World political map from OpenStreetMap data

Dražen Tutić ♦, Tomislav Jogun ♦, Ana Kuveždić Divjak ♦ and Martina Triplat Horvat ♦

Faculty of Geodesy, University of Zagreb, Zagreb, Croatia

ABSTRACT
The paper presents the development of an automated procedure for creating a small-scale world political map from OpenStreetMap (OSM) data, and the map itself. A novel approach was used for cartographic processing, while the fitness for use of OSM data for this task was evaluated. It was anticipated and shown that creating a world political map from OSM data is a methodologically and technologically demanding task. The result was a political map of the world at a scale of 1:30,000,000, showing independent states, dependencies and areas of special sovereignty as in the OSM data set, with no adaptation to specific political recognition issues. A high degree of automation built on open-source software was achieved. The resulting map is an intermediate stage of production, requiring modest manual intervention for the final map. By allocating a code to the community (http://github.com/GEOF-OSGL/OSMPoliticalMap), we have provided opportunities for its continuing development.

1. Introduction

Children and adults often first encounter maps in geography books and atlases in the form of a small-scale world political map, showing countries in different colours, with the characteristics of a prototypic map (MacEachren, 1995). This means it vividly presents basic information on countries of the world. Political maps belong to a long mapping tradition that began with Abraham Ortelius’ atlas Theatrum Orbis Terrarum in the sixteenth century (Eckert, 1925). A political map is often used in papers on geoscientific research to show territorial coverage and the names of spatial units (Raisz, 1948).

Scientific interest in political maps has decreased in last few decades, mostly for two reasons. The first is the prevailing opinion that the theoretical and methodological background has already been firmly established. Yet, ever since political maps began to be produced, the question of the geographical content to be shown has been pertinent. At the very least, a world political map should show the land borders of political entities, capitals, world coastlines and a graticule. However, other geographical features are sometimes included, for example, hydrography, places, and geographical names. The greatest area of disagreement is in relation to the main content, that is, political entities, and concerns the various international statuses of certain entities. There are also many areas open to dispute, where borders are undefined. The second reason is the declining rate at which new political entities are established. This is indeed true, if we compare the present day with the twentieth century, but since the world is in a state of constant political flux, political maps deserve appropriate attention.

New potential in geospatial data collection, in near real time, has facilitated the development and updating of world political maps. Modern information and communication technology allows immediate access to final cartographic products and geospatial data. So it could be concluded that the production of a world political map is a relatively simple task. Although the Internet has facilitated mapping and the availability of maps, it also introduced liberalization, which in many cases has reduced the quality of these products and services. The current situation among a plethora of world political maps on the Internet is that not many of them are of high enough quality to comply with cartographic principles. If the term ‘world political map’ is entered as an Internet search item, many results show world maps in cylindrical projections, which are not at all suitable for world maps. Cartographic generalization and map aesthetics are also sometimes inappropriate.

On the positive side, high-quality, free editable contemporary world political maps can be found. One such political map of the world is by the cartographer Tom Patterson. On his Shaded Relief site (http://www.shadedrelief.com), three political maps of the world are given, in the Patterson, Natural Earth and Natural Earth 2 map projections. Political and territorial entities are depicted by vignetting, that is, colouring
narrow zones along international boundaries. This leaves enough room to represent shaded relief, rivers, lakes and populated places. Patterson’s world maps derive from the Natural Earth vector data, namely coastlines, rivers, lakes, country boundaries, and populated places. Natural Earth is a public domain map data set available at scales of 1:10,000,000, 1:50,000,000, and 1:110,000,000. It provides the geometry and attributes needed to create visually pleasing small-scale world maps (http://www.naturalearthdata.com).

Another political world map is published by the Central Intelligence Agency (https://www.cia.gov/library/publications/the-world-factbook/graphics/ref_maps/political/pdf/world.pdf). The map is in the Robinson projection and shows only political and territorial entities, cities, coastlines/oceans and large lakes while rivers and other physical and geographical features are not shown. This lack of additional content is not a shortcoming, because it makes the political and territorial organization of the world stand out clearly. The major inadequacy of this map is its generalization of coastlines.

OpenStreetMap (OSM) is one of the most popular mapping services on the Internet (https://wiki.openstreetmap.org/wiki/Main_Page). OSM data are free and presumably based on local knowledge, kept up-to-date and representing the situation on the ground. It should be noted that it differs considerably from data sources traditionally used to produce political maps, because it focuses much more on the accurate representation of reality in detail, rather than to the suitability for any single purpose. In other words, it is excellent for mapping streets or buildings, but if used at the global level, there are two major challenges. The first is the huge amount of constantly accumulating data (mainly at the local level), and the second is data consistency. There are examples of large scale (http://wiki.openstreetmap.org/wiki/Using_OpenStreetMap/Map_examples) (O’Brien, 2009) and small scale maps derived from OSM data (http://www.imagico.de/map/techniques_en.php), but none is a world political map.

Therefore, the main idea of our research was to evaluate the applicability of OSM data for the automated creation of a small-scale, world political map. It should be emphasized that it was not our intention to reflect any particular point of view regarding political entities, but to model the cartographic process based on input data. We did not aim to produce a final map, but to provide an intermediate one, requiring considerably less manual editing to produce a final world political map.

The main objective of this work was to create a small-scale, Main Map, respecting scientific cartographic principles, and to automate the process of extraction, generalization and data visualization.

2. Methods
The methodology involved map planning, data preparation and processing, quality assessment, dealing with exceptions and designing a map suitable for printing.

2.1. Map planning – content, format, presentation medium and mathematical elements (map projection, scale)
It was decided that the main content of Main Map should be entities which had been assigned official ISO 3166-1 codes (independent countries, dependent territories and special areas of geographical interest). This list was created through cooperation between the United Nations and the International Organization for Standardization, so it can be considered as relevant and neutral. At the end of 2016, it covered 249 entities (http://www.iso.org/iso/country_codes), with the exception of the Republic of Kosovo, whose unofficial code XK is used by European Commission and other organizations (http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Country_codes).

Therefore, the total number of political and territorial entities shown on the map was 250. Other features to be included were coastlines, lakes, rivers, cities, and the names of the oceans and seas.

The Winkel Tripel map projection with the standard parallel at 50°28′ latitude was selected. Francula (1971) has demonstrated the suitable properties of this map projection for world maps, and since 1998, the National Geographic Society has used it for world maps (http://www.csiss.org/map-projections/microcam/mapnews.htm). Map projection for all data can be changed using the dedicated Python script and a map projection can be set in advance using PROJ4 syntax. The Winkel Tripel projection is defined by the parameters: proj=winitri, ellps=WGS84, lat_1=50.4597762522. The software used (QGIS, GRASS GIS) does not allow this map projection to be set as a reference (due to the lack of reliable equations for inverse mapping), so the equidistant cylindrical projection was assigned to the project file in order to process data. That meant that on the fly map projection changes to the derived data could not be performed.

Although a scale of 1:20,000,000 would be more suitable for this purpose, the selected map scale was 1:30,000,000, assuming that the map would fit onto an A0 sheet of paper. For a world map with very uneven sizing and arrangement of territorial units, from countries as large as continents to very small, densely arranged countries, this was the smallest scale at which it was still possible to achieve content readability without enlarging certain parts. Scale is a parameter of the automated procedure and affects only
cartographic generalization operators. Users can set different map scales in advance.

2.2. Research workflow

The research methodology and tasks in each phase of the map production process are given in Figure 1.

2.3. Data preparation and processing

The primary source of geometric and attribute data was raw data from OSM, stored in the `planet.pbf` file. We aimed at the simplest data processing possible, based on the OSM recommendations for mapping and tagging features. Thus, we were able to evaluate the suitability or fitness for use of OSM data and their consistency with this specific cartographic application for making a world political map.

The automated process followed the logical sequence of map design, that is, data downloading, filtering, cleansing, map projection transformation, generalization and symbolization. This process was repeated for each object group (mathematical map elements, coastlines, political and territorial entities, lakes, rivers, cities, and oceans, seas and bays), including adjustments for specific properties.

Relatively small changes to the OSM database were needed to show certain objects on the map, for example, missing ISO codes for political or territorial units. New features were added to the group of oceans and seas, and in some places, duplicated objects were deleted. Rivers were the only object group that required special processing, but the results were only partly satisfactory. According to OSM recommendations, each river should be defined as a single relation containing ways of river centreline (http://wiki.openstreetmap.

Figure 1. Flowchart of the automated map creation procedure.
Vector-based processing and generalization did not produce acceptable results, due to topological inconsistencies and data completeness quality. Finally, we applied raster-based generalization inspired by the work and results of Christoph Hormann (http://www.imagico.de/map/water generalize_en.php). More satisfying results with vector- or raster-based generalization could be achieved if the relations for rivers in OSM become more consistent. Cities were selected based on population figures (>500,000) and their capitals according to their roles as the administrative centres of entities. Objects representing cities were unevenly distributed due to the rapid expansion of large population density areas in developing countries during the last 25 years, and also due to the varying definitions of a city in different parts of the world (UN, 2014). In this object group, attention should be paid to the emergence of new entities with large populations, which are actually parts of larger cities. To address this problem, at least in part, we created a clustering operator which concentrated the most populated entities within a radius of 40 km. This ensured that symbols did not overlap and made label placement less complicated.

Problems occurred in the political entities object group due to conflict between the rigid and simple criteria for automated filtering (ISO 3166-1 code), and their dynamics in the OSM. An example is Bir Tawil, a tiny trapezoid of desert on the border between Egypt and Sudan, which, according to international treaty, cannot be claimed by either country (Jennings, 2012), so it was not given a symbol on the map.

2.4. Cartographic generalization

Cartographic generalization is the most challenging process in the automation of any map production. The map generalization operators required were selection, line simplification and smoothing, exaggeration and displacement.

Selection referred to the choice of object type, and choice of object according to attribute value and/or geometric characteristics (length of a line, surface of an area). Selection was achieved using simple relational operations.

Line simplification using the Douglas-Peucker algorithm was applied to reduce the amount of data (points) prior to final generalization. This algorithm does not provide visually acceptable results, so a simplification and smoothing algorithm with an area-preserving property (Tutić & Lapaine, 2009) was used. It takes the map scale as an input parameter, which makes it especially suitable for application in automated cartographic systems.

Exaggeration was used for small islands which are also territorial units, enlarging them to minimum dimensions using an algorithm written as part of this software. Groups of small islands were represented by the three largest islands among them, enlarged to minimum size. Optimal parameters for this operator should be investigated in the future.

Displacement was not used in the automatic, but in the final phase of map editing, for cities and small islands that overlapped, or were too close to each other for successful label placement.

2.5. Map language

OSM enables storing the names of objects in multiple languages. It is a prerequisite for the automatic creation of derived maps with names in the desired language. The key ‘name’ is used for local feature names, while ‘name:<ln>’, for example, ‘name:en’ (English) or ‘name:hr’ (Croatian) is used for the names of features in a specific language. The user sets the desired map language in advance and for each feature on the map, the keys ‘name’ and ‘name:<ln>’ are extracted. When the value in the selected language is not given, the value ‘name’ is used. The best way to make a complete map in a specific language is to add names to OSM features directly.

2.6. Cartographic visualization

In traditional world political maps, two common approaches to colouring political-territorial entities are used — filling polygons and vignetting border areas. On this world map, a combination of both was used. Political-territorial entities were shown in different pastel hues used to create primary visual differences. By vignetting, that is, colouring a narrow zone along a border, a subtle emphasis was added to boundaries.

The initial colouring of political-territorial entities was achieved by using the TopoColour plugin for QGIS, based on the four-colour theorem (http://github.com/nyalldawson/topocolour). Manual edits were necessary for some neighbouring entities at sea (if they were the same colour) and for Greenland and Antarctica, which are usually white or blue.

The colours used were initially based on the ColorBrewer qualitative pastel colour scheme (http://colorbrewer2.org). However, some modifications were made to achieve the appropriate visual hierarchy and overall appearance of the map.

To achieve the visual hierarchy of labels, it was decided to use the PT Font family (http://en.wikipedia.org/wiki/PT_Fonts) which has both serif and sans serif counterparts, and was also suitable for the fully unified appearance of the map. For example, PT Serif (http://fonts.google.com/specimen/PT+Serif) was used for the names of political-territorial entities, so they could be clearly differentiated from the names of cities, shown in PT Sans Narrow (http://
fonts.google.com/specimen/PT+Sans+Narrow). Letter size and spacing were functions of the entity area, defined by the conditional expressions in QGIS. Manual correction of automatically placed labels in QGIS was necessary. In areas with a high density of political-territorial entities, as in the Caribbean (Figure 2), tick lines were used to place labels and connect them to features. Text hyphenation was performed manually, depending on the number of words and shape of the entity.

Attributes that were set manually (colour, label coordinates and text wrap) were stored in the reference map data set and copied to the newly created map data set. Thus, users do not need to repeat this manual work, at least when the same language (English) is used as on the reference map.

3. Conclusions

The result of this research was a Main Map created exclusively from OSM data. This was the first such attempt to elaborate on geometry processing and evaluate the fitness for use of OSM data for this purpose.

The map was created at the scale of 1:30,000,000 in the Winkel Tripel map projection. An algorithm with area-preserving properties was applied to the generalization of lines and polygons, and a new algorithm was written for the generalization operator for the exaggeration of small islands. Independent states, dependencies and areas of special sovereignty were shown as in the OSM data set, with no adaptation to specific political recognition issues.

Although the resulting map was of high quality, relevant, up to date and scientifically based, it served primarily to develop an automated procedure and to accelerate map production of this type. It should be considered as an intermediate, rather than a final product, though considerably fewer manual interventions are necessary for the production of a final map. Since it is completed through automation, the role of the cartographer is reduced to map-planning, running a programme for automated map production, data quality control, small changes in geometry and attributes, adjusting the positions of the labels, and the final visualization. There is potential for improvement in the quality and speed of data processing. With the anticipated development of OSM, parts of some scripts dealing with exceptions could be eliminated, making the process more straightforward in the future. For example, coastlines in the OSM data set are already highly consistent and there is no need to deal with exceptions.

Software

Open-source programmes and modules for spatial data (GRASS GIS, QGIS, ogr2ogr, osmfilter, osmtojson, GDAL/OGR, pyproj) are used for data extraction and processing. Automation is mostly achieved as an orchestration of existing modules and functions, resulting in more robust solutions, considering data complexity and the actual process.
We opted for open-source software because it was free, upgradable to our needs, and provided a good insight into all data processing procedures. Moreover, by releasing the code to the community, we have provided opportunities for ongoing improvement and development in this area of research.

The software architecture for the automated production process of a world political map is hierarchical and branched, and consists of Linux shell scripts (*.sh) and programmes written in Python (*.py) for general and special tasks. General Python programmes cover all or most object groups (e.g. changing the map projection, simplification, or generalization), while special Python programmes solve specific geometric-attribute problems within an object group, or even an object itself (e.g. creating the inner ring of the Caspian Sea in the Eurasian polygon in the coastlines object group). The main executive script _0_PoliticalMap.sh runs from the command line and invokes other scripts and Python programs in the chain. An example of the procedure for creating the coastlines object group is given in Figure 3.

The complete programmes and scripts used in the research were placed on the GitHub as OSMPoliticalMap (http://github.com/GEOF-OSGL/OSMPoliticalMap), and published under the terms of the GNU General Public License 3.0 (http://www.gnu.org/licenses/gpl.html).

Data

The data for Main Map were OpenStreeMap data in the planet.pbf file. The planet files remained under the same licence as the master OSM geo-database from which they were extracted – currently the Open Database License. The planet files previous to 12 September 2012 have a Creative Commons Attribution-ShareAlike 2.0 licence.

The processed data for the designed map and the map in .pdf format were published on GitHub and licensed under the Open Database licence and Creative Commons Attribution-Share Alike 4.0 licence, respectively (http://creativecommons.org/licenses/by-sa/4.0).

The final map was produced using Quantum GIS.

Acknowledgements

The authors would like to thank all volunteers, organizations and institutions who participated and contributed to the development of open geospatial software and data, and without whom this research and paper could not have been achieved.

Disclosure statement

No potential conflict of interest was reported by the authors.
ORCID
Dražen Tutić  http://orcid.org/0000-0001-9336-2892
Tomislav Jogun  http://orcid.org/0000-0002-8685-8328
Ana Kuveždić Divjak  http://orcid.org/0000-0003-1059-8395
Martina Triplat Horvat  http://orcid.org/0000-0002-8304-7474

References
Eckert, M. (1925). Kartenwissenschaft. Forschungen und Grundlagen zu einer Kartographie als Wissenschaft (2nd ed.). Berlin: Walter de Gruyter.
Frančula, N. (1971). Die vorteilhaftesten abbildungen in der atlaskartographie (Unpublished doctoral dissertation).
Raisz, E. (1948). General cartography (2nd ed.). New York, NY: McGraw-Hill Series in Geography.
Tutić, D., & Lapaine, M. (2009). Area preserving cartographic line generalization. Cartography and Geoinformation, 8(11), 85–100.
UN. (2014). United Nations, department of economic and social affairs, population division. World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352).

Institut für Kartographie und Topographie der Rheinischen Friedrich-Wilhems-Universität Bonn, Bonn. Jennings, K. (2012). Maphead: Charting the wide, weird world of geography wonks. New York, NY: Scribner.
MacEachren, A. M. (1995). How maps work. New York, NY: The Guilford Press.
O’Brien, O. (2009). Creating and maintaining street orienteering maps using openstreetmap. In Proceedings of GISRUK (pp. 415–420), University of Durham, Durham.