Jean-Luc Arnaud*

Production of georeferenced data – use, cost and accuracy

Keywords: Cartographic Heritage; georeferenced data; CartoMundi

Summary: The tools operated for research of cartographical documents by geographical location (*CartoMundi*, MapRank…) work with georeferenced data. The production of the georeferenced data must operate with several parameters but past cartographers have rarely indicated the means and the tools used during their work. In this context, the production of georeferenced data is as more difficult as the documents are old.
This paper discusses different methods to produce georeferenced data for old documents. We examine their advantages and their defaults, about their accuracy and their cost. It shows that very simple methods can be very effective.

The tools operated for research of cartographical documents by geographical location (*CartoMundi*, MapRank…) work with georeferenced data.
The surface of the earth is approximately spherical and the maps are printed on flat sheets of paper. The means of transition from one system to another are numerous. But no one of them is perfect and they present defaults and advantages, for this reason they have been multiplied during the last three centuries.
The production of the georeferenced data must operate with several parameters but past cartographers have rarely indicated the means and the tools used during their work. In this context, the production of georeferenced data is as more difficult as the documents are old. Most of the time, it is possible to get a great accuracy but it's costly and it's not always necessary. In fact, we don't use old maps to prepare the construction of a high speed railway or to fire shells. For the works in the field of history or to locate old documents with a web site, some accuracy problems are sometimes of little importance.
This paper discusses different methods to product georeferenced data for old documents. We examine their advantages and their defaults, about their accuracy and their cost. It shows that very simple methods can be very effective. It shows also that their results can be satisfactory even if they seem approximate when they are evaluated with modern instruments.
On one hand, according to the rules of geomatics, the restitution of a sheet index or of the perimeter of a map needs several information.
- data about the model used to describe the sphericity of the earth – ellipsoid (Bessel, Clarke, Fresnel…)
- data about the projection used (cylindric of Mercator, conic of Lambert, Peters, Albers, Bonne, polyconic…)
- prime meridian (Paris, Greenwich, Roma, Constantinople, Pulkova…), they are numerous and some of them are not very well documented.
- data about the implementation of the projection (center of the projection, origin of the grid, unit used: grades or degrees for angular measures, miles, *verstes*, kilometers… for linear measures).
On the other hand, according to the standard for cataloguing cartographic documents, these data can be well described. But we notice that in the catalogues themselves, the fields dedicated to these data are, most of the time, incomplete when they are not empty. This first remark bears witness of a gap between the standard and the information available on the cartographic documents.

* Senior researcher at National Center for Scientific Research – CNRS-Telemme, Aix-en-Provence, France; Founder and head of *CartoMundi* – Online Promotion of the Cartographic Heritage http://cartomundi.eu
[jlarnaud@mmsh.univ-aix.fr]

Available information

In fact, for most of old documents (more than 60 years old) and for numerous more recent documents, information required for georeferencing is incomplete or totally absent.
If they exist, these data are not stated on the documents themselves and they are not easily reachable. Most of the maps published between the mid-nineteenth century and today have been produced by armies. For militaries, information about georeferencing is strategic. We can't pull a shell with a map of which we don't know the process of construction. So this information has not been largely published. Research in archives would be long, costly and would not always be successful. For the old documents, difficulties are numerous. 1. Authors of the maps have often used approximate methods, more manually adapted than calculated. 2. Several data are approximate. For example, for the prime meridian called *Ferro* or *île de Fer*, two values are available. For some, this meridian is placed 20 degrees West from Paris, for others, it's 0° 23' 14'' closer of Paris. A lot of documents are based on Ferro prime meridian, particularly the maps of Central Europe published during the second half of the nineteenth century by the Austro-Hungarian Army. This difference is not negligible. At the Paris latitude, it corresponds to more than 30 kilometers. 3. Some methods have not been preserved and they are lost. For example, that is the case for the famous map of France prepared by the Cassini family during the second half of the eighteenth century. Evidently, this map has been built on the basis of an ellipsoid but, in spite of historians' research, we did not find any information about it. The engineers of French National Institute for Cartography and their Belgium colleagues did not achieve the restitution of Cassin's parameters.

In this context, one wonders how it was possible to publish several versions of this map online. Publishing a map online, assumes that a numerical reproduction is available. Then, it is possible to proceed to its georeferencing. The process is based on landmarks identified both on the old document and on a new georeferenced one. This method, also called rubber sheeting, is practiced for about twenty years. But it works sheet by sheet and it does not allow finding the parameters used for the construction of the document.
This method is a kind of roundabout way for georeferencing. Unfortunately, it is not usable widely because, among the millions of documents held by the map libraries and for which we have incomplete or no information about georeferencing, the part which has been scanned is very small, probably less than 1 %. In spite of the recent improvements of the scanners technologies, in spite of the lowering of the scanning costs, we can't expect an increasing of more than one percent per year in the next 15 years.
Otherwise, to manage correctly the reproductions, most of the libraries don't scan the documents which are not well catalogued. In this field, the delay is enormous. So, in 2011, the Maps and Plans Department of the French National Library, which hold about one million documents, had to prepare a list of 5000 maps, copyright free, and well catalogued for scanning. It was not so easy to achieve this list.
According to my experience, this mode of management will rapidly change during the next years. Map librarians will learn how to use the numerical version of the documents for cataloguing. That is what we are doing with *CartoMundi*. In cooperation with researchers of La Rochelle University in the field of automatic recognition of characters and symbols, we are developing a Assisted Cataloguing Program Base on numerical reproductions. This program was presented at the 2012 LIBER Conference.

My third point deals with the question of the quality of the georeferencing work.
In France, the Cassini's map has been the subject of several experiments. Its ways of use are numerous. This map can be used for a large panel of investigations and scales, from the entire national territory to very local questions. According to the aim of the investigation, quality requirements for georeferencing are not the same, the cost and the results are also different. Let me expose three examples.

* To establish the mosaic for the whole France covered by 170 sheets, the operators of the IGN have selected mostly 3, sometimes 4 landmarks, per sheet. On the base of about 500 landmarks, they got a general document which was published online.
* The engineers in charge of the geographical information system of the Val d'Oise County have used only two sheets of Cassini's map. Each of them has been georeferenced with 30 landmarks per sheet. Their coordinates have been determined from the 1:25 000 scale map of France.
* For a municipality in the surroundings of Toulon (not more than 25 square kilometers), the quarter of the relevant sheet has been georeferenced with about 20 landmarks determined on the base of a topographic survey.

At this very thin scale, the rubber sheeting operates at several levels:
- it positions and deforms the original document according to the new system,
- it corrects survey mistakes; in this case, a problem of orientation of the coastline,
- eventually, it can also compensate for deformation of the support (in this case, a sheet of paper which was wet before printing and then, was dried.

But it's not possible to know the part of each correction.
These three examples are different by the means used, by the cost and by the results. I let you imagine the cost of a treatment of the entire collection according to the process developed in the third example.
So, it can be useful to remind that for several fields of research in history, georeferencing is not necessary. The history of the road networks is older than the new technologies of reproduction and georeferencing. For most of the works, a referencing based on 3 or 4 landmarks per sheet is good enough. On the contrary, for the comparisons of documents from different periods through an historical GIS, higher accuracy is required. But, in this case, the geographical zone is smaller.
On this base, there is not good or bad georeferencing process, there are different results, each corresponds to specific aims.
After this information, I will expose the ways we are following in the context of *CartoMundi*.
First of all, *CartoMundi* (http://cartomundi.eu) is an online collective catalogue of cartographic documents. It is based on several peculiarities. The most important is its tool for researching documents by geographical location. This tool allows to select documents on the base of the territory that each of them represents.
To achieve this, we add to each bibliographical description, a geographic field which describes the perimeter of the corresponding document. This field can be requested by geographical question and its content can be superimposed on a skeleton map. To get this result online, we use mash up technologies and the Google map planisphere.
It would be possible to use other planispheres online like Yahoo map or Bing map for example. They are all of interest but they all present the same default about their accuracy.
As you know, there is no information about coordinates on these planispheres. That is not by chance; this information would show a lot of problems! To realize how these maps are approximate in terms of location, we can compare the vectorial map with the raster satellite image. The difference is sometimes several hundreds of meters. Google earth is not better. The position it displays for the Royal Observatory of Greenwich is a good example. According to the image online, the prime meridian of Google earth is placed at about 40 meters East from the dome of the observatory. There are two possibilities: 1. The prime meridian used by Google earth is the Greenwich one; in this case, it must fit the center of the dome; 2. The prime meridian is the WGS84 one; in this case, the error is not 40 meters, the meridian should be placed 82 meters farther on the East.
In this context it is not necessary to search for a high accuracy in referencing the perimeters of the maps. For this reason, in spite of the lack of information for a lot of old maps, we prefer to get approximate results than to stop working on these documents.

How we proceed when the data are insufficient? We use three main methods

* Limits of the sheets follow a grid of parallels and meridians

In this case we list the coordinates of each line and each column. We don't forget the identification of the prime meridian. Then, these coordinates are mapped in a GIS. If we have information about the ellipsoid we consider them, if we don't have we regard the earth as a perfect sphere.

* Maps with a sheet index printed on a skeleton map

In this case, we proceed with the numerical version of the sheet index as with a topographical map. We use landmarks and rubber sheeting method. Then, in the GIS, we draw the limits of the sheets on the base of the distorted raster image.

* Maps without coordinates and without sheet index.

For these documents we proceed in two stages. First, on a sheet of paper, we draw manually a sheet index. This drawing is a reduction of the original map; it respects the proportion and the orientation of the original. Then we select the landmarks and we measure the position of each landmark from the edges on the original sheets. We draw these points on the sheet index according to the same ratio of reduction. In a second stage, this document is processed according to the second method.

This method is not relevant for cartographic series based on a polyedric projection. But this kind of projection has not been used before the 1860's. At that time, most of the maps bear coordinates. On the contrary, it is very useful for older series, particularly for a lot of small series – from 6 to 20 sheets – published between 1750 and 1850.

Conclusion

Evidently, these methods are far from the rules and the calculations developed by the geomatician engineers. I must notice that it was not so easy to convince our geomatician that this way of work can be useful. Several times he told me that, according to the available data, georeferencing is not possible. We always found a solution.

These methods are based on adaptation manually – do it yourself – more than calculation. On this base, there is a relation between the process of production of the documents and the method used for their referencing. The more they have been produced manually, the more we have to proceed on the same way.

The accuracy we get with these methods is also in relation with the accuracy of the planisphere available online. These approximations are even more insignificant when the scale of the documents referenced is smaller. For large scales, cadastral maps for example, the most important problem is not the referencing, it is the accuracy of the planisphere. In this context, for local data, we planned to use *géoportail* (for France, Spain, Belgium…) rather than planispheres which are better related to small scale series.