Section 7 Technical issues: internet, web services infrastructure and applications

Chapter 16 **Web services and applications - open source options vs commercial options: criteria for selection**

Pier-Giorgio Zaccheddu

16.1 Introduction

Access to consistent and reliable multilingual geographical names is essential for a number of uses including postal services, emergency services, navigation, tourism, property purchases, the mass media and applications such as Google Earth/Maps. In all of these areas, geographical names provide one of the most important keys for referencing and accessing a variety of related information. However, today, a patchwork of heterogeneous international, national and regional geographical names related web services and applications exist. Not all of them are compatible with the requirements and standards for integration in international, national and regional spatial data infrastructures.

From a spatial data infrastructure or database point of view, the entity 'name' might be one of many attributes related to the geographical object/feature. UNGEGN has continuously adapted its geographical names standardization programme as an essential part of spatial data infrastructures. It leads to a significant improvement in this portion of the geospatial data management framework promoted by the United Nations Global Geospatial Information Management (UN-GGIM).

This section deals with the display and publication of the content of geographical names databases in different output options, through web services or integrated in web applications.

The following sections will provide information about the prerequisites and tools for creating web services and applications for the publication of geographical names databases within regional, national and international spatial data infrastructures. The information is far from being exhaustive, but it might provide an overview of issues and considerations for geographical names experts.

16.2 The design of web services and applications

First of all, a web service and a web application need a database to be based on. The database needs a certain structure in order to be understood by the web service or application. How a geographical names database should be designed is explained in chapter ‘Technical issues: database management’.

Web services are the vehicle for accessing geographical names database content through an application, integrated in a Geographical Information System (GIS) or as an essential part of a spatial data infrastructure.

The German geoportal (Geoportal.de) reveals the contents of the national spatial data infrastructure (GDI-DE). It is the central point of access to the data and services of the GDI-DE. Users can search within the central search engine, which contains around 140,000 decentrally maintained sets of metadata on geospatial data sets and services from across all levels of public administration in Germany.

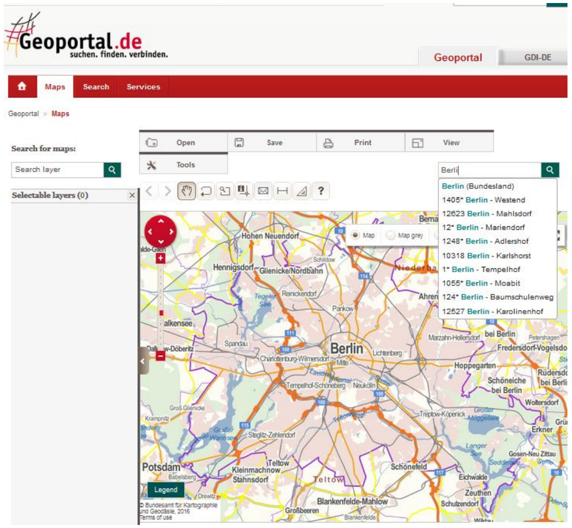

Figure 16-1 The German national geographical names database (GN-DE) published as a web service as part of the national spatial data infrastrucuture (GDI-DE) and visualized through the Geoportal application (Geoportal.de)

Different books and different organizations give different definitions of Web Services. Some of them are listed here [1]:

“A web service is any piece of software that makes itself available over the internet and uses a standardized Extensible Markup Language (XML) messaging system. XML is used to encode all communications to a web service. For example, a client invokes a web service by sending an XML message, then waits for a corresponding

XML response. As all communication is in XML, web services are not tied to any one operating system or programming language."

"Web services are self-contained, modular, distributed, dynamic applications that can be described, published, located, or invoked over the network to create products, processes, and supply chains. These applications can be local, distributed, or web-based. Web services are built on top of open standards such as TCP/IP, HTTP, Java, HTML, and XML."

"A web service is a collection of open protocols and standards used for exchanging data between applications or systems. Software applications written in various programming languages and running on various platforms can use web services to exchange data over computer networks like the Internet in a manner similar to inter-process communication on a single computer. This interoperability (e.g., between Java and Python, or Windows and Linux applications) is due to the use of open standards."

To summarize, a complete web service is, therefore, any service that:

- Is available over the Internet or private (intranet) networks
- Uses a standardized XML messaging system
- Is not tied to any one operating system or programming language
- Is self-describing via a common XML grammar/syntax
- Is discoverable via a simple find mechanism

In a nutshell, web services allow various applications to talk to each other and share data and services among themselves. Thus, different applications can use the same web service if the required standards and protocols are used.

In computing, a web application or "web app" is a client-server software application in which the client (or user interface) runs in a web browser (e.g. Microsoft Internet Explorer or Mozilla Firefox) [2]. Unlike traditional desktop applications, which are launched by the operating system, web apps can only be accessed through a web browser.

The difference between a website and a web application is simple. A website is defined by its (static) content while a web application is defined by its interaction with the user. E.g. a website can plausibly consist of a static content repository that's accessible for all visitors, while a web application depends on interaction and requires user input and data processing. For example, a daily news site would be a website, but a spreadsheet or a collaborative calendar would be a web application.

Web browsers support many programming languages. Web apps have to be programmed in a language the browser can understand. There is no "best language for developing web applications". The following is a list (not complete) of languages that are commonly used to programme web applications [3]:

- Python: high-level programming language focusing code readability.
- Java: the server-side language for large-scale websites with a high volume of traffic.
- JavaScript: used to express animation, interactivity and other dynamic effects.
- Ruby: similar to Python, emphasizing short and simple code.
- PHP: a general-purpose server-side scripting language.
- CSS: allows a programmer to define the layout of multiple webpages at once.
- HTML: the standardized markup language that structures and formats content on the web.

HTML and CSS are not really programming languages. HTML is simply the format consumed by web browsers to parse and render webpages. CSS is used to write rules that specify the appearance of the page (fonts, colours, background, etc.).

A special web service, addressing a specific profile or use case for publishing geographical names data, is the so-called "gazetteer service". In this context, geographical names databases and gazetteers should be distinguished from each other, as gazetteers are only one of many possible outcomes derived/produced from a names database. A gazetteer can be defined as a list, a report or a repository of location information that is used to search for specific locations.

The term 'gazetteer' in a spatial data infrastructure (SDI) context is considered as *"any geospatial dataset which contains 'spatial identifiers'"*. These can be geographical names, postal codes or other indexes for indirect spatial referencing. The intended use of 'gazetteers' in the European INSPIRE initiative (using 'geographic identifiers') followed the International Organization for Standardization (ISO) 19112 standard. The schema from ISO 19112 was slightly amended to correct errors in that schema and to allow for a better integration in INSPIRE as an SDI. 'Gazetteers' here were simply intended as a channel to publish spatial data from the INSPIRE themes that allows others to use them in indirect spatial referencing. It is obvious, that this technical SDI view on 'gazetteers' is different from the UNGEGN view on 'gazetteers': *"List of toponyms arranged in alphabetic or sequential order, with an indication of their location and preferably including variant names, type of (topographic) feature and other defining or descriptive information."* [4]

Within UNGEGN recommendations and policies, and indeed within the wider research literature, commonly accepted definitions for the terms 'official' and 'unofficial' do not seem to be available as they relate to gazetteer data. Rather, there seems to be a proliferation of terminology used to define both types of data which are incorporated into gazetteers, and the gazetteers themselves – ranging from official and authorised to unofficial and informal. The need for the officially sanctioned gazetteers to be of a high quality in terms of accuracy and completeness of available data is increasing rapidly [5]. In a nutshell, a gazetteer enables the user to search and find a location of interest within an application. A gazetteer service uses location data (usually in object/feature based format) with information related to location like a location name, coordinates and/or postcodes [6].

There is growing interest in the development of a common object/feature-based model for access to named objects/features, often referred to as a 'gazetteer' as well. Two major activities form the basis of this standard, an Open Geospatial Consortium (OGC) Best Practice for Gazetteer Services – Application Profile of the Web Feature Service Gazetteer (WFS-G) Implementation Standard, which is an OGC Discussion paper on gazetteers, and an ISO draft standard for geographic identifiers [7]. Since this resource is so important, there is growing global interest in sharing and updating geographical names across a standard web-based service interface that is not controlled by any one organization or group. OGC has met this need for open accessibility to geographical feature data via its creation of the Web Feature Service (WFS) standard. The WFS standard defines an interface for specifying requests for retrieving and updating geographical features across the Web using platform-independent calls. And of interest to the geographical names community, members of the OGC have been actively developing a special profile of the WFS which is being designed specifically to support geographical names data through a WFS-G profile. However, the WFS-G profile is still a Best Practice document and has not been adopted as an OGC standard for different reasons, amongst others, that schema elements, attributes, etc. should be defined and described in more detail. Furthermore, the suggested closed code list for type/status of name ("official", "variant") needs to be reconsidered in order to make the schema applicable beyond the USA and Canada. A compromise (INSPIRE) for European purposes was "official", "standardized" and "other" for current names, and "historical" for names not in use anymore. Only a few countries in Europe provide their geographical names with status 'official' while the names are 'standardized'. Also, the value "variant" is out of the scope of name's status (e.g. "official").

16.3 Open source options vs commercial options: criteria for selection

Generally, one can find commercially oriented and open source oriented web feature (gazetteer) services software and implementations. It is deemed impossible to define selection criteria for any decision-making regarding their selection. From the technical point of view there might be advantages and disadvantages with every software programme or tools considered for the services architecture, but these are very specific and depend on the envisaged implementation. Very often the existing IT infrastructure and services architecture in an organization determine the usage of web services software/tools. For example, if an organization uses ESRI Arc GIS for the maintenance of the spatial data – including geographical names – the publication of web services through the tool ESRI Arc GIS Server is obvious. Nevertheless, technical requirements can make the usage of open source software necessary as well.

Organizations providing a commercial web services' offering and charging for its usage, most probably do offer a free-of-charge (light) version of the services or applications as well. This is general practice. One example for an open gazetteer service is provided via Ordnance Survey (OS). OS OpenData provides access to a 1:50,000 scale gazetteer product that can be used to generate gazetteer services [8]. Another example is GEOnet Names Server (GNS), provided by the United States National Geospatial-Intelligence Agency (NGA). It is an official repository of standard spellings of global geographical names. All geographical features within the database contain information about location, administrative division and quality and are ISO compliant. The database can be used for a variety of purposes, including establishing official spellings of foreign place names, cartography, GIS and finding places [9].

Figure 16-2 Search for geographical names through the GEOnet Names Server (GNS)

GNS provides a number of OGC services including a WFS-G. This can be implemented into any WFS enabled client and is therefore extremely useful for geoportals looking to provide gazetteer services and search tools. Furthermore, the GNS repository of global geographical names, coordinates, and extents can be used to create gazetteer services. All data and services are open and available to download free of charge from the website.

In summary, geographical names are used within different web services and applications:

- As search criteria (location), e.g. in a geoportal, for rescue services, geocoding [10], geoparsing [11] and navigation.
- As geographical identifiers, e.g. in gazetteer services.
- For visualization, e.g. as an information layer in viewing services.
- In standardization, translation, and compilation of maps, reports, documents and articles. For instance, reliable information on the correct spelling and the status of names is required by press agencies and map producers.
- For the processing of spatial data sets, e.g. for integration of historical data.
- In human and social sciences, e.g. in linguistic research, onomastic science, archaeology and etymology.

Figure 16-3 News portal of the European Commission using geographical names in different languages to search for information

For all these reasons and in order to allow for these various use cases the publication of geographical names databases as web services or web applications is broached.

From a spatial data infrastructure or database point of view, the entity 'name' is only one of many attributes related to the geographical object/feature. One and the same geographical object/feature may be described by many different names, each one of which may again be pronounced, transcribed, transliterated or otherwise rendered graphically in different ways, which may be considered 'official' at different levels at different times. UNGEGN has continuously adapted its geographical names standardization programme as an essential part of spatial data infrastructures, leading to a significant improvement in this portion of the geospatial data management framework promoted by the United Nations Global Geospatial Information Management (UN-GGIM) as geographical names are definitely very pertinent to it.

Within UNGEGN a trend of activities in the countries/divisions focused on the establishment of multi-functional or multi-usable geographical names databases, services and applications to provide GN data for different purposes has been identified – e.g. providing geographical names as an essential dataset to the national or regional SDI or for the support of specific services and applications. Web services technologies for the geographical names database provision, visualization and dissemination are increasingly used and thus support the vision of a (national or regional) Spatial Data Infrastructure (SDI). This very positive trend of supporting multi-purposes of geographical names use is recognized by UNGEGN Resolution VIII/6 'Integration of Geographical Names data into National and Regional Spatial Data Infrastructures' of the Eighth United Nations Conference on the Standardization of Geographical Names (Berlin, 2002). Res VIII/6 […] *"recommends that standardized geographical names data should be better considered in the establishment of national and regional spatial data infrastructures (SDIs) and included in their design, development and implementation."* [12]

16.4 References

[1] tutorialspoint, http://www.tutorialspoint.com/webservices/what_are_web_services.htm, last accessed 09/2016

[2] Wikipedia, https://en.wikipedia.org/wiki/Web_application, last accessed 09/2016

[3] 10 Best Programming Languages of 2015, Devsaran, https://www.devsaran.com/blog/10-best-programming-languages-2015-you-should-know, last accessed 09/2016

[4] Glossary of Terms for the Standardization of Geographical Names, UNGEGN, 2002, http://unstats.un.org/unsd/geoinfo/UNGEGN/docs/pdf/Glossary_of_terms_revised.pdf

[5] "The Four Faces of Toponymic Gazetteers", Submitted by Australia, Tenth United Nations Conference on the Standardization of Geographical Names, 2012, E_CONF.101_57_The Four Faces of Toponymic Gazetteers

[6] Explanations to gazetteer services, http://learningzone.rspsoc.org.uk/index.php/Learning-Materials/Introduction-to-OGC-Standards/5.3-Example-of-gazetteer-services, last accessed 09/2016

[7] Best Practices Document "Gazetteer Service - Application Profile of the Web Feature Service Best Practice", Copyright © 2012 Open Geospatial Consortium; To obtain additional rights of use, visit http://www.opengeospatial.org/legal

[8] OS OpenData, http://www.ordnancesurvey.co.uk/business-and-government/products/opendata-products.html, last accessed 09/2016

[9] GEOnet Names Server (GNS), http://geonames.nga.mil/gns/html/, last accessed 09/2016

[10] Wikipedia, "Geocoding is the computational process of transforming a (postal) address description to a location on the Earth's surface", https://en.wikipedia.org/wiki/Geocoding, last accessed 09/2016

[11] Wikipedia, "Geoparsing is the process of converting free-text descriptions of places (such as "twenty miles northeast of Jalalabad") into unambiguous geographic identifiers, such as geographic coordinates expressed as latitude-longitude.", https://en.wikipedia.org/wiki/Geoparsing, last accessed 09/2016

[12] UNCSGN Resolutions 1967 – 2012, http://unstats.un.org/unsd/geoinfo/UNGEGN/docs/RES_UN_E%20updated_1-10%20CONF.pdf, last accessed 09/2016