ThemaMap: a Free Versatile Data Analysis and Visualization Tool

Jacques Madelaine, Gaétan Richard, Gilles Domalain

ThemaMap: a Free Versatile Data Analysis and Visualization Tool

Jacques Madelaine Gaétan Richard
GREYC - Université de Caen Basse-Normandie, CNRS
{Jacques.Madelaine, Gaetan.Richard}@unicaen.fr

Gilles Domalain
IRD Sète
Gilles.Domalin@ird.fr

Abstract

The ThemaMap project aims to provide a friendly and versatile thematic cartographic tool available as an Open-Source free software. It allows various classic and fancy data visualizations. Data table manipulation allows simple statistics indicators computing or data (alphanumeric and geographic) grouping and filtering.

This paper explains how it can be a good candidate tool to make Open Data accessible to everyone.

I. Context

The directive 2003/98/EC [1] of the European Parliament of 13 June 2013 on the re-use of public sector information, known as Open Data, will allow citizens to have access to more and more public data through an increasing number of web sites offering data to the public, as the French portal [3].

However, this data is often given in a "raw form" (presented as a table and stored as a CSV file). This format is suitable for programs and allows easy use and manipulation of the data, but with no benefit for their accessibility or readability. Indeed, the exploitation of raw data requires often specialized skills and therefore is not directly accessible. This defect is also pointed out by the authors of the directive, since the direct use of data by a general public is not considered. It is suggested to use intermediaries such as companies specialized in data processing.

It is especially the case for geolocalized information. Maps are a meaningful visual representation of that kind of data, but in order to be synthetic it needs for some more computations : various groupings, means and other statistics indicators. As a proof, during the hackathon organized on the initiative of the Open Knowledge Foundation [2], two workdays of several specialist engineers were required to produce maps of traffic wrecks and maps combining at town level the number of police force and their tax potential.

ThemaMap [4] could be a tool to achieve (more) quickly this work. It enables the user to represent visually geolocalized data and make it accessible to everyone in a attractive manner. Moreover, ThemaMap allows a *free* interpretation in the sense of free software: it is possible to study, modify and distribute the representation along the data.

II. Computer aided thematic visualization

A first functionality of the ThemaMap project is the ability to create maps visualizing geolocalized data. To achieve this, the project support several usual geographic data format such as Esri shape files, MIF files, CSV files with latitude/longitude columns. It is also able to connect using a JDBC connector to various relational database management systems such as PostGIS, MySQL, Microsoft Access or SQL-Server.

The geometries are easily represented. Colors, lines width are customized interactively using a visual interface. Statistical variable representing a density may be depicted with choropleth [1]. Other statistical variables may be represented using discs or regular polygons of proportional sizes. Multiple data sets are represented using charts: histograms, pie, lines or radar charts. Flow maps are also available and can be used to visualize GPS traces. Furthermore, all the symbols, including the charts elements, may be colored with respect to another variable and are fully customisable.

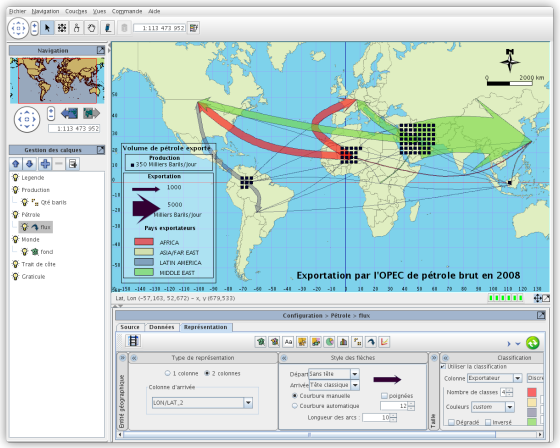

Figure 1: *ThemaMap GUI*

The wide range of visualization tools along with the interactivity allows the user to create, test and improve its maps until it is satisfied. In particular, when confronted to the mass of open data, this tool helps to quickly see interesting data and possible interpretation. The visualization can be either saved as a whole or exported in several picture formats.

Aside for those classical representations, the ThemaMap project allows inclusion of additional representation by using some generic tools provided inside the editor or by adding it directly in the open-source code.

Due to the quantity (and quality) of data, it is often impossible to represent all the interesting in-

[1] A *choropleth* is a map where areas are shaded in proportion to the measurement of a statistical variable

formation on a single map; the project can generate sets of maps according to some criteria and even animate them.

III. Data manipulation

The data used to make the map is always presented to the user as a table. Several data manipulation tools, the so-called adapters, may transform and adapt the data table.

Figure 2: *Available data manipulation tools*

The table may be sorted, columns or lines may be hidden. It is also possible to compute a new data column defined with a formula. The formula language is similar to the one present in spreadsheet applications, except that, here, the formula is the same for all the cells of the column. In addition to the basic arithmetic operators, the characters strings manipulations, the usual statistical analysis tools (mean, max, average, ...), ThemaMap also supports many different classifiers and many of the geographical primitives defined by the OpenGis consortium for SQL [5].

One other aspect of data manipulation consists in combining, selecting and grouping the desired data. These manipulations are also available on the geometries: intersection of layers, transformation of a sampling points set to a regular grid, with the possibility of kernel smoothing, and spatial aggregation.

These manipulations can be difficult in the case of open data since the quantity of data can be quite large. ThemaMap allows to import selectively and incorporate several sources allowing a real use of the data cube.

In ThemaMap, the data manipulation adapters act as a pipeline. This allows to separate the different steps and to be able to reuse some of them. It is also possible to change on the fly any of those adapters. In particular, animations are treated as such. It is then easy to offer selections of some specific subsets along with making manipulations of the result.

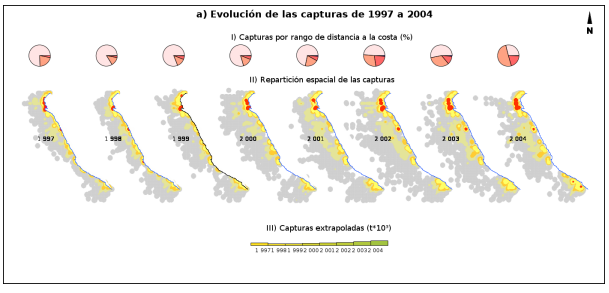

Figure 3: *Changes over time of a phenomenon*

IV. Towards the general public

For the moment, ThemaMap is a functional prototype framework to study and diffuse Open geolocalized data. It already allows intermediaries – persons who have the ability to analyze and understand the content of the data e.g. journalists, analyst, teachers – to deal more efficiently with the mass of open data. They can analyze and publish their interpretations to the general public. It also facilitates exchange, discussion and cooperation between these people.

It also gives to the general public a quick access to a global vision of specific subsets of geographical data using the default representations. People may then have access to the specific portion of data which concerns them (in the sens of geographical locality). In this direction, one very interesting work would be to create a full enhanced *atlas* with available data.

The ultimate goal is to smooth the gap between the data and the general public. In this context, the intermediaries do not only produce a map, they can give access to the whole creation process of the map. According to their ability, people will just see the resulting visualization, or they may explore the method used to construct it. In that latter case, they can follow up. They may modify and manipulate this visualization to better understand it as well as its context. Even more, they may create their own visualizations. This process allows a thorough apprehension of the problematics underlying the given data.

References

[1] Directive 2003/98/ec. `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:EN:PDF`.

[2] Open knowledge foundation hackathon. `http://fr.okfn.org/2014/08/09/retour-sur-le-premier-hackathon-sur-les-donnees-du-ministere-de-linterieur/`. Accessed: march 2015.

[3] Plateforme ouverte des données publiques françaises. `http://www.data.gouv.fr/`.

[4] Themamap. `https://themamap.greyc.fr/en`.

[5] Opengis implementation specification for geographic information - simple feature access - part 2: Sql option. `http://portal.opengeospatial.org/files/?artifact_id=25354`, August 2010.