Swiss World Atlas: Development and automated setup of a geographical name index database for cross-media atlas applications

Thomas Koblet, Lorenz Hurni

ETH Zurich, Institute of Cartography and Geoinformation – Zürich, Switzerland

Abstract. The first edition of the Swiss World Atlas was released in 1910. Since then numerous enhancements und updates were realized. The first entirely digitally produced edition was published in 2002. During these years, the map production workflow has changed fundamentally. In contrast, the creation of the geographic name index is not entirely different from its establishment in earlier years. Still, for the 2002 edition the geographical name index was typed and amended manually with attributes like latitude, longitude, administrative area or type of geographical object.

For the next edition of the Swiss World Atlas the workflow to create the index will be automated and simplified using an innovative, self-developed ArcGIS tools (written in Python with the ArcPy package), which helps to build up and handle an index database. The functionality of these tools includes translation of geographical names (German, French, Italian, English), definition of the location (country and administrative subdivision), calculation of the coordinates, saving of the entities to the new index database and derivation of the printable version of the geographical name index.

The benefits of continuously building up database are manifold. Besides potential of saving time in final stage of the atlas production, a database allows the creation of a printed index based on various rules. Furthermore, an index database can be used as a search database when complementing a printed atlas with a digital atlas. By the example of the Swiss World Atlas interactive, it is shown in this paper that a search functionality requires the same database structure and attributes as the generation of an index for a printed atlas.

Keywords: Atlas, Atlas Index, Automation, Cross-Media Publishing, Database, Toponymy

1. Introduction

The Swiss World Atlas (SWA) is a widely used and well-established school atlas with a history of more than one hundred years. Considering Switzerland's linguistic diversity, the SWA exists in three languages (German, French and Italian). The first edition was released in 1910 and since then numerous enhancements and updates were published (Hurni 2004, Cron et al. 2011). During these years the map editing process has changed fundamentally from analogous to digital methods (Hurni & Wondrak 2011). The first entirely digitally created edition of the SWA was published in 2002 (Spiess) and with the current edition from 2010 (Hurni & Spiess 2010) more and more data is stored in spatial databases. In contrast to the rapid progress of map authoring techniques the compilation of a name index, which is crucial for the navigation in an atlas, is not completely different from its establishment in earlier years. Still, for the last editions of the SWA the geographical name index was typed and amended manually with attributes like latitude, longitude, country membership or geographical type.

In order to simplify and automate the process of creating a geographical name index for the next edition of the SWA, an innovative ArcGIS Toolbox, written in Python with the ArcPy package was developed, in order to build up and handle an index database containing a great amount of geographical entities linked to hundreds of maps in the printed atlas.

In case of cross-media publishing the availability of an up to date index database becomes more and more important. By the example of the Swiss World Atlas interactive, which was first released in 2010 (Cron et al. 2011, Haeberling et al. 2011), it is shown in this paper that an interactive search functionality requires the same database structure and attributes as for the generation of an index for a printed atlas.

Feldfunktion geändert
Feldfunktion geändert
Feldfunktion geändert
Feldfunktion geändert
Feldfunktion geändert
Feldfunktion geändert
Feldfunktion geändert

2. Index Database

The index database is built with the user-friendly SQLite database system (Owens 2007). SQLite has a python API (sqlite3), which makes it attractive for the use with ESRI's ArcPy package. In addition, SQLite has also a graphical user interface, hence it allows the user to edit the database tables manually.

2.1. Extracting the existing Data

Before building a new SQLite index database as much information as possible was extracted from the existing geographical name index, which was available as an Adobe InDesign file. This data includes coordinates of an

entity, geographical type, country, elevation (mountains and mountain passes only) and a page number, which indicates the page in the atlas. Figure 1 shows a typical extract of the German name index.

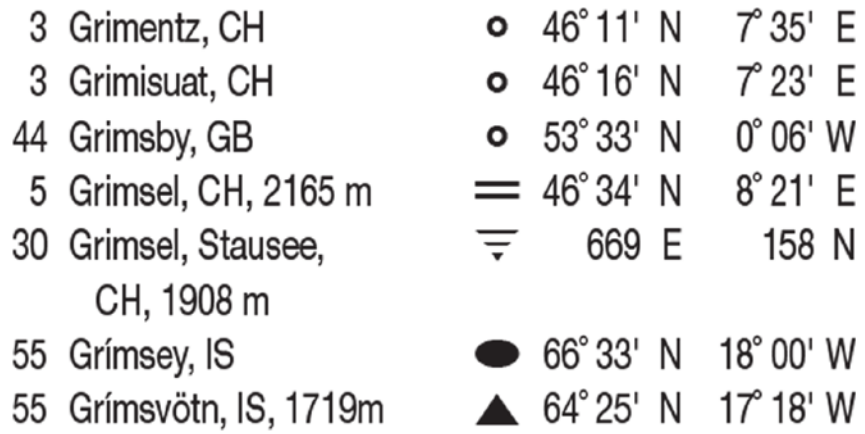

Figure 1: Extract from the German name index. The first number indicates the page, on which an entity is mentioned. It is followed by its exonym and its country membership. For mountains and mountain passes (e.g. Grimsel) the elevation is given. The symbol in the next column represents the geographical object (e.g. Grímsvötn is a mountain). At the end of each entry the coordinates are given.

The fact that the SWA is available in three languages (German, French, Italian) allowed to extract this data in three languages. Afterwards, the three tables were linked by their coordinate information. Although this step is not free of errors (the name indices of the different language editions were done by different editors) it was possible to establish a table with more than 17'000 locations in three languages (plus English if available[1]). This table with the information from the different existing name indices is called *index_old*. The table *index_old* serves for the various tools as input data.

[1] If an entity has the same exonym in German, French and Italian it is supposed to be named similarly in English. Furthermore, if the English exonym was given in the name index, it was as well written into the table.

2.2. Database Schema of the new Index Database

Figure 2 shows the database schema of the new SQLite index database. Table *index_new* will contain all the entities that appear in the next edition of the SWA. The database table *page_appearance* describes which entity is found on which pages in the atlas. One entity typically appears on various maps and is thus linked to the table *atlas_content*. This allows developers to easily define rules in order to determine which page number is finally printed in the name index.

The main table *index_new* has basically the same attributes as the table *index_old*. The attribute *old_geotype* is replaced by *new_geotype* because the geographical objects (also called *geotypes*) have a new naming in the *index_new* table (linked by the table *type*). The attribute *population* is added to *index_new* and contains the population of cities. The attribute *page* from the *index_new* table was not used because it is not clear, if this page number indicates the page where an entity appears first or if it indicates the map with the smallest scale. Finally, the page number in the printed geographical name index is generated based on the *page_appearance* table.

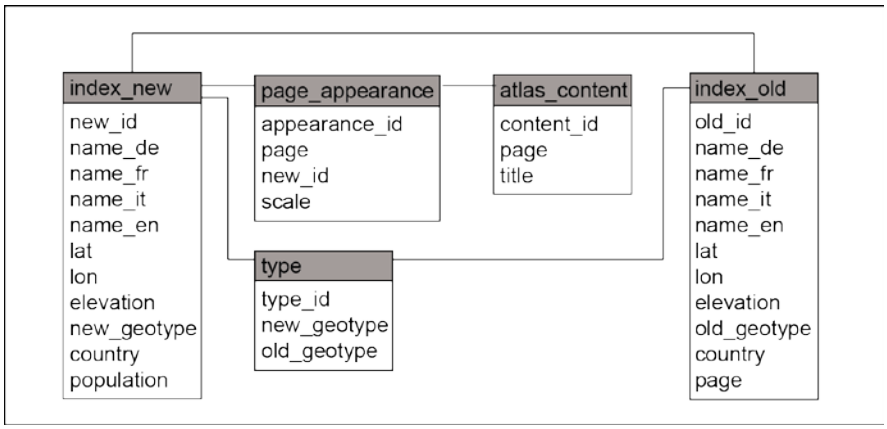

Figure 2: Database schema of the new index database. The geographical entities from table *index_new* are linked to the maps of the SWA (*atlas_content*) by table *page_appearance.*

3. Filling the Index with Data

The index tools described in this section are organized in an ArcGIS Toolbox. An ArcGIS Toolbox is a set of tools that can be distributed among other users, working with ESRI ArcGIS (ESRI 2012). The tools in this

Feldfunktion geändert

toolbox are written in Python with the ArcPy package, a Python library provided by ESRI to automate geoprocessing, and the sqlite3 module for the connection to the SQLite database.

The core of the toolset is described in section 3.4. It creates the new index database and links the entities to a map in the printed atlas. If some of the required attributes are missing (German name, French name, Italian name, geographical object and country) the tools described in sections 3.1 – 3.3 help to complete the data by means of the existing *index_old* table. Finally, the index for the printed atlas is derived on the basis of the tables *index_new* and *page_appearance* and a set of rules defined by developers.

The input data for creating the new index is typically a number of point features with a German name attribute, stored as a shapefile[2], which allows us to easily calculate the coordinates of the entity. Table 1 shows the attribute table of a typical input dataset.

name_de	lat	lon
Kilimandscharo	-3.066667	37.359167
Warschau	52.216667	21.033333
San José	9.924722	-84.078056

Table 1. Attribute table of a shapefile that can be used as input data for the index tools. The attributes *lat* and *lon* are not required as they are part of the geometry stored in the shapefile.

3.1. Check Geographical Object Tool

name_de	lat	lon	geotype
Kilimandscharo	-3.066667	37.359167	Mountain
Warschau	52.216667	21.033333	Major city
San José	9.924722	-84.078056	City

Table 2. Attribute table after the usage of the *Check Geographical Object Tool*. Additionally a log file is created, which informs the user that various entries with German name San José exist.

If the input data lacks the geographical object (as in table 1), the *Check Geographical Object Tool* checks if the entities exist in the *index_old* table

[2] Shapefile: quasi standard format for geospatial vector data

based on the *name_de* attribute. If exactly one match is found, the geographical object stored in *index_old* is added to the attribute table (table 2). If more than one match is found, the geographical object is taken from the first entry in *index_old* and a warning is written to a log file.

3.2. Add Country Code Tool

The next step in completing the input data is calculating the country membership and comparing it with the specification in the existing *index_old* table. This tool calculates the country membership with a spatial join analysis. The input features are joined based on their spatial relation with a dataset containing the administrative boundaries. For entities within Switzerland the detailed swissBOUNDARIES3D dataset (Swisstopo 2013) is used. Entities outside Switzerland are joined with the Global Administrative Areas dataset (GADM 2012). The information about the country where an entity is located is written to the attribute table of the input data (table 3). In addition, the administrative sub-unit (e.g. federal state) is also written to the attribute table. This attribute is not used at the moment for creating the atlas index, but could be used in the future.

Feldfunktion geändert

Feldfunktion geändert

Name_de	lat	lon	Geotype	country_code
Kilimandscharo	-3.066667	37.359167	Mountain	TZA
Warschau	52.216667	21.033333	Major city	POL
San José	9.924722	-84.078056	City	CRI

Table 3. Attribute table after the usage of the *Add Country Code Tool*. The country membership is stored as ISO 3166 ALPHA-3 code. The administrative sub-unit is also added to the attribute table (not shown in this example table).

Then, the entities are again checked against the *index_old* table. If an entry with the same German name and the same geographical object is found, the calculated country membership is compared with the specification in the *index_old* table. If the data differs a warning is written to a log file and it has to be revised manually.

3.3. Translate Entity Tool

To complete the preparation of the input data, the German term is translated to French, Italian and English (although English is not required). The exonyms are again taken from the *index_old* table. To avoid mismatching translations, the country membership and the geographical object are also requested parameters for the *Translate Entity Tool*. Table 4 shows an ex-

ample, where mismatching translations can be avoided due to the attributes *country_code* and the *geotype*.

name_de	name_it	name_fr	geotype	country_code
Rom	Rome	Roma	Major City	ITA
Rom	Rom	Rom	River	CHE

Table 4. The two entities have the same German exonym. The correct French and Italian exonyms can be extracted from the *index_old* table due to the *geotype* and *country_code* attributes.

3.4. Link with new Database Tool

The *Link with new Database Tool* is the main tool of the toolset. It creates the table *index_new* and links the entities with the atlas pages. It performs two actions. First, it checks if an entity is already in the *index_new* table. Therefore this tool performs the following tests:

1) Does an entry with the same German name, the same country and the same geographical object exist in the *new_index* table?

   No: The entity is added to the new index database. Yes: the second test is performed:

2) Based on the coordinates stored in the shapefile and the coordinates stored in the *index_old* table the distance between the entity in the input data and the found match in the *index_old* table is calculated using following formula:

$$dist_{AB} = \arccos(\sin(lat_A) * \sin(lat_B) + \cos(lat_A) * \cos(lat_B) * \cos(lon_B - lon_A))$$

   The distance is then multiplied by the scale of the specific map to get the distance between the two points in the printed map. If the distance in the map is below an arbitrary threshold it is considered the same entry and not written to the database. Accordingly the entity is considered as new and stored in the database if the distance is above this threshold.

The second step of the *Link with new Database Tool* links the entity from the input shapefile with the specific map in the atlas (table *page_appearance*). This relationship is then used in the *Create Index Tool* (next section).

3.5. Create Index Tool

Finally, the index based on the tables *new_index*, *atlas_content* and *page_appearance* for three languages is created. The name index consists

of the name, the coordinates, the geographical object, the country membership and exactly one page number. It is yet undefined which page number will be printed to the name index. This tool allows the user to define and select rules for the name index creation. The following rules are included in the *Create Index* tool:

- *First appearance*: The page number of the first appearance in the printed atlas is used.
- *Scale dependent*: The page number of the map with the smallest scale that contains the entity is used
- *Type dependent*: It can be defined, which map types precede others. E.g. an entity that appears in a general map and in a city map will have the page number of the general map in the geographical name index, because general maps precede city maps.

4. Using the Index Database as Search Functionality in an interactive Atlas

Development and production of school atlases is currently in a stage where GIS systems and databases get more and more important (Hurni 2004). Together with the increasing trend to use computer technologies in the classrooms (Welling & Stolpmann 2012) the importance of interactive school atlas systems is raising. When developing such an interactive atlas as a complementary product to a printed atlas the index database can be adapted as a search database.

The already existing SWA interactive has a search functionality (Cron et al. 2009), which is currently independent from the printed atlas. However, it serves as an example that the same database structure and attributes are required for generating a printed geographical name index. An index database can be efficiently adapted for an interactive atlas in case of cross-media publishing. When browsing maps by geographical entities, the database (table *new_index*) is queried and relations to maps are listed (table *page_appearance*) (Figure 3).

Feldfunktion geändert

Feldfunktion geändert

Feldfunktion geändert

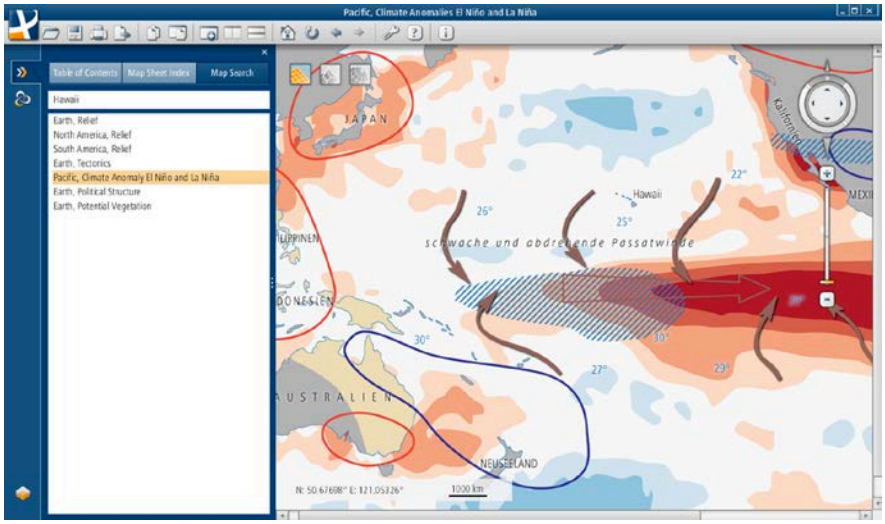

Figure 3: Searching for Hawaii in the SWA interactive. A drop-down list shows the matching maps.

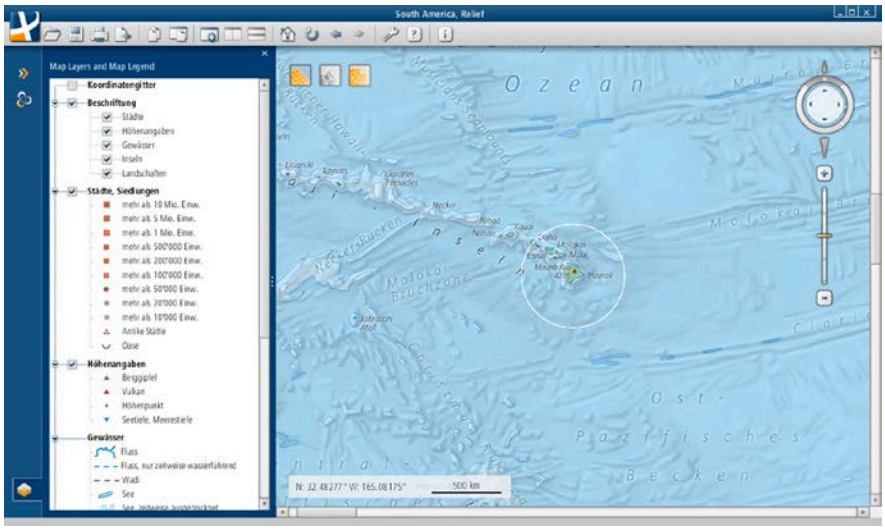

Figure 4: Visualizing Hawaii in the map 'South America, Relief' of the SWA interactive. After choosing the map from the drop-down list (figure 3), the geographical entity is visualized with a white circle.

In a printed atlas the search result is a line of text that contains corresponding maps and coordinates. The user then needs to browse the specific page and find the desired entity based on the coordinates. In an interactive atlas the specification of the coordinates in the database can be used to visualize the search result (Figure 4).

5. Conclusion and Outlook

A geographical name index is often fulfilled at the end of the atlas authoring process and often typed and amended manually. The presented self-developed tools improve this workflow by continuously building up an index database that contains all the geographical entities that appear in the next edition of the SWA. Moreover, it contains the relationships between each entity and its corresponding maps. The necessary data is saved time-efficiently to the index database. Finally, the extraction of the geographical name index for a printed atlas is done automatically on the basis of a set of rules.

In case of cross-media publishing such an index database is of greater value. The search functionality for an interactive atlas system can be efficiently established, due to the requirement of the same database structure and attributes as for the generation of a geographical name index in a printed atlas.

References

Cron, J., Marty, P., Bär, H.R., Hurni, L. (2009) Navigation in school atlases: functionality, design and implementation in the "SWISS WORLD ATLAS interactive". Proceedings of the 24th International Cartographic Conference ICC 2009 Santiago de Chile.

Cron, J., Moll, P., Haeberling, C., Hurni, L. (2011) 100 Jahre "Schweizer Weltatlas" und neuer "Schweizer Weltatlas interaktiv". Kartographische Nachrichten 61-1:32-35.

ESRI (2012) ArcGIS Help 10.1 - Welcome to the ArcGIS Help Library. http://resources.arcgis.com/en/help/main/10.1/. Accessed 12 March 2013.

GADM (2012) Global Administrative Areas (boundaries). University of Berkeley, Museum of Vertebrate Zoology and the International Rice Research Institute. http://www.gadm.org/. Accessed 12 March 2013.

Haeberling, C., Bär, H.R., Hurni, L. (2011) The new web-based "Swiss World Atlas intractive" - characterisation of the cartographic representations and the functionality for a modern geographic education. Proceedings of the 25th International Cartographic Conference ICC 2011 Paris (France).

Hurni, L. (2004) Vom analogen zum interaktiven Schulatlas: Geschichte, Konzepte, Umsetzungen. Wiener Schriften zur Geographie und Kartographie 16:222-231.

Hurni, L., Spiess, E. (2010) Schweizer Weltatlas / Atlas Mondial Suisse. Schweizerische Konferenz der kantonalen Erziehungsdirektoren, Zürich.

Hurni, L., Wondrak, S. (2011) Map authoring in the age of multidimensional geodata processing. Proceedings of the 25th International Cartographic Conference ICC 2011 Paris (France).

Owens, M. (2007) Query anything with SQLite: virtual tables are the foundation of powerful features. Dr. Dobb's Journal 31(12):24-28.

Spiess, E. (2002) Schweizer Weltatlas / Atlas Mondial Suisse. Schweizerische Konferenz der kantonalen Erziehungsdirektoren, Zürich.

Swisstopo (2013) swissBOUNDARIES3D. Bundesamt für Landestopografie swisstopo. http://www.swisstopo.admin.ch/internet/swisstopo/en/home/products/landscape/swissBOUNDARIES3D.html. Accessed 12 March 2013.

Welling, S., Stolpmann, B.E. (2012) Mobile Computing in der Schule - Zentrale Herausforderungen am Beispiel eines Schulversuchs zur Einführung von Tablet-PCs. Jahrbuch Medienpädagogik 9:197-221.

Feldfunktion geändert