

Pan-European datasets: challenges and lessons from SABE

Authors: Heinz Bennat*, Ingrid Naumann*, Claude Luzet**, Nick Land**

* Bundesamt für Kartographie und Geodäsie (BKG - Federal Agency for Cartography and Geodesy) Frankfurt am Main, Germany; fax: +49 69 6333 441 email: sabe@ifag.de

**EuroGeographics; tel: +33 (0)1 64 15 32 39, fax: + 33 (0)1 64 15 32 19, email: nick.land@eurogeographics.org

Abstract:

The SABE (Seamless Administrative Boundaries of Europe) dataset is a well-known reference dataset of the wider Europe. Assembled from the official national datasets of more than 30 countries, SABE is the dataset that gives spatial meaning to a host of otherwise ‘*shapeless*’ data, such as population, economy, agriculture, environmental, statistical data. The SIRE database maintained by EUROSTAT (the European Commission Directorate for the Statistics), taking advantage of SABE data is a particularly good example of the kind of applications that SABE enables. The next update of SABE, currently in preparation, will coincide with the administrative situation of Europe for the latest census that was completed during the year 2001 for most European countries: It is to be a tool extremely precious both for the governance of Europe, and for the private sector (marketing, insurance, utilities, ...).

Creating, maintaining, distributing SABE has brought to EuroGeographics a few challenges on interoperability ‘*from local to regional*’, some of them successfully conquered, some others still being worked on. What has been easily conquered are the most technical aspects, which, while not at all trivial, have reasonable solutions: Geodetic reference systems, data model, logical and semantic models, format issues etc. One lesson learnt is that we need to work further on making national framework data product series more harmonised from the start, therefore making the pan-European needs directly addressed by maintenance of national data, as opposed to the nowadays prevailing model of post-harmonisation, by EuroGeographics or private sector players.

Challenges that are still on the agenda include the data and pricing policy for the SABE product, where harmonisation of the different national policies is a delicate handiwork. Another challenge is the definition of the national boundaries and their compatibility and topology with other datasets (EuroGlobalMap, EuroRegionalMap) that EuroGeographics is assembling: An issue of technical significance (lines comparison, cross-resolution interoperability, generalisation, ...), but also of high political sensitivity in some instances.

The lessons learned from SABE will be of great assistance for tackling the future challenges of the development of the European Spatial Data Infrastructure.

1 Introduction

As Europe becomes more closely linked, pan-European topographic data will become increasingly important. Multi-national companies and international organisations need to be able to monitor their activities using reliable and comparable cross-border data. On the political level the awareness of the possibilities of geo-information is rising. Especially usage of multi-national coverage of geographical data sets as a vehicle to support international marketing and free-trading will increase. Applications as environmental protection or route planning do not stop at national boundaries.

National Mapping Agencies (NMA) as the providers of basic topographic information have to cope with that trend. To provide a **legal and organizational framework** the European NMAs founded **EuroGeographics** on 1 January 2001 as the successors of **CERCO** (Comité Européen des Responsables de la Cartographie Officielle) and **MEGRIN** (Multipurpose European Ground-Related Information Network).

To meet the requirements of a common European market the European National Mapping Agencies joined in the association **EuroGeographics** have established several projects to integrate and harmonize pan-European geographic digital data sets.

- Seamless Administrative Boundaries of Europe (SABE), a pan-European dataset of administrative units compiled from national contributions,
- EuroRegioMap, a pan-European 1 : 250,000 scale topographic dataset using the experiences of PETIT,
- EuroGlobalMap, the European contribution to the Global Mapping Project.

All those projects aim at producing seamless, homogeneous and consistent digital datasets on the basis of standardized models with uniform spatial reference. Data acquisition is dominated by the principle of not digitizing maps again but to found strictly on the existing digital resources of the National Mapping Agencies. Thus, much importance is attached to the aspects of data integration and harmonization.

Experiences from managing the SABE project will be presented as well as technical problems of integrating and harmonizing datasets from different sources giving examples of problems and solutions with regard to different data models, scales, accuracy, semantic meaning, projection systems, exchange formats.

2 EuroGeographics

CERCO functioned as useful forum (a club of General directors of European National Mapping agencies) and mainly serves the purpose of information exchange with some working groups dealing with special subjects like legal and economic affairs, Geodesy and GPS, updating databases, and quality. The need to have a unique focal point with permanent staff for operational and commercial activities in the growing geoinformation market led in 1991 to the creation of MEGRIN. Because of different legal background only 19 of the 35 CERCO member NMAs were able to join MEGRIN, too. This was one of the reasons to start the evaluation about a reunion of both organizations in 1999. The outcome was the creation of **EuroGeographics** as an Association based on French law **to strengthen the co-operation between the European NMAs**. EuroGeographics continues with the MEGRIN projects and includes also the Working Groups of CERCO. In addition, EuroGeographics will support its members in convincing their own governments for adequate national GI policies and support the development of a European Geographic Information Infrastructure.

Membership in EuroGeographics is limited to one representative per country (Active Member) but additional organizations could become "associated member". The highest decision making body is the General Assembly (GA), which will meet at least once a year and elects by its active members the President and complementary members of the Management Board (MB). Active Members whose subscription for the coming year exceeds 10% of the total subscription voted for the said year each have the right to appoint a member to the Management Board for a two-year period (such are Germany, Great Britain and France). In addition four members will be elected by the Active Members of the General Assembly from among the representatives of the Active Members. The President is elected on a one-year-term from among the members of the MB by the Active Members.

The Management Board represents the Association in all circumstances and has full power to manage it. For the day-to-day management the MB mandated an Executive Director, who is supported by a Head Office Team located in Marne-la-Vallée near Paris. This team is responsible for general administration and co-ordination, representing the organisation at conferences, etc. external communication and supports and advises EuroGeographics projects and Expert groups (in past - Working groups).

The following **EuroGeographics projects** which are currently running could be summarized as activities to support the **building of a European Geoinformation Infrastructure** by

- contributing to the EC initiative INSPIRE (Infrastructure for Spatial Information in Europe)
- improving **access to geo information** by establishing **metadata services**,
 - like **GDDD** (Geographical Data Description Directory) created in 1994 to establish a descriptive listing of all the principal geographical databases available from the official National Mapping Agencies (NMAs) of Europe, LaClef, ESMI),
 - its successor **LaClef** aiming to include for example, a multilingual service, metadata of a higher resolution than at present, and a prototype electronic commerce service,
- providing **topographic information** by
 - **EuroRegionalMap**, a pan-European 1 : 250,000 scale topographic dataset
 - **EuroGlobalMap**, the European contribution to the Global Mapping Project, and
- providing information about **administrative boundaries** in the project **SABE** (Seamless Administrative Boundaries of Europe), maintaining a pan-European dataset of administrative units.

Although EuroGeographics's Members vary greatly in terms of their national responsibilities, they share a number of common interests. Whenever Members consider that joint discussions or project work could develop solutions or improvements, EuroGeographics **Expert Groups** were established. Their programme of work is agreed at the annual General Assembly.

Currently there are three Groups active:

- Experts on Legal and Commercial Issues.
- Experts on Geodesy producing the current European control frameworks (EUREF and EUVN).
- Experts on Quality Management Data Specifications and Standards exchanging experiences in improving such matters in both administrative and technical fields.

More information about the Expert Groups, projects, and EuroGeographics in general could be obtained from www.EuroGeographics.org

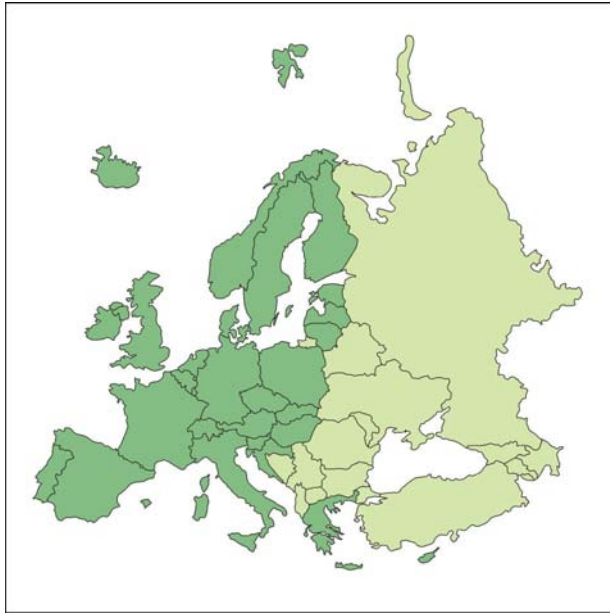
To **EuroGeographics**, its projects and Expert Groups, NMAs are contributing in various ways financial support, manpower and input in form of expertise and data sets. On the other hand NMAs benefit from the framework provided by EuroGeographics in gathering experiences in various fields, i.e. improved access to latest developments in cartography and surveying, qualifying of staff by working in an international environment; taking part in revenues from commercial activities, etc. In the following experience in multinational projects will be reported based on the EuroGeographics project SABE (Seamless Administrative Boundaries of Europe).

3 Experiences from the SABE project

SABE is EuroGeographics pan-European dataset of administrative units. The project started in 1992 and in the meantime versions showing the administrative situation of 1991, 1995 and 1997

are available and successfully commercialized by EuroGeographics. The production of a new version current with the situation at the 2001 European census dates is on the way.

Fig. 1: SABE97 Version 2.1 coverage



The SABE97 dataset covers the following countries:

Austria, Belgium, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Great Britain, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, The Netherlands, Northern Ireland, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland.

Coastline is delivered for:

Croatia, Finland, Germany, Great Britain, Greece, Ireland, Northern Ireland, Norway, The Netherlands, Poland, Sweden.



The dataset has been compiled from source data provided by Europe's official national mapping organisations with the best available semantic quality. In its latest version (SABE97 Version 2.1) it contains the geometry and semantics of the administrative hierarchies of 30 countries (see Figure 1): a total of over 100,000 polygons. Each country has its own specific administrative hierarchy composed of a different number of levels (see Table 1),. SABE contains all levels of national administrative hierarchies from the highest (country) to the lowest. SABE is the first pan-European boundary dataset available at this level of detail and assembled from high quality data and features.

Table 1: Number of most detailed administrative units and number of administrative levels in the SABE countries (Contents of SABE97 V2.1)

| Country | most detailed admin. units | No of Admin. levels |
|-----------------|----------------------------|---------------------|
| Austria | 2374 | 4 |
| Belgium | 593 | 5 |
| Croatia | 538 | 3 |
| Cyprus | 616 | 3 |
| Czech. Republic | 6234 | 4 |
| Denmark | 276 | 3 |
| Estonia | 247 | 3 |
| Finland | 452 | 3 |
| France | 36577 | 6 |
| Germany | 14739 | 6 |
| Great Britain | 10430 | 6 |
| Greece | 1034 | 4 |
| Hungary | 3152 | 3 |
| Iceland | 124 | 3 |
| Ireland | 3440 | 3 |

| Country | most detailed admin. units | No of administrative levels |
|------------------|----------------------------|-----------------------------|
| Italy | 8085 | 4 |
| Latvia | 590 | 3 |
| Liechtenstein | 4 | 2 |
| Lithuania | 56 | 3 |
| Luxembourg | 118 | 3 |
| Northern Ireland | 582 | 6 |
| Norway | 435 | 3 |
| Poland | 2489 | 4 |
| Portugal | 4018 | 4 |
| Slovakia | 3534 | 5 |
| Slovenia | 5973 | 3 |
| Spain | 8101 | 4 |
| Sweden | 288 | 3 |
| Switzerland | 2932 | 5 |
| The Netherlands | 572 | 3 |

SABE is delivered as individual country files which create a seamless and consistent dataset. The term **consistent** refers to the contents, to the structure, to geo-referencing, and time referencing of the data, although with so many independent data sources there are variations in the currency of the data. The term **seamless** means that there are no gaps or overlaps between polygons initially derived from different sources. Coordinates are two-dimensional, geographical in degrees (longitude, latitude) with decimal fraction. The spatial reference system is WGS 84 (ETRS89) with ellipsoid GRS 80. No map projection is applied.

Additional information could be obtained from the **SABE User Guide** that is available at the EuroGeographics Internet home page.

Organizational background

While MEGRIN projects were managed directly by the Central Team the management tasks of EuroGeographics projects are delegated to member NMAs. i.e. in the MEGRIN project SABE the Bundesamt für Kartographie und Geodäsie (BKG) was involved as a "Service Centre", responsible for the data harmonization only. With the change in project structures BKG role was transformed into the "Project Co-ordinator", which means the day-to-day management of the project was transferred from the Head Office to BKG. The experiences of both phases are summarized below.

Data harmonisation

The contributions provided by national mapping agencies have been transformed into a uniform structure and uniform positional reference system, line-filtered to a uniform resolution and are edge matched at international boundaries.

The harmonization process from national contribution to SABE data format is controlled by a conversion procedure consisting in several steps that are documented. The main steps are :

- Pre-processing: (check national contribution for completeness with regard to the requirements of the SABE data model)
- Transformation of coordinates from national map projection to geographic coordinates (WGS84).
- Processing of geometry to the SABE format (mosaic of lowest level administrative boundaries, harmonization with neighbouring countries, codes, names, administrative hierarchy, residence of authority, exclaves, condominiums)
- Generalization to map scale 1 : 100,000
- Final quality check

A brief view to the quality files of the 91, 95 and 97 version shows that the number of NMAs delivering the required format has increased from version to version. In parallel the number of inconsistencies detected decreased significantly.

For the 97 version for two countries a transformation from the national map projection and datum to SABE specifications was necessary. Seven contributions were generalized to SABE30 geometric resolution. Six countries were edited interactive to change geometric features and for six countries additional work was performed to correct the semantic information.

In two cases the conversion of diacritical characters to ISO 10646 Latin 1 needs additional editing. The integration of the Central and East European countries (CEEC) into future versions of SABE will increase the problems with diacritical characters. For the next version SABE offers to the NMAs to include the geographical names in their original language and the best suited ISO national character set in parallel to the ISO 10646 Latin 1 version.

Only for six countries inconsistencies were detected by the final check routines, mainly missing codes and names, which were clarified by the NMAs except in one case.

There are still differences in scales, reference systems and map projections used to derive national contributions to SABE. Due to its federal constitution even within Germany the scale of source data varies between the Federal States from 1 : 5,000 to 1 : 200,000.

The source data scales used for SABE deliveries varied between 1 : 5,000 and 1 : 750,000. More than 25 different reference systems and nearly 20 different map projections were used as sources for SABE contributions.

Most data delivery was in time. But the delay of one country could affect the complete harmonization process. It is obvious that additional delays were caused by incomplete or missing metadata information and necessary transformation parameters. The latest data delivery for SABE97 arrived at BKG in April 1998 and the first version was ready one month later, except one country. Due to problems with transfer format and data contents it needed additional 12 month to include this data set and SABE97 therefore was available on the market in May 1999. One of the project aims is to recover the costs of producing SABE. Each extra effort within the harmonization process leads to additional costs making it more difficult to reach this aim and hampers the up-to-date availability on the market improving the chances of SABE competitors from the private sector.

Project management tasks

With the re-organisation of projects parallel to the transition from MEGRIN to EuroGeographics one of the first tasks for BKG in its new role as Project Co-ordinator for SABE is to prepare the data delivery for the next SABE version (SABE2001) and to provide the legal background for working on behalf of the contributing NMAs and to suggest solutions for pre-financing the project costs. At the General Assembly at Malmö 2000 all heads of NMAs agreed to the continuation of SABE and promised to support the necessary activities. In December 2000 BKG distributed a first request to the NMAs to nominate a responsible contact person for NMA contributions to SABE2001. It was disappointing to notice how slowly NMAs respond on this request even after two reminders. About 20% of the NMAs did not answer until end of April 2001. One reason for this disappointing response time might be that the census dates in Europe are varying from 1st January 2001 until July 2002.

SABE marketing

The most important customer for SABE is Eurostat. SABE enables Eurostat to visualise and analyse statistical data in its spatial context for own Eurostat responsibilities but also as a service for other EC bodies.

Statistical analysis is also the main application of customers. One example is a big European car manufacturer who bought SABE to combine it with demographic data and consumer behaviour data to improve its marketing strategy and network of distributors.

A well known publisher of tourist guides used SABE as a basis for developing connections between hotels, restaurants, sight seeing attractions and towns to build a tourist information system.

The SABE pricing policy was revised in the beginning of 2002 and significantly decreased allowing more customers to obtain the data. EuroGeographics aims to avoid direct SABE competition to the products of the contributing NMAs. Therefore customers requesting only one country are advised to contact the responsible NMA. The SABE revenues in the past two years covered about 60% of the harmonisation costs. SABE is marketed via NMAs and valued added resellers (VARs), which are mainly private companies (except Eurostat), GI/GIS solution providers. Currently there are six SABE VARs and the number of partners will be increased in near future. The largest part of revenue still comes from sales via EuroGeographics Head Office, but income from distributors continues to grow. It is likely that the trend will continue in the future, i.e., distributors revenue will grow and EuroGeographics revenue will decrease. This is because distributors are able to provide

more value to SABE data (in combination with statistical data or by providing it in different formats) and to maintain better customers relations

Compared to products of private competitors SABE has several advantages, i.e the sources are the official administrative data sets of the NMAs and the status is well documented. In general SABE should be recognised as "definitive" administrative data set of Europe. It is constantly updated and supported by responsible public organisations (e.g. - NMAs)

Challenges

SABE is now a mature product delimited in future improvements by decisions in the very beginning. When the SABE production started there was no possibility to check the consistency of administrative borders and topographic features. Therefore harmonization is performed by a simple rule. The border of the state that delivered the higher resolution is used as basis for matching adjacent countries. This did not influence the usage of SABE for statistical purposes but if someone needs to combine SABE with topographic data there is no guarantee for consistency of borders and topographic objects.

However the issue was taken into account and harmonisation will be processed meeting the prerequisites for the new EuroGeographics projects EuroGlobalMap and EuroRegioMap which will provide topographic data sets at scale 1 : 1.000.000 and 1 : 250.000 respectively. Users will not accept if e.g. a border is permanently crossing a river where it is clear the middle of the river is the international boundary. Therefore for both projects the first step was to match international borders and topographic features.

Theoretically these new data sets could be used to derive improved SABE versions. But today this is impossible because EuroRegionalMap (scale comparable to the SABE high resolution version) covers only the regions of 6 NMAs and EuroGlobalMap (scale identical to the SABE low resolution version) will have in its first version only the highest administrative units.