

# STATUS OF DATA QUALITY IN EUROPEAN NATIONAL MAPPING AGENCIES

*Par Mr Antti JAKOBSSON*

*Chair, EuroGeographics Working Group on Quality - National Land Survey of Finland*

*PO BOX 84 FIN-00521 Helsinki - Finland*

*Tel. +358 20541 5177*

*Fax. +358 20541 5070*

*E-mail: antti.jakobsson@nls.fi*

*Dr François VAUGLIN*

*Secretary, EuroGeographics Working Group on Quality until 2001*

*COGIT Laboratory - IGN 2/4 avenue PASTEUR 94165 Saint-Mandé Cedex - France*

*Fax: + 33 1 43 98 81 71*

*E-mail: Francois.Vauglin@ign.fr*

---

## Résumé

EuroGeographics regroupe des représentants des agences cartographiques européennes. Sa mission est de représenter ces agences et de coordonner leurs efforts pour le développement d'une infrastructure d'information géographique au niveau européen. Plusieurs groupes de travail ont été créés pour l'échange d'information et la discussion de sujets liés à l'information géographique -et, notamment, un groupe de travail sur la qualité. Ce groupe s'intéresse à trois thèmes : le management de la qualité et ISO 9000, la normalisation, la qualité des données. Le sous-groupe sur la qualité des données a pour but le partage d'expériences et des méthodes utilisées pour la rédaction des spécifications des BDG, la mesure de la qualité, et l'évaluation des besoins des utilisateurs et de l'adéquation des produits à ces besoins.

Ce sous-groupe de travail a utilisé le cycle de vie des produits comme base pour étudier l'ensemble des questions liées à la qualité des données et pour préparer un questionnaire touchant à chaque étape.

Le questionnaire, qui fut rédigé en 1999, a reçu les réponses de 18 pays, dont les USA. Cet article relate les résultats de ce questionnaire.

226 jeux de données sont concernés par les réponses au questionnaire. Toutes les agences cartographiques ayant répondu produisent des jeux de données topographiques. La moitié produit aussi des jeux cadastraux. Le questionnaire couvrait aussi d'autres types de données. Généralement, les motivations de la production de données sont divisées selon des obligations légales ou des obligations commerciales. Les besoins des utilisateurs ont été identifiés principalement à l'aide d'enquêtes auprès des utilisateurs.

La question de la spécification et de la production des jeux de données est présentée dans l'article. Globalement, les jeux de données produits répondent à des spécifications de produits, mais il y a un clair besoin de standardisation. Les cartes papier sont toujours considérées comme la source principale de saisie. Les agences cartographiques définissent des critères généraux sur la qualité des données, mais beaucoup indiquent que des critères détaillés seront utilisés dans un avenir assez proche. Le sous-traitance de la production semble relativement généralisée (les 3/4 des agences ont recours à la sous-traitance). De manière surprenante, la baisse des coûts n'est pas la raison la plus souvent invoquée pour justifier la sous-traitance, mais c'est l'accélération de la production.

Finalement, les procédures d'évaluation de la qualité et, plus généralement, les questions qui touchent la qualité des données sont présentées dans l'article. Les agences savent bien que des normes sur l'information géographique existent, mais le besoin d'éducation sur l'évaluation de la qualité se fait sentir. Quelques composantes de la qualité et leurs indicateurs ne sont pas bien connus. Garantir la qualité des données à l'utilisateur reste une question majeure. Quoi qu'il en soit, les résultats de l'évaluation de la qualité ne sont généralement pas transmis à l'utilisateur; ce qui implique que les procédures d'évaluation ne fonctionnent pas correctement. Les attitudes actuelles sur la qualité et les besoins de spécification adaptée sur la qualité indiquent que des solutions aux questions qui restent posées sont envisageables dans le futur.

## Abstract

EuroGeographics (formerly CERCO) brings together representatives from the European National Mapping Agencies (NMAs). Its mission is to represent Europe's NMAs and co-ordinate their efforts to develop the European Geographic Information Infrastructure. It has created several working groups for exchanging information and discussing issues relating to geographical information – including the Working Group on Quality. This group is considering three subjects: 'Quality management' and 'ISO 9000', 'Data quality issues and Standards'. The Sub-Working Group on Data Quality (SWG) aims to share experiences in methods used to write dataset specifications, measure quality, assess user requirements and evaluate fitness for use.

The SWG used product life cycle as a basis for further study of other issues relating data quality and to prepare a questionnaire based on different steps in the product life cycle.

The questionnaire on data quality was prepared in 1999 – and received answers from 18 countries, including the USA. This paper is based on the results of the questionnaire.

The questionnaire covered a total of 226 datasets. All NMAs produce topographic datasets and about half also produce cadastral datasets. The questionnaire also covered other datasets produced by the NMAs. The reasons for producing the datasets was divided between legal obligations and market reasons. The user requirements were identified mainly through customer interviews.

The specifications and production of the datasets are discussed. Generally, datasets have a product specification, but there is a clear need for standardization. Graphical maps are still considered to be the main source of datasets. NMAs have defined general criteria for quality of datasets, but detailed criteria will be used in the future. Subcontracting the production appears to be quite common. Nearly 75% of the NMAs subcontract part of their production. Surprisingly, cost is not the major reason cited, but rather the acceleration of production.

Finally, quality evaluation procedures and data quality issues are discussed. The NMAs recognize the geographic standards are being developed, but evidently some education is needed to evaluate quality. Some quality elements and subelements are not well known. Assuring customer of data quality appears to be a major issue. However, quality results are not forwarded to customers, which implies that evaluation procedures are not functioning properly. Current attitudes on quality and the need for proper quality specification indicate the possibility that remaining issues will be solved in the future.

---

## 1. Introduction to EuroGeographics and the Working Group on Quality

EuroGeographics (formerly CERCO, Comité Européen de Responsables de la Cartographie Officielle) brings together representatives from the European National Mapping Agencies (NMAs). Its mission is to represent Europe's NMAs and co-ordinate their efforts to develop the European Geographic Information Infrastructure. Currently, it has members from 37 nations. EuroGeographics has two products:

- the Geographical Data Description Directory (GDDD) – a metadata service available on the Internet ([www.eurogeographics.org](http://www.eurogeographics.org)), and
- a dataset of administrative areas of Europe (SABE, the Seamless Administrative Boundaries of Europe)

It also has a number of on-going projects covering global and pan-European initiatives as EuroGlobalMap (1:1 million) and EuroRegionalMap (1:250,000).

It has created several working groups for exchanging information and discussing issues relating to geographical information. One of these, the Working Group on Quality, was created in September 1997 and is working on three main subjects: 'Quality management and ISO 9000', 'Data quality issues' and 'Standards'. The mission of the Sub-Working Group on Data Quality (SWG) is to share experiences in methods used to write dataset specifications, measure quality, assess user requirements and evaluate fitness for use.

The Working Group has created a network of experts dealing with quality management and data quality among member organizations. It has produced two guidebooks: Good Reasons for Implementing a Quality Management System and a Handbook for Implementing a Quality Management System. This paper focuses on the report covering the questionnaire on data quality made to the member organizations in 1999.

## 2. Introduction to the data quality questionnaire

The SWG organized a workshop in Paris in 1998. As a result of this, the SWG used product life cycle as a basis for further study of different issues regarding data quality. The approach was presented by Dassonville, Vauglin, Jakobsson and Luzet (1999).

After comparison of the NMAs' production organizations, 5 major steps were identified:

1. To identify user requirements
2. To specify products meeting the requirements of step 1
3. To produce data meeting the specifications of step 2
4. To check what has been correctly produced in step 3
5. To deliver data checked in step 4 to users and to prepare for testing if requirements were correctly identified in step 1.

The objectives of the questionnaire were:

- To determine the major data quality issues for NMAs
- To share information on data quality procedures used at the NMAs in a product's life cycle
- To arrange seminars on data quality that would focus on issues in which the NMAs need more information.

The questionnaire was posted to the NMAs by mid-July 1999, and answers were requested by September 1999. The first draft report was presented at the Working Group meeting in Ankara in November 1999. The final result was presented in connection with the Working Group seminar in Stockholm 2000.

### 3. Material and methods

The questionnaire was posted to the NMAs in Europe; there were 21 answers received from 18 countries. Of the answers, 20 were from Europe and one was received from the USA. The NMAs were:

- Bundesamt für Kartographie und Geodäsie (BKG), Germany
- Czech Office for Surveying, Mapping and Cadastre (CUZK), Czech Republic
- Estonian National Land Board, Estonia
- Federal Directorate of Cadastral Survey, Switzerland
- Federal Office of Metrology and Surveying (BEV), Austria
- Federal Office of Topography (BLT), Switzerland
- Ministry of National Defence, General Command of Mapping (GCM), Turkey
- Geodesy, Cartography and Cadastre Authority of the Slovak Republic, (ÚGKK)
- Institut Cartogràfic de Catalunya, Spain
- Institut Géographique National (IGN-F), France
- Institute of Aerial Geodesy, Lithuania
- Institute of Geodesy, Cartography and Remote Sensing (FÖMI), Hungary
- Instituto Geográfico Nacional (IGN) España, Spain
- Kort & Matrikelstyrelsen (KMS), Denmark
- National Land Survey of Finland (MML)
- National Land Survey of Sweden (Lantmäteriet)
- Ordnance Survey (OS), Great Britain
- Portuguese Institute for Cartography and Cadastre (IPCC), Portugal
- Surveying and Mapping Authority of the Republic of Slovenia (GURS)
- Topografische Dienst Nederland (TDN), The Netherlands
- United States Department of Commerce, Bureau of the Census, USA

## 4. Results

Here we present shortly the main results; a full analysis can be found in the report, which will be made public later this year on the EuroGeographics website.

### 4.1 Datasets and user requirements

All of the NMAs produce topographic datasets and about half also produce cadastral datasets. Other datasets

produced include geodetic datasets, address files, land-use data and geologic data. In all, the NMAs identified a total of 226 different datasets produced.

The reasons for producing the datasets could be divided into legal obligations and market reasons. User requirements were identified mainly through customer interviews.

### 4.2 Dataset specifications and production

Most datasets have a product specification. Product specifications are based on a methodology developed in-house. Currently, there is no international standard on how to create specifications. Some datasets are based on international product specifications such as the Corine, the Digest etc. Product specifications may include:

- Data dictionaries
- Content specifications
- Data Capture specifications
- Technical specifications
- User guides.

The reference set used as a comparison for the dataset when checking data quality, can be the real world or other datasets, graphical maps or other sources. Graphical maps and the real world are used as references. Sometimes both the real world and other sources are used as a reference; e.g. a graphical map can be used in the data compilation process and then the final product is checked in the field. If a graphical map is the only source used in quality evaluation, it can mean that the currency of the dataset cannot be evaluated. In the future, when datasets are updated the real world should be used as a reference.

15 NMAs (72%) are currently subcontracting some parts of their production. Some NMAs, 3(14%), subcontract only very minor parts, while 3 NMAs do not subcontract at all. Aerial photography (57%) and data compilation (48%) are the most subcontracted processes in the production of base topographic datasets, others include quality evaluation etc. Quality evaluation is typically subcontracted in about 5-10% of dataset types. The entire production is subcontracted most in cadastral (19%) and base topographic data (14%), and least in derived topographic (5%) and imagery (5%) datasets. The reasons for subcontracting include:

- Production acceleration (57%)
- Lack of resources (52%)
- Political decision (43%)
- Cost benefit (19%).

The requirements given to subcontractors include: technical specifications, legal contracts and qualification of the subcontractor. Aerial photography and data compilation are most likely subcontracted. Topographic base datasets are subcontracted more than the others.

### 4.3 Quality elements

We used the quality elements or parameters as defined in the European Prestandard (ENV 12656) and in the ISO 19113 Draft International Standard. The question was: *Which quantitative quality elements and parameters (according to CEN/TC 287 and ISO TC 211 definitions) do you use? / Give also the quality measure that you use.* Table 1 presents the results.

Table 1. Use of quality elements

| <i>Quality element</i>                                                                                                                                                                                                                                                                                        | <i>% of NMAs using</i>                      |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------|
| <b>Nonquantitative quality elements</b>                                                                                                                                                                                                                                                                       |                                             |
| Purpose                                                                                                                                                                                                                                                                                                       | 57                                          |
| Usage                                                                                                                                                                                                                                                                                                         | 48                                          |
| Lineage                                                                                                                                                                                                                                                                                                       | 71                                          |
| <ul style="list-style-type: none"> <li>▪ Source</li> <li>▪ Process step</li> <li>▪ Producer organization</li> <li>▪ Date of production</li> <li>▪ User-defined</li> </ul>                                                                                                                                     | 81<br>48<br>62<br>57<br>5                   |
| User-defined                                                                                                                                                                                                                                                                                                  | 5                                           |
| <b>Quantitative quality elements</b>                                                                                                                                                                                                                                                                          |                                             |
| <b>Completeness</b>                                                                                                                                                                                                                                                                                           | 52                                          |
| <ul style="list-style-type: none"> <li>▪ Omission</li> <li>▪ Commission</li> <li>▪ User-defined</li> </ul>                                                                                                                                                                                                    | 43<br>38<br>5                               |
| <b>Logical consistency</b>                                                                                                                                                                                                                                                                                    | 62                                          |
| <ul style="list-style-type: none"> <li>▪ Topological consistency, Format consistency</li> <li>▪ Domain, Geometric, Semantic consistency</li> <li>▪ User-defined</li> </ul>                                                                                                                                    | 52<br>43<br>5                               |
| <b>Positional accuracy</b>                                                                                                                                                                                                                                                                                    | 62                                          |
| <ul style="list-style-type: none"> <li>▪ Absolute accuracy</li> <li>▪ Relative accuracy</li> <li>▪ Relative horizontal</li> <li>▪ Relative vertical</li> <li>▪ Adjacent entities</li> <li>▪ Gridded data positional accuracy</li> <li>▪ Accuracy of geographic identifiers</li> <li>▪ User-defined</li> </ul> | 67<br>29<br>43<br>33<br>24<br>19<br>24<br>5 |
| <b>Temporal accuracy</b>                                                                                                                                                                                                                                                                                      | 38                                          |
| <ul style="list-style-type: none"> <li>▪ Last update</li> <li>▪ Temporal validity, Temporal consistency</li> <li>▪ Time measurement, Rate of change</li> <li>▪ Temporal lapse</li> </ul>                                                                                                                      | 52<br>19<br>14<br>10                        |
| <b>Thematic accuracy</b>                                                                                                                                                                                                                                                                                      | 43                                          |
| <ul style="list-style-type: none"> <li>▪ Quantitative attribute, Qualitative attribute, Classification correctness</li> <li>▪ Accuracy of spelling</li> <li>▪ Agreement for an attribute</li> <li>▪ Likely misclassification, Consistency of abbreviations</li> </ul>                                         | 38<br>29<br>24<br>19                        |

In essence, the terminology used appears to be unclear to the NMAs. Some subelements are used more than the element itself, and some others are not standardized in the ISO 19113. The use of very common quality elements appeared to be quite low, e.g. positional accuracy is used only 62% of the time.

#### 4.4 Quality evaluation

The question was: *Which quality elements and parameters do you test and what types of methods do you use to evaluate quality?* The NMAs responded that positional accuracy is evaluated in 71%, completeness in 63% and thematic accuracy in 46% of the dataset types produced. Temporal accuracy and aggregate results are not commonly evaluated.

Sampling methods are not based on the use of standards: 67% of the NMAs use no standards while only 19% do. The questionnaire listed 2 standards (ISO 2859 and ISO 3951) that could be used. ISO 2859 is used by 1 NMA and 2 use their own standard; 1 uses a standard, but does not specify which one. The different sampling methods that are used in

- Simple random sampling 36%
- Stratified sampling 45%
- Multistage sampling 9%
- Other 1%

Aggregation is not used. Only 1 NMA stated the usage of aggregation for themselves and for their customers.

Quality criteria, including target value(s) for quality, are used in 58% of the dataset types. Cadastral and imagery datasets use commonly defined quality criteria.

#### 4.5 Data delivery

71% of NMAs offer no warranty for quality. Only 29% of NMAs offer warranties covering aspects such as technical correctness, legal responsibility and correctness of delivery. 59% of the datasets are delivered without quality results; imagery datasets more often include quality results (56%). 60% of users require quality results for some datasets. 80% or more users of cadastral datasets need quality results while 50% of users of derived topographic datasets need quality results.

#### 4.6 Opinions

In the questionnaire some statements were given to determine the general opinions regarding quality in NMAs. International Cartographic Association (ICA) used the same statements earlier (Östman 1997). There is very strong consensus that NMAs have recognized the need for proper data quality specification (mean 4,1) and for educational material concerning data quality specifications and assurance routines (mean 4,2). The average for the following statements was about 3. Figure 1 depicts the results.

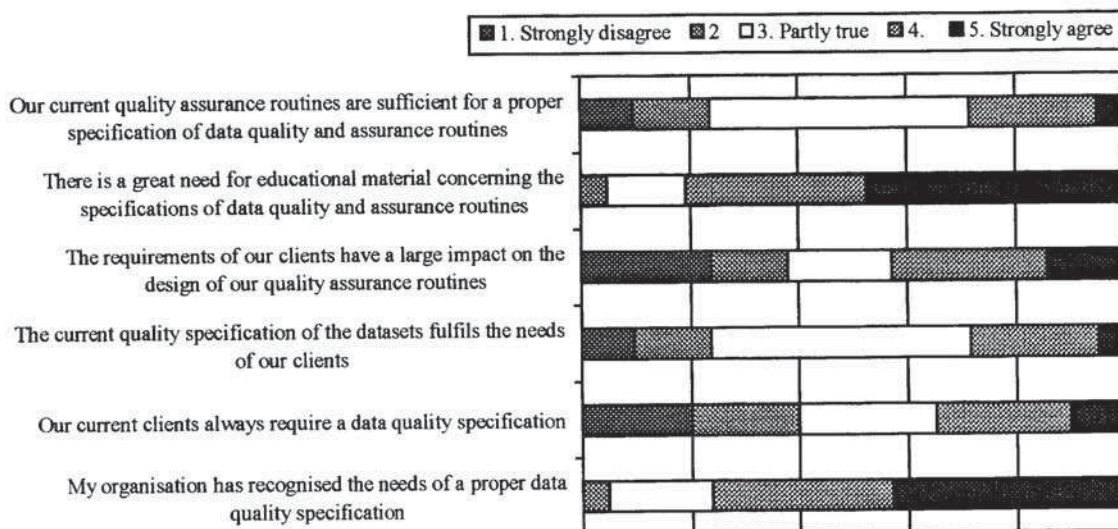


Figure 1. Opinions regarding quality

### 5. Discussion

In December 1996 the ICA and its Commission on Spatial Data Quality posted a worldwide questionnaire to

producers of spatial data. The questionnaire went to 288 NMAs, and a total of 56 responded, most from Europe (59%). Table 2 gives a short summary of the results (Östman, 1997).

Table 2. Results of the ICA questionnaire: percentage of answers indicating subjective evaluation or no evaluation as a quality assurance routine (Östman, 1997)

| Quality parameters  | Percentage |
|---------------------|------------|
| Positional accuracy | 43%        |
| Thematic accuracy   | 48%        |
| Temporal accuracy   | 68%        |
| Completeness        | 56%        |
| Logical consistency | 32%        |

When we compare these results to our results regarding quality evaluation answers, it appears that positional accuracy, and completeness are more commonly evaluated now than before, although the difference is not very clear.

The questionnaire raised the following key issues among the NMAs:

- Usage of standards is not very common and the issue needs to be raised
- Need for better process management, especially in quality evaluation processes
- Need for data management improvements, especially in data specification
- Need for better metadata, e.g. quality results
- Training of personnel.

NMAs need to determine how forthcoming geoinformation standards should be used in their processes, especially data administration standards covering product specification, quality and quality evaluation.

The questionnaire showed that there is great need for improving the current situation. As the examples show, NMAs in Europe understand the importance of quality, but so far have not really invested in quality evaluation. The reasons are, of course, that the market has not demanded more and NMAs have received funding from government budgets. However, now that the geographic information market is evolving and datasets are coming into everyday use, the role of the customer is changing. Customers are no longer professionals who know how the information is gathered and what to expect. NMAs therefore must take quality seriously.

Quality management and quality information together constitute a significant competitive advantage for producers. Furthermore, better quality often saves costs and increases process output.

#### Acknowledgements

The authors wish to acknowledge the experts of the EuroGeographics Working Group on Quality and especially all the National Mapping Agencies that responded to the questionnaire.

#### Authors' contributions

The main author is Mr. Antti Jakobsson who is the Chairman of the EuroGeographics Working Group on Quality and the SWG. He has been responsible for preparing and analyzing the questionnaire on data quality. He is currently employed by the National Land Survey of Finland (NLS) and was chief developer of the Topographic Database System in 1992. Currently he is working with quality management issues in the NLS. He is a postgraduate student in the Helsinki University of Technology and is currently preparing a doctoral thesis.

François Vauglin was responsible for analyzing the questionnaire together with Antti Jakobsson. He is working at the French IGN-F. At the COGIT Laboratory, he is leading the research initiative on geographical data quality. He was a secretary for the CERCO Working Group on Data Quality 1999-2000. He graduated in 1993 as a survey engineer and has since worked on a PhD thesis on the mathematical modelling of positional accuracy, graduating as doctor in 1997. His current research interests are focused on geographical data quality, more specifically on the geometrical, semantic and topological aspects of quality.

#### References

- Dassonville, L., Vauglin, F., Jakobsson, A., Luzet, C. 1999, Quality management, data quality and users, metadata for geographical information, *In Proceedings of the International Symposium on Spatial Data Quality*, Hong Kong.
- ENV 12656, 1998, Adopted European Prestandard.
- ISO TC211, 2001, ISO/DIS 19113, Geographic Information – Quality Principles, Text for DIS 19113, The International Organization for Standardisation.
- ISO TC211, 2001, ISO/DIS 19114, Geographic Information – Quality Evaluation Procedures, International Standard Committee Draft.
- Jakobsson, A., 1999, Quality Evaluation of Topographic Datasets –Experiences in European National Mapping Agencies, *In Proceedings of the International Symposium on Spatial Data Quality*, Hong Kong.
- Östman Anders (1997). The specification and Evaluation of Spatial Data Quality, Proceedings of the 18<sup>th</sup> ICA International Cartographic Conference, pp. 836-847.