reference theme, we selected some feature types using the FACC data dictionary, and we gave a definition of each feature type using FACC definition. We examined availability, coverage, maintenance, descriptions, attributes, positional accuracy and the source for roads, railways, lakes/ponds, watercourses, coastlines/shorelines, named locations, contour lines, administrative boundaries, cadastral parcels and benchmarks. In this context availability means that a feature type exists in a dataset or that it could be derived by some method (e.g. generalization). It does, however, not mean that the whole country is covered. For most feature types, there seemed to be a corresponding feature type at the national level. Respondents felt that roads, administrative boundaries, watercourses, lakes/ponds and contour lines were either directly available or derivable from the datasets. Here we did not take into account that two countries do not have any railways in reality or that 5 countries do not have a coastline. Altogether the availability was over 60% for all other feature types except addresses that had 48% availability. Median availability of a feature type was 82%. Non-availability of built-up areas, interchanges and addresses was 15% - 21%.

The coverage of the feature types in the countries covered is presented in Figure 5. Typically roads, administrative boundaries, railways, lakes/ponds, watercourses, coastlines/shorelines, named locations, contour lines, administrative areas, and benchmarks are covered in the whole country. However, some of the countries do not have railways (Iceland, Malta, Cyprus) and some do not have coastlines (Austria, Czech, Hungary, Slovak Republic, Switzerland) in reality, and here we have changed the coverage to 100% in these feature types. Coverage values are typically close to 100% or 0%. Therefore average and median values do not reflect a typical situation when there is considerable deviation in the values.



	Addres s	Interchange	Parcel	Built-up are	Bench mar	Building	Administra	Contour Line	Named	Coastline/S	Watercour	Lake/Pond	Railway	Administra	Road
Median	0	0	70	90	100	95	100	100	100	100	100	100	100	100	100
Average	38,2	42,9	53,3	59,8	69,4	69,9	70,3	73,6	75	76,7	77,1	78,3	78,6	81,2	82

Figure 5. Coverage of feature types in countries

We can conclude that those countries that have feature types available have very good coverage (median 100 % in all feature types). Interestingly the order of the feature types changes when 0% is not counted. It seems that benchmarks, administrative boundaries and areas, coastlines/shorelines, named locations, roads, addresses and railways typically cover the whole country if available.

We also calculated the coverage of all feature types at the national level. For topographic feature types (indicated in Figure 6 in red) we included coastlines/shorelines, watercourses, lakes/ponds, contour lines (land), built-up areas, roads, interchanges and railways. Each feature type was given the same weight, while in reality the importance of a feature type varies. Again when a country did not have a coastline or railway, we used 100 % coverage. Some countries did not provide any information at feature type level, and their position in the table is therefore rather low.

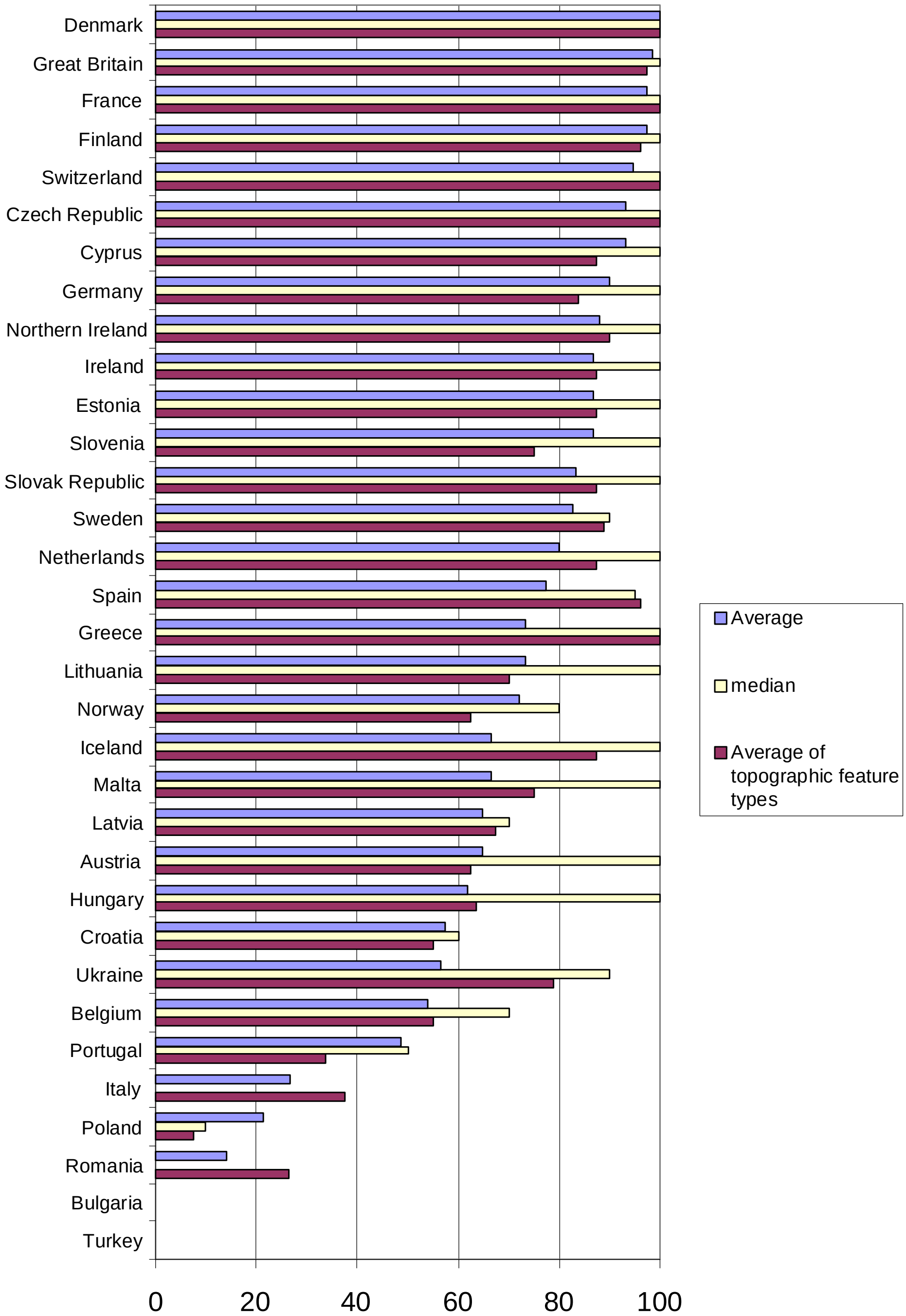


Figure 6. Coverage of all feature types at national level

Maintenance of feature types

Address information is the feature type updated most frequently in the countries. 73% of the countries update these data continuously. Parcels and benchmarks come second and third. If we also consider annual and bi-annual update frequencies, then the most updated feature types are: addresses, administrative boundaries, administrative areas, parcels, interchanges, roads and buildings. Benchmarks and buildings are either updated very frequently or every few years. The reason is presumably that those feature types are either part of the topographic or cadastral datasets. Coastlines/shorelines, Watercourses, Lakes/Ponds, Contour lines, Built-up areas, Roads and railways are typically updated every 5 - 10 years.

DISCUSSION

NMCAs play a key role in providing reference themes in Europe. According to our survey 92% of the datasets were produced by a NMCA. Topographic databases were an important source for many reference feature types. Some of the countries still have information that is based on scanned maps, and in these countries updating presents a real challenge. It was evident that most of the countries had not done any co-operation to guarantee that information is continuous at national borders. The level of detail and attribute information varies in the countries. There is a clear need for harmonised specifications in order that national data for use on European level can be provided. The GiMoDig (Geospatial info-mobility service by real-time data integration and generalisation) project, within the framework of which methods for delivering geospatial data to the mobile user by means of real-time data integration and generalisation were developed and tested, has demonstrated that a common data model for topographic data is feasible at some level (11). The results of the questionnaire reffered indicates that many countries have similar feature types, but we could not actually compare the semantics or how the feature types actually represent the real world. The author has suggested a solution that entails a multi-tier harmonization of data sets at the global, European, national, regional and local level (12). At the national level in Finland, the implementation of the National SDI Policy (13) has begun, and here the harmonisation of core datasets has a key role. Eurogeographics has begun to build common specifications (14), and as a result of the new proposed directive for spatial information INSPIRE (Infrastructure for Spatial Information), the specification of the reference themes will start.

CONCLUSIONS

NMCAs produce most of the reference themes at the national level. Other organizations produced 8% of the databases in the reference themes covered. Some of the themes were entirely produced by NMCAs, such as hydrography, transportation and contour lines. If we also include separate cadastral agencies, then administrative and cadastral parcel will be included in the list. Most of the reference themes are included in topographic or cadastral databases. Countries have begun introducing new technology and converting the reference databases to object-based databases. The majority of the cadastral databases are already object-based and about 40% of the topographic databases are object-based. Data content has not yet been harmonised on a transnational level, nor has geometric interoperability at the national borders been achieved. ISO TC 211 standards are not yet implemented, but the issue is being addressed.

Although topographic data sets existed in every country, some countries did not appear to have a common database. We inferred that a topographic database existed in 23 countries. The most typical scale was 1:10 000. Cadastral databases existed in 23 countries. Digital elevation models were produced in 26 countries with a typical resolution of 10 to 20 metres. Orthoimagery was available in 27 countries, typically with a 0.5 or 1 meter resolution.

The questionnaire was set out to determine the availability and descriptions of some feature types. Most feature types were available at the national level, and a corresponding feature type could be described. Availability was over 60% for all other feature types, aside from addresses with 48% availability. Median availability for a feature type was 82%. We conclude that benchmarks, administrative boundaries and areas, coastlines/shorelines, named locations, roads, addresses and railways typically cover the whole country if available. Update frequency is dependent on the feature type. Addresses were updated continuously by 73% of the countries. The most up-to-date feature types were addresses, administrative boundaries and areas, cadastral parcels, interchanges and roads, and buildings.

REFERENCES

1. A. Jakobsson (ed.): Reference Data Sets and Feature Types in Europe. Eurogeographics, at http://www.eurogeographics.org/eng/documents/ref_data_ver101_part_A.pdf (2005) (accessed April 28th, 2005)
2. European Commission: Proposal for a Directive of the European Parliament and of the Council establishing an infrastructure for spatial information in the Community (INSPIRE). European Commission, at http://inspire.jrc.it/sdic_call/EN.pdf (2004) (accessed April 28th, 2005)
3. D. Rase, A. Björnsson, M. Probert, M-F. Haupt (eds): Reference Data and Metadata Position Paper. Eurostat, at http://inspire.jrc.it/reports/position_papers/inspire_rdm_pp_v4_3_en.pdf (2002) (accessed April 28th, 2005)
4. DGIWG: FACC data dictionary. at http://www.digest.org/DownloadDigest.htm (accessed April 28th, 2005)
5. Eurogeographics: EuroRegional product specification. Eurogeographics (2003) at http://www.eurogeographics.org/eng/documents/D51_ProductSpecification_v31.PDF (accessed April 28th, 2005)
6. Eurogeographics: SABE Seamless Administrative Boundaries of Europe, User Guide. Eurogeographics at, http://www.eurogeographics.org/eng/documents/ug2001v11_final.pdf (accessed April 28th, 2005)
7. A. Illert, S. Afflerbach: Global Schema Specification, GiMoDig project (2004), at http://gimodig.fgi.fi/pub_deliverables/Gimodig_D531.pdf (accessed April 28th, 2005)
8. ISO: ISO 19110 Geographic Information - Methodology for feature cataloguing. 55 pages (2005)
9. ISO: ISO 19115 Geographic Information –Metadata, ISO First edition. 140 pages (2003)
10. ISO: ISO/DIS 19131 Geographic Information -Data Product Specification (2004)
11. S. Afflerbach, A. Illert, and T. Sarjakoski, 2004: The Harmonisation Challenge of Core National Topographic Databases in the EU-Project GiMoDig. Proceedings of the XXth ISPRS Congress, July 2004, Istanbul, Turkey. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXV(B4/IV):129-134, at http://www.isprs.org/istanbul2004/comm4/papers/328.pdf (accessed April 28th, 2005)
12. A. Jakobsson: Framework and Requirements for Management of Topographic Data in Europe. In: Virrantaus, K. and H. Tveite (eds.), ScanGIS'2003 Proceedings of the 9th Scandinavian Research Conference on Geographical Information Science, June 4-6, 2003, Espoo, Finland, pp. 91-102. An electronic version available at http://www.scangis.org/scangis2003/papers/
13. National Council for GI: National Strategy for Geoinformation. Publications of the Ministry of Agriculture and Forestry 10/2004 (in Finnish)
14. C.Luzet, N.Land, H.van der Vegt: Eurospec – A Cornerstone for the Building of the European Spatial Data Infrastructure. Proceedings of the XXth ISPRS Congress, July 2004, Istanbul, Turkey. International Archives of Photogrammetry, Remote Sensing and Spatial Information Science, at http://www.isprs.org/istanbul2004/comm4/papers/537.pdf (2004) (accessed April 28th, 2005)

The author, Mr Antti Jakobsson, is employed by the National Land Survey of Finland (NLS). He is working on the development of national mapping, data quality and quality management. He is also the chairman of the EuroGeographics Expert Group on Quality. Mr Jakobsson has participated in the work of the work groups of ISO TC 211 developing the ISO 19113 and 19114 standards. He is a postgraduate student at the Helsinki University of Technology, and is currently preparing a doctoral thesis on data quality and organizational and quality management issues regarding topographic datasets.