



LIGUE DES BIBLIOTHÈQUES EUROPÉENNES DE RECHERCHE  
ASSOCIATION OF EUROPEAN RESEARCH LIBRARIES

*18th Conference of the LIBER Groupe des Cartothécaires - Maps Expert Group  
Barcelona, Spain 17-18 April 2012*

## **CartoTXT - Cataloguing the series in a semi-automatic way**

Jean-Luc Arnaud (Senior researcher at CNRS / Head of CartoMundi)

In mostly map libraries of the world, cataloguing operations are very behind. So, hundreds of thousands documents are not known and condemned to stay in their drawers. However, thanks to the development of new tools for processing and reproduction, the needs of these documents are always increasing.

On a second hand, large scanners become every day cheaper and more efficient. Map libraries are now entering in a new era: a massive digitization process. In this context the lag of non-catalogued documents is a very huge problem because:

1. We do not digitize non-catalogued documents
2. There is no way to get money or forces to catch up.

In this context, the program *CartoMundi*, which initial development is now ending, is setting up the development of a new tool - CartoTXT - to help map curators for cataloguing the series, in a semi-automatic way. This tool will work from digitized copies of the documents. So it will not slow down digitizing process, on the contrary it will promote it.

CartoTXT will proceed, in a semi automatic way, from batch of maps digitized in series. Cartographic series have two very important peculiarities in this process.

1. Metadata are all and always in the margins of the sheets, that is to say outside the zone dedicated to the cartographic representation.
2. The different statements for a series are not very numerous, a lot of them are the same for all the sheets.

On this base, CartoTXT will combine tools of optical characters recognition (OCR) with *thesaurus* of statements. After the initial process, the cataloguer will work on a double screen computer. On the first hand the system displays the digitized version of the document, on the second hand it displays the metadata. For each sheet and each field the cataloguer will check the proposition of the system. It will be interactive: after the treatment of the first sheets of a series, it will have in memory the main statements and their corresponding field. For the statements which are different on all the sheets (proper number, proper title, dates...) the system will not record their value but their position in the sheet, their type, size, color...).

On the base of the widening of the available documentation, the main goal of CartoTXT is to offer to the users currently bad known information, and to allow the development of new fields of research from cartography.