

24 May 2005

English only

**Eighth United Nations Regional Cartographic
Conference for the Americas**

New York, 27 June-1 July 2005

Item 8 (b) of the provisional agenda*

**Reports on achievements in geographic information in
addressing national, regional and global issues, including
spatial data infrastructures**

**EuroGeoNames: the vision of integrated geographical
names data within a European SDI****

* E/CONF.96/1.

** Prepared by Sievers, Jörn and Zaccheddu, Pier-Giorgio (Germany).

EUROGEONAMES – THE VISION OF INTEGRATED GEOGRAPHICAL NAMES DATA WITHIN A EUROPEAN SDI

Sievers, Jörn and Zaccheddu, Pier-Giorgio

Federal Agency for Cartography and Geodesy (BKG), Richard-Strauss-Allee 11, 60598 Frankfurt am Main, Germany
E-Mail: pier.Zaccheddu@bkg.bund.de or joern.sievers@bkg.bund.de

1 INTRODUCTION

The first access to geoinformation is via geographical names. They are used for the search and overview on maps and in other spatially related products. They are considered to be the main search item within Geographic Information Systems (GIS) and search engines for spatially referenced information. Their clear and consistent use is important for administrative tasks in EC itself and in all European countries, for economy, postal services, telecommunication, health, risk management, safety and rescue services, transportation, tourism as well as for the purpose of popular education or for the use in the mass media.

Presumably, in some countries databases are already including geographical names data. They are mainly used by surveying and mapping authorities supporting processes in cartographic map production and they mainly consider national linguistic specifics. But, in some cases various feature categories of geographical names (e.g. populated places, administrative units, hydrography, mountains, islands, etc.) are very often not yet based on the same data model and are therefore not yet compatible with each other. Additionally, the distributed administrative responsibilities for specific names categories (e.g. statistical offices may be responsible for administrative units or hydrographic offices for hydrographic features) has to be taken into account. Some geographical names databases have not been created to satisfy cartographic purposes but more linguistic and other purposes, e.g. to preserve the cultural heritage. Furthermore, not all of these national files of geographical names may even be available in digital format, as some of them are published as simple analogue gazetteers. If they exist in digital form, these files may be stored in complex databases or in digital file lists (e.g. MS Excel files) or even in text documents (e.g. Word format). Additional toponymic attributes to geographical names, e.g. exonyms¹ (figure 1), the pronunciation, the language or script of geographical names are currently hardly available.

¹ Example for exonyms: a German could prefer to start his Internet search by entering the German name “Prag” (German exonym) instead of the Czech name “Praha” (endonym). E.g. the Italian exonym would be “Praga” and so on.

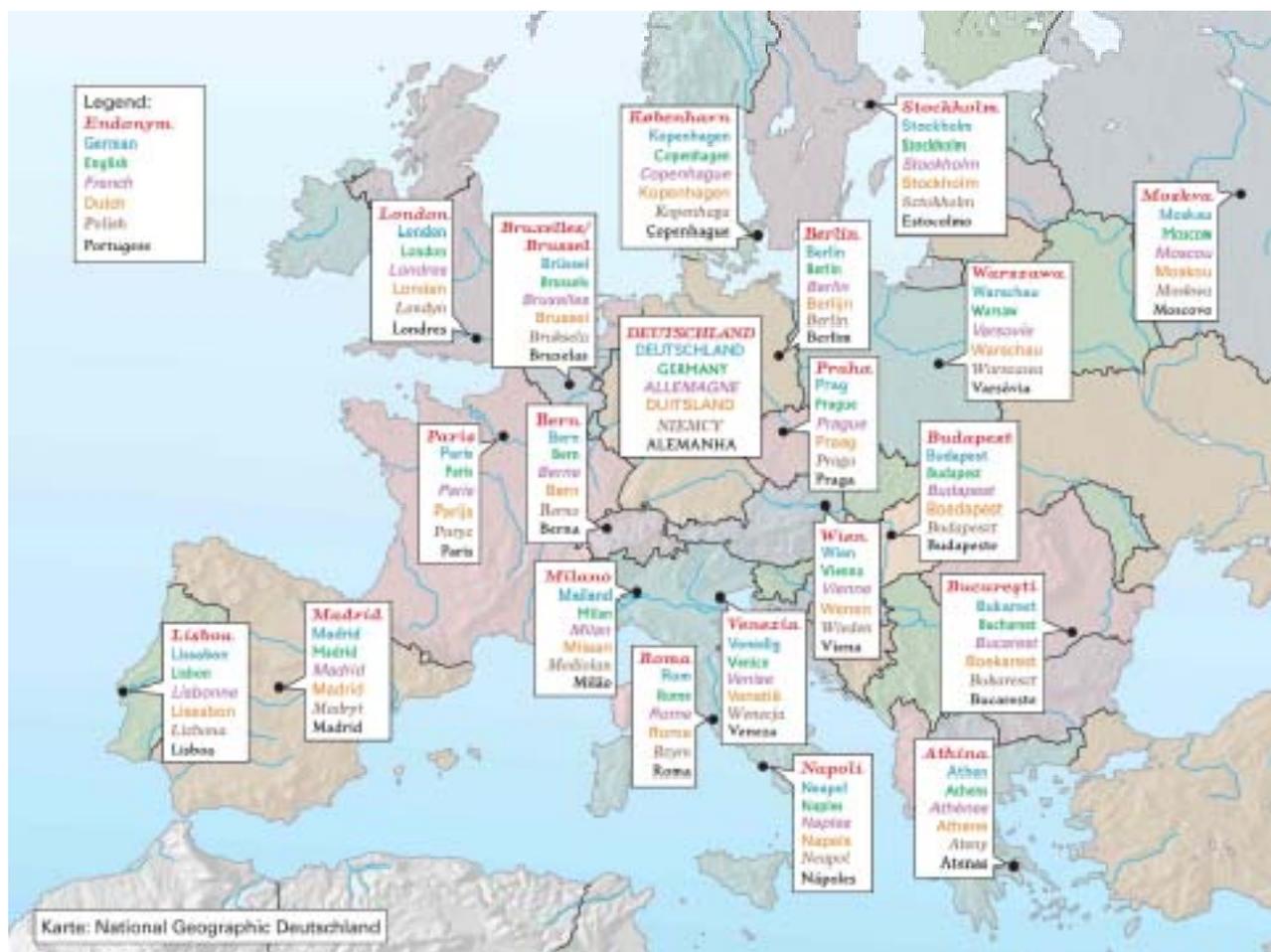


Figure 1 – Names of European cities spelt in their local language (endonym) and in other European languages (exonym)

2 CONCEPT OF THE EUROGEONAMES PROJECT

2.1 Motives for the EuroGeoNames project

INSPIRE is a legal framework being developed by the Commission services with officials and experts in Member States and accession countries from national, regional and local levels. It is to be implemented throughout the European Union (EU) from 2006 onwards with different types of geographical information gradually harmonised and integrated, resulting in a European Spatial Data Infrastructure (ESDI). In the context of the current INSPIRE initiative resulting in a European Spatial Data Infrastructure (ESDI), geographical names are considered to be one of the three most important data components (priority common basic data) [1]. EuroSpec is the EuroGeographics² programme that will prepare, influence, and contribute to the implementation of INSPIRE, from the member organisations perspective [2]. EuroSpec, with its related projects, such as EuroRoadS, RISE, etc. is already addressing a number of the INSPIRE reference information priority components, such as roads, elevation, hydrography, direct referencing systems. However geographical names have not been given attentive focus yet.

Soon after the establishment of the United Nations Conferences on the Standardization of Geographical Names in 1967 guidelines were adopted for the creation of national names authorities and for the production of national gazetteers with a

² Association of the European National Mapping and Cadastral Agencies (NMCA).

view to the harmonisation and publication of standardized official geographical names. But, the technical prerequisites for rendering diacritics of different languages in (digital) files and databases satisfactorily only exist of late. Therefore, the Eighth United Nations Conference on the Standardization of Geographical Names (Berlin, 2002) recommends that standardized geographical names data shall be considered in the establishment of national and regional spatial data infrastructures and included in their design, development and implementation [3].

As a recently accepted project of EuroGeographics, EGN will also be linked with the EuroGeographics products EuroGlobalMap (EGM), EuroRegionalMap (ERM) and Seamless Administrative Boundaries of Europe (SABE). These European reference data products currently have only a selection of place names from all European countries, portrayed for every country in each official national (native) version and with an English-like transcription, but without rendering correctly some language-typic specifics, like diacritical marks and special characters.

Therefore, the Dutch- and German-speaking Division of the United Nations Group of Experts on Geographical Names (UNGEGN) and the German Federal Agency for Cartography and Geodesy (BKG) initiated the project EuroGeoNames (EGN), the vision of integrated geographical names data within a European SDI [4]. EGN shall ultimately be an Internet service linking geographical names official sources across Europe.

2.2 Why additional toponymic attributes?

Very often emotional discussions take place because of the array of characteristics and associations linked to any geographical name. The most obvious characteristic of geographical names is the identification of geographical entities. Furthermore, toponyms contain multiple attributes such as political, linguistic, and cultural aspects amongst others.

Especially the usage of exonyms causes controversial reactions. In the past different forms of use and misuse of exonyms did occur, e.g. due to the influence of political changes in the creation of exonyms. But on the other hand, exonyms can be considered as proof for cultural diversity and openness [5]. Noting that although the general goal of the United Nations is to limit the use of exonyms, in several countries there has been a tendency to increase the number of exonyms. In the European context exonyms become attractive while helping to promote the cultural diversity and the multilingualism.

Wrong spelling of geographical names e.g. in editorial contributions may divert parts of the potential of attention from the respective proper text contents and wrong spelling of geographical names e.g. in advertising texts and commercial contributions may entail claims for damages by the advertising customers on the broadcasting corporations. In the case of traffic announcements geographical names should be pronounced such that unambiguous localization is possible, thus providing correct information on the relevant traffic situation to the road users. Especially in the print media various examples can be found daily, for which the unclear use of geographical names caused e.g. a misled by a navigation system.

2.3 Current scenarios for geographical names data sources

Figure 2 shows three possible patterns of storing geographical names data. Geographical names data can be stored in a stand-alone database containing all/some feature categories (administrative units, populated places, hydrography, landscapes, etc.) and therefore including all relevant geographical names data (e.g. a Geographical Names Database – GNDB, case 1). Or they can be stored in distributed databases kept and maintained separately, e.g., some containing administrative units, others hydrographic features or other geographic features (case 2). These data sources are very often still incompatible because an encoding system (e.g. by a unified object identifier) as well as common standards and protocols for data exchange are missing. Or thirdly, the geographical names data may already be integrated in geographic information systems, sometimes featuring different scales, and can be connected with other geographic information (e.g. other topographic data or statistical data) at best within a regional or national spatial data infrastructure (SDI) (case 3).

Geographical names related data sources – current scenarios (selected)

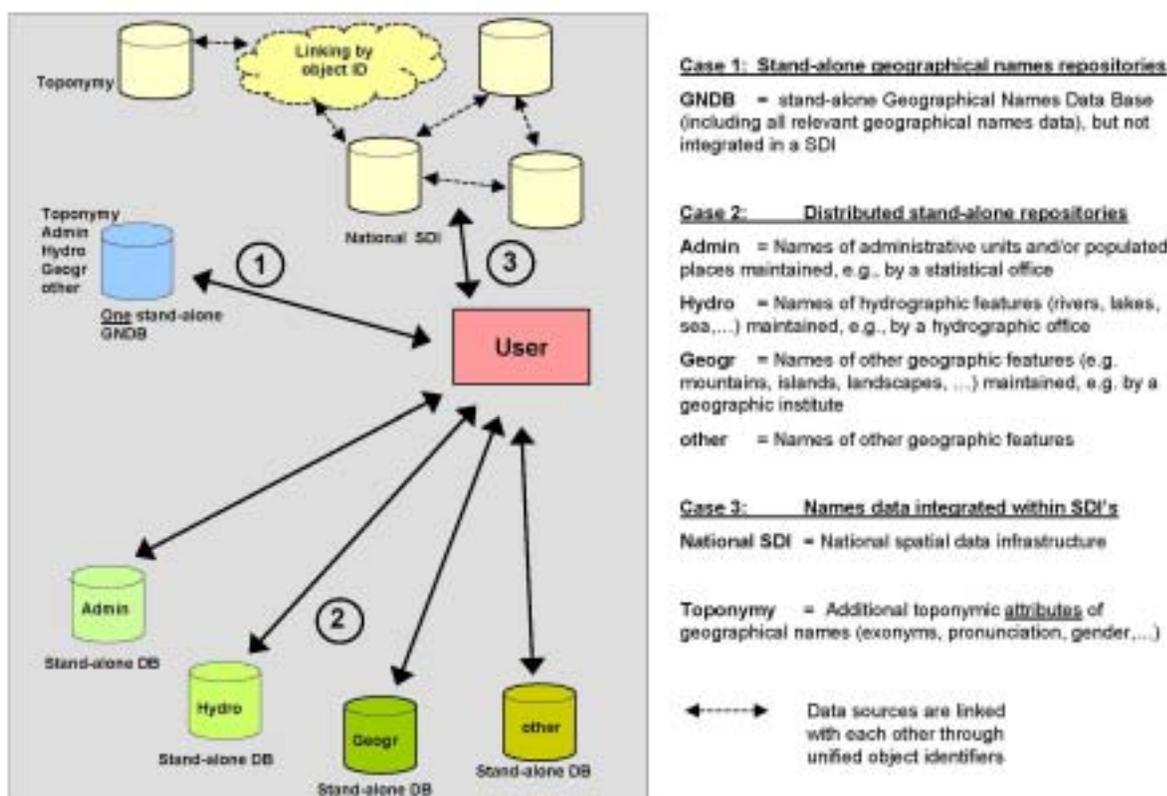


Figure 2 – Three typical examples of storing geographical names data, showing the distributed access to databases

2.4 The vision of EuroGeoNames (EGN)

The project EuroGeoNames (EGN) was initiated to integrate geographical names data better in a European SDI. EGN shall ultimately be an Internet service linking geographical names official sources across Europe. Generally, names searches shall be possible for all official European languages including the officially recognized minority languages. The EGN service shall increase the availability and the usability of (standardized) official national geographical names as well as their value-adding and profit-making by commercial enterprises. The data will be integrated as part of spatial data infrastructures. In European countries, where official sources are not yet available other data repositories (e.g. the US GEOnet Names Server (GNS), <http://earth-info.nima.mil/gns/html/>, etc.) could be considered to be taken into account [6],[7].

EGN would neither affect the independence or the organization mode nor the responsibility of the national agencies that produce and maintain geographical names repositories. It would, however, extend the applications of these national, regional or thematic databases by bringing them together in a continental context and allowing transnational applications independent of the many borders that cross Europe.

Generally, the set-up of EGN shall follow a step-wise approach. In a first stage, the density of acquisition for the geographical names concerned shall be focussed on the representation of features contained in the European data currently available at scale 1:250,000. If applicable, also data available in the countries at scale 1:50,000 or larger shall be considered. This target is in line with the INSPIRE principles.

The development of the EGN service will be a core task of the EGN project. The deployment of the ISO/OGC-based service will help to publish, find and deliver, use and study geographical names data through the Internet across Europe.

The future technological architecture for the network – shown in Figure 3 – will comprise the standards, technologies, specifications and procedures to represent, transform and generally accommodate the integration, maintenance and use of geographical names information in digital format [8],[9]. EuroGeoNames specifications, architecture and implementation will follow the guidelines for data harmonisation and specifications developed by INSPIRE and the EuroSpec initiative of EuroGeographics.

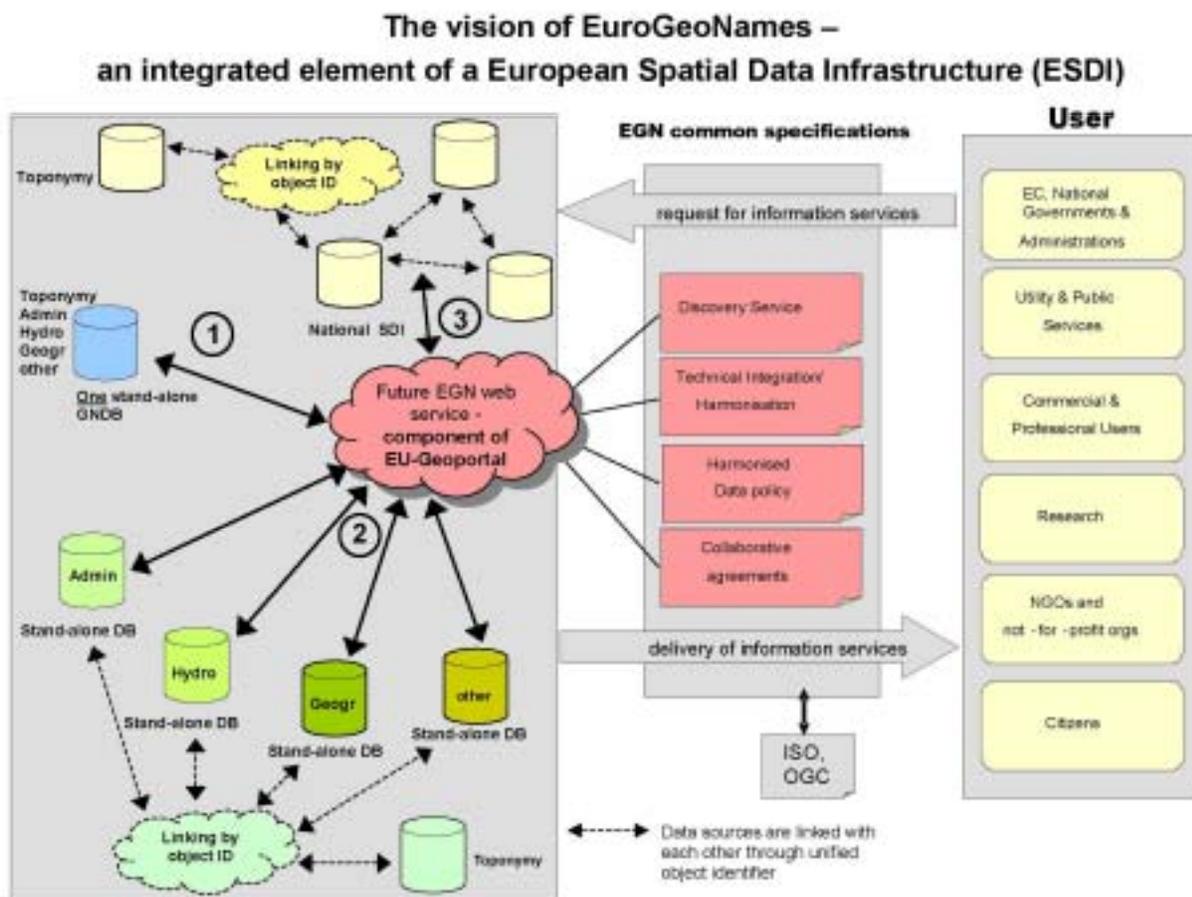


Figure 3 – Connection of the data sources (shown in figure 2) with each other by the future EGN network/service. This service will also provide links to other services.

A proposal for geographical names data which should be provided mandatory by the countries as well as for toponymic attributes which are additionally attractive for the EGN service is given in the following table:

Mandatory geographical names data	Additional toponymic attributes
Official geographical name	Exonym
Coordinates of the object	Pronunciation
Unique object ID	Gender
Object class	Language/script
Resolution, scale	Synonyms and/or non-official names forms
	Historical names

Table - Mandatory geographical names data and additional toponymic attributes

Due to the fact that currently satisfying measures do not exist for the categorization of exonym use, the publication of pronunciation guides for endonyms, and the formulation of guidelines ensuring a politically sensitive use of exonyms (aiming at a reduction of exonyms), the EuroGeoNames project could close these gaps and help to preserve the cultural heritage of countries and regions.

Generally, the advantages of an EGN service supporting all European languages are obvious when thinking e.g. of Internet searches for geographic information related to border-crossing features like the river “Danube” in Europe. Germans and Austrians may like to start their internet search with the geographical name in German „Donau“ whereas a Hungarian may like to start with „Duna“, a Serbian or Bulgarian with „Dunav“, a Romanian with „Dunărea“ and a Slovakian or Ukrainian with „Dunaj“ (see Figure 4).



Figure 4 – Some official local names forms (endonym) of the Danube river

The main bottleneck which has to be solved by commercial companies within the EGN project is the spelling of geographical names and their display through the Internet, no matter what the platform, no matter what the program, no matter what the language.

Hence, the EuroGeoNames project aims at stimulating commercial companies to develop and deploy an adaptable, interoperable Internet service incorporating the Unicode standard³ [10]. This is, amongst others, a new and still not fully exploited field and opportunity for commercial companies occupied with GIS, with the development of geoportals or with publication of geographical names related products.

Furthermore, incorporating Unicode into client-server or multi-tiered applications and websites offers significant cost savings over the use of legacy character sets. Unicode enables a single software product or a single website to be targeted across multiple platforms, languages and countries without re-engineering. It allows data to be transported through many different systems without corruption.

3 CURRENT STATUS OF THE EUROGEONAMES PROJECT

3.1 The Survey/Inventory on European Geographical Names data (SI-EGN)

The detailed European survey/inventory on geographical names data (SI-EGN) is a prerequisite to set up EuroGeoNames (EGN), investigating the availability, quality, accessibility and responsibility for national official geographical names data.

The SI-EGN Consortium, consisting of the BKG, as project coordinator, together with the National Mapping and Cadastral Agencies (NMCAs) from Austria and Slovenia as well as with the Eurogeographics Head Office (EGHO) and ESRI Germany, has started actions for SI-EGN in January 2005. A detailed questionnaire has been sent out by EGHO, both to all NMCAs being members of EuroGeographics as well as directly to experts nominated by the NMCAs as a national expert to the EuroGeoNames project. Additional interviews with the EuroGeographics VARs, distributors and main customers as well as with other GI stakeholders have also been prepared in March 2005 for an initial assessment of the market potential of the future EGN service.

First results of the evaluation and assessment of SI-EGN are expected in Mid-2005. As its main outcome it shall provide the draft content for the future EGN service. This includes the decision on mandatory geographical names attributes for EGN as well as a selection of additional toponymic attributes required for EGN.

3.2 Next steps

The EuroGeoNames project aims at setting up a customisable and interoperable Internet service that will enlarge the availability and accessibility of these geographical names. Together with commercial partners and based on the interviews mentioned in 3.1, EGN needs to define a business case to develop and exploit this service. To be successful in the eContent plus funding programme of the European Commission this will be the main task for the near future.

REFERENCES

- [1] Proposal for a Directive of the European Parliament and of the Council establishing an infrastructure for spatial information in the Community (INSPIRE), Brussels, 23.7.2004, COM(2004) 516 final, <http://inspire.jrc.it/home.html>
- [2] <http://www.eurogeographics.org>
- [3] Res VIII/6 (2002) Integration of Geographical Names Data into National and Regional Spatial Data Infrastructures, Eighth United Nations Conference on the Standardization of Geographical Names
- [4] Working Paper No. 40 (2004), EuroGeoNames (EGN) - A Project on a European Geographical Names GIS Network, Twenty-second Session of the United Nations Group of Experts on Geographical Names (UNGEGN)
- [5] Res II/28 (1972), List of exonyms (conventional names, traditional names), Second United Nations Conference on the Standardization of Geographical Names
- [6] Reports "Arbeitsberichte" on the Database of Geographical Names of Europe (1998 – 2002), Geographic Institute of Humboldt-University Berlin, Germany

³ Unicode is a new character coding system designed to support the worldwide interchange, processing, and display of the written texts of the diverse languages and technical disciplines of the modern world.

- [7] GEOnet Names Server (GNS), <http://earth-info.nima.mil/gns/html/>
 - [8] “Geodateninfrastruktur” (2005), Bernhard/Fitzke/Wagner, ISBN 3-87907-395-3
 - [9] ISO/DIS 19112, Geographic information – Spatial referencing by geographic identifiers (2001)
 - [10] <http://www.unicode.org/standard/WhatIsUnicode.html>
-