Database management

Module 6: Data Processing and Management
International Training on Toponymy
Bali – Indonesia

Jasper Hogerwerf
Content

Preparing the names database – General issues

▪ What are the purposes of a database?
▪ Names database and Geographic Information System (GIS)
▪ Names databases and the Web
▪ Use cases for Web applications

Description of database characteristics – Database management

▪ Characteristics of databases
▪ Realization of databases: open source vs. proprietary
▪ Data model and database modelling
What are the purposes of a database?

- Gazetteer
- GIS
- Map
- Names DATABASE (national)
- Names DATABASE (international)
- Linked data

**Names DATABASE**

**GIS**

**Map**
What are the purposes of a database?

- Gazetteer
- Maps
- GIS
- Names DATABASE (national)
- Names DATABASE (international)
- Linked data
What are the purposes of a database?

Gazetteers in the Netherlands

20 years ago: Digital military gazetteer
Topografische Dienst

Current: TOPnamen annotation gazetteer
Kadaster
What are the purposes of a database?

Gazetteer in New Zealand

Gazetteer web map application

Toitū Te Whenua
Land Information New Zealand

https://gazetteer.linz.govt.nz/
What are the purposes of a database?
Names database and Geographic Information System (GIS)

- Where am I?
- Where do I find…?
- Where is the next…?
- How do I get to…?
- How far is it to…?
- Where does this way lead to?

Data models consist of feature definitions + relations including spatial reference to points, locations, areas or regions as specific feature

→ necessary information is called geo(graphic) information
→ 80% of all information is estimated to be spatially referenced
Geographic Information System (GIS)
a GIS is an organized collection of computer hardware, software, (geographic) data and personnel.

→ designed to capture, store, update, manipulate, analyse and display all forms of geo referenced information.

(see www.GIS.com)
Geo(graphic) information:
Information that is referenced to the earth's surface, whether by coordinates (direct referencing) or by identifiers such as addresses or postal codes or geographical names (indirect referencing).

Geo(graphic) data / spatial data:
Computer-readable geo information

<table>
<thead>
<tr>
<th>Vector data model (feature data)</th>
<th>Raster data model (coverage data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>points, lines and polygons (areas)</td>
<td>gridded data (scanned maps, satellite images, orthophotos)</td>
</tr>
</tbody>
</table>
Names database and Geographic Information System (GIS)
Names database and Geographic Information System (GIS)
Names database and Geographic Information System (GIS)

BRT Achtergrondkaart (BRT Background Map) as ‘NL Maps’ service integrated with:
- Google Maps
- Leaflet
- Mapbox
- OpenLayers

https://app.pdok.nl/viewer/ (in Dutch only)

INSPIRE Geographical Names layer with all names from BRT - TOP10NL
Names database and Geographic Information System (GIS)

Example of a separate names database / register:

Danske Stednavne

Styrelsen for Dataforsyning og Infrastruktur, Denmark

https://sdfi.dk/data-om-danmark/vores-data/danske-stednavne (in Danish)
What are the purposes of a database?

Eighth United Nations Conference on the Standardization of Geographical Names, 2002

Resolution VIII / 6

Integration of Geographical Names Data into National and Regional Spatial Data Infrastructures

The Conference,

....

Recommends,

that standardized geographical names data should be considered in the establishment of national and regional spatial data infrastructures and included in their constructions.
What are the purposes of a database?

A **spatial data infrastructure** (SDI) is a **data infrastructure** implementing a framework of **geospatial/geographic data**, **metadata**, users and tools that are interactively connected in order to use spatial data in an efficient and flexible way.

Another definition is:

[...] the technology, policies, standards, human resources, and related activities necessary to acquire, process, distribute, use, maintain, and preserve spatial data.

What are the purposes of a database?

https://www.pdok.nl/ (in Dutch)

https://www.nationaalgeoregister.nl/ (in Dutch)
What are the purposes of a database?
What are the purposes of a database?
Names databases and the Web

UNEGGN World Geographical Names database search for geographical names of countries and capitals

https://uneggn-dashboard-hja6b0dafqh0c6dd.z01.azurefd.net/

http://unstats.un.org/unsd/geo info/geonames/
Names databases and the Web

<table>
<thead>
<tr>
<th>Language</th>
<th>Short Name</th>
<th>Formal Name</th>
<th>Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indonesian</td>
<td>Jakarta</td>
<td>Jakarta</td>
<td></td>
</tr>
<tr>
<td>Arabic</td>
<td>دكايرا</td>
<td>Jakarta</td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>忍京</td>
<td>Jakarta</td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>Jakarta</td>
<td>Jakarta</td>
<td></td>
</tr>
<tr>
<td>French</td>
<td>Jakarta</td>
<td>Jakarta</td>
<td></td>
</tr>
<tr>
<td>Russian</td>
<td>Джакарта</td>
<td>Jakarta</td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>Yakarta</td>
<td>Jakarta</td>
<td></td>
</tr>
</tbody>
</table>
Names databases and the Web
Names databases and the Web

https://www.geonames.org/
Names databases and the Web

https://geonames.nga.mil/geonames/GeographicNamesSearch/
Names databases and the Web

Toponamenzoeker (Topo Names Finder)

Web application to search for geographical names in the Key Register Topography (BRT) of the Netherlands

https://www.toponamenzoeker.nl
Names databases and the Web
Names databases and the Web

Developing towards linked open data:

1 star: Map with names online

2 stars: Names database available online in Esri format (e.g. File Geodatabase)

3 stars: Names database available online in open source format (e.g. Geopackage)

4 stars: Names, attributes and values available as URI (linkable web data)

5 stars: URIs of names database linked to other linked open data
Content

Preparing the names database – General issues
- What are the purposes of a database?
- Names database and Geographic Information System (GIS)
- Names databases and the Web
- Use cases for Web applications

Description of database characteristics – Database management
- Characteristics of databases
- Realization of databases: open source vs. proprietary
- Data model and database modelling
Use cases for Web applications using names data

International purchase of properties with Internet property services
http://www.viviun.com/

⇒ Geographical names databases or web services can provide additional location-based information

Location based information on **Pozzuolo, Toscana (IT)** possibly directly from the Italian data repository!
Use cases for Web applications using names data

Location apps and websites

Navigation software

Web shops using address data
Content

Preparing the names database – General issues

▪ What are the purposes of a database?
▪ Names database and Geographic Information System (GIS)
▪ Names databases and the Web
▪ Use cases for Web applications

Description of database characteristics – Database management

▪ Characteristics of databases
▪ Realization of databases: open source vs. proprietary
▪ Data model and database modelling
Characteristics of databases

What do you have to consider (or to know) before you start creating a database?
Characteristics of databases

Data Capture

Data Storage

Data Processing

Data Output

from cards, lists, books, fieldwork, maps, other databases, ....
Characteristics of databases

Main software requirements:

1. Operating system

   software program that manages the hardware and software resources of a computer.

   **Examples:** Microsoft Windows, Apple iOS (proprietary), Linux (open source = free of charge and usage)

2. Database

   collection of records stored in a computer in a systematic way, so that it can be consulted to answer questions.

   **Examples:** Microsoft Access, SQL Server, Oracle (proprietary), MySQL, PostGreSQL / PostGIS (open source)

3. Geographic Information System (GIS) software

   tool that allows users to select, analyze, edit, manipulate and visualize data from a database.

   **Examples:** ESRI ArcGIS, Hexagon Geomedia, MapInfo (proprietary), QGIS, GRASS GIS (open source)
Characteristics of databases

Important: Code files!

**ISO8859 character set** or **Unicode**?
in (Microsoft Windows) operating system

Different interpretation of characters possible in

- Database products: MS Access or similar
- GIS software: ESRI ArcGIS or similar
Content

Preparing the names database – General issues

- What are the purposes of a database?
- Names database and Geographic Information System (GIS)
- Names database and the Web
- Use cases for Web applications

Description of database characteristics – Database management

- Characteristics of databases
- Realization of databases: open source vs. proprietary
- Data model and database modelling
Realization of databases

Operating System:
Microsoft Windows?  
Linux?  (open source)

Database:
Microsoft Access?  
MySQL?  (open source)  
PostgreSQL/PostGIS?  (open source)

Geographical Information System
ESRI ArcGIS?  
Quantum GIS?  (open source)
Realization of databases

Database: Microsoft Access? (proprietary)

• Microsoft Access is a **pseudo-relational database management** system from Microsoft that combines the relational Microsoft Jet Database Engine with a graphical user interface and software-development tools.
• Access stores data in its **own format** based on the Access Jet Database Engine. It can also import or link directly to data stored in other applications and databases.
• Website: [https://www.microsoft.com/en-us/microsoft-365/access](https://www.microsoft.com/en-us/microsoft-365/access)
Realization of databases

Database:
MySQL? (open source)

“The world's most popular open-source database”
• MySQL is a relational database management system (RDBMS) that runs as a server providing multi-user access to a number of databases.
• The MySQL development project has made its source code available under the terms of the GNU General Public License, as well as under a variety of proprietary agreements.
• MySQL is a popular choice of database for use in web applications.
• Website: http://www.mysql.com/
Database:
PostgreSQL/PostGIS? (open source)

- PostgreSQL, often simply Postgres, is an object-relational database management system (ORDBMS).
- PostgreSQL is not controlled by any single company — a global community of developers and companies develops the system.
- PostGIS adds support for geographic objects to the PostgreSQL object-relational database. In effect, PostGIS "spatially enables" the PostgreSQL server, allowing it to be used as a backend spatial database for geographic information systems (GIS), much like ESRI's SDE or Oracle's Spatial extension.
- Website: http://www.postgresql.org/
  http://www.postgis.net/  - http://postgis.refractions.net/
Realization of databases

Geographical Information System
ESRI ArcGIS? (proprietary)

• Esri® ArcGIS® facilitates collaboration and lets you author data, maps, globes, and models on the desktop and serve them for use on a desktop, in a browser, or in the field, depending on the needs of your organization.
• ArcGIS support and educational services consist of technical maintenance programs, software releases and updates, technical support, online support services, publications, training, and consulting services.
• Website: https://www.esri.com/en-us/arcgis/about-arcgis/overview
Realization of databases

Geographical Information System

Quantum GIS? (open source)

- QGIS provides data viewing, editing, and analysis capabilities.
- Quantum GIS (QGIS) is a user-friendly open-source Geographic Information System (GIS) licensed under the GNU General Public License.
- QGIS runs on Linux, Unix, Mac OSX, and Windows and supports numerous vector, raster, and database formats and functionalities.
- Website: http://www.qgis.org/en/site
Realization of databases

**Open Source Software Collection**

Open source software collection: OSGeo Live


OSGeo-Live is a self-contained bootable DVD, USB thumb drive or Virtual Machine, that allows you to try a wide variety of open-source geospatial software without installing anything. It is composed entirely of free software, allowing it to be freely distributed, duplicated and passed around.
Content

Preparing the names database – General issues

▪ What are the purposes of a database?
▪ Names database and Geographic Information System (GIS)
▪ Names database and the Web
▪ Use cases for Web applications

Description of database characteristics – Database management

▪ Characteristics of databases
▪ Realization of databases: open source vs. proprietary
▪ Data model and database modelling
Data model and database modelling

“UNEGGN schema” (proposal): elements

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Data Type</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>RecordID</td>
<td>Index</td>
<td>This number is assigned automatically by the system. Do not change it.</td>
<td></td>
</tr>
<tr>
<td>UID</td>
<td>Number</td>
<td>A unique identifier assigned to the name (or named feature). In our exercise, each group may start to count from 1 and add the group number multiplied by 100, e.g. 201 for the first name by group 2.</td>
<td>20001 might indicate e.g. 2000 for feature type ‘Populated places’</td>
</tr>
<tr>
<td>Name</td>
<td>Text</td>
<td>Geographical name in Romanian</td>
<td></td>
</tr>
<tr>
<td>Latitude</td>
<td>Number</td>
<td>Geographical Coordinates, in degrees decimal.</td>
<td>35,…</td>
</tr>
<tr>
<td>Longitude</td>
<td>Number</td>
<td></td>
<td>-0.5…</td>
</tr>
<tr>
<td>FeatureCode</td>
<td>Text</td>
<td>Feature codes, from an existing feature code table</td>
<td>DDS</td>
</tr>
<tr>
<td>AdminUnit</td>
<td>Text</td>
<td>Name of the administrative unit where the name is situated in Departement d’ Oran, Department of Oran</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>Text</td>
<td>Language of the name</td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td>Text</td>
<td>Field remarks, meaning of the name, language of the name, historical names if any</td>
<td>e.g. capital of a political entity</td>
</tr>
<tr>
<td>VariantName</td>
<td>Text</td>
<td>Enter variant names, if any</td>
<td>Wilaya d’Oran, Wilaya d’ Oran</td>
</tr>
<tr>
<td>MapSheet</td>
<td>Number</td>
<td>Reference to a map sheet in a topographic map series, e.g. 1:250.000</td>
<td>80</td>
</tr>
<tr>
<td>Source</td>
<td>Text</td>
<td>Source of the information on the name: - Informant - Interviewer</td>
<td>Mr. XY, old person at xyz, interview by group 1</td>
</tr>
<tr>
<td>Status</td>
<td>Text</td>
<td>The status of the name. In our case, the names are not yet approved by the Board.</td>
<td>not approved</td>
</tr>
<tr>
<td>Pronunciation</td>
<td>OLE-Object</td>
<td>Audio-files of the pronunciation of the geographical name</td>
<td>e.g. *.wav -file</td>
</tr>
<tr>
<td>Location information</td>
<td>OLE-Object</td>
<td>Digital pictures of the location</td>
<td>e.g. *.jpg – image file</td>
</tr>
</tbody>
</table>
Data model and database modelling

The European (INSPIRE) GN schema: elements

- name(s) (text, spelling)
- geometry
- feature type
- unique identifier

**mandatory**

- language {three letter codes from ISO 639-3 or -5}
- nameStatus {official, standardised, historical, other}
- link to relatedSpatialObject
- script {four letters codes defined in ISO 15924}
- nativeness {endonym, exonym}
- transliterationScheme
- grammatical gender {masc., fem., neuter, common}
- grammatical number {singular, plural, dual}
- pronunciation
- sourceOfName
- typeLocal
- lifecycleInfo (begin/end of the object in the source DB)
- …

**voidable**
A Named Place, representing a real world entity referred to by a Geographical Name
e.g. “the City of Athens” type = built-up area
geometry = \{X, Y\}

is associated with one or several Geographical Names,
i.e. proper noun applied to the feature

1) “Athína” language = Greek
   nativeValue = Endonym

2) “Athens” language = English
   nativeValue = Exonym

and may have one or several Spellings of Name,
i.e. proper way of writing the name

1.1) text = Aθína
    script = Greek

2.1) text = Athens
    script = Latin

1.2) text = Αθína
    script = Latin
Standardization in Europe

Not standardized and very different

Almost harmonized, but with national specialities

© Andreas Illert
Information on geographical names data

List of geographical names databases: https://unstats.un.org/unsd/ungegn/nna/geo-names/

Get involved in UNGEGN

Working Group on Geographical Names Data Management

https://unstats.un.org/unsd/ungegn/working_groups/wg2.cshtml
Thank you for your attention!