Petr Přidal*, Petr Žabička**

Tiles as an approach to on-line publishing of scanned old maps, vedute and other historical documents

Summary

This paper describes a complete process of publishing of historical maps using a collection of freely available open-source tools. Special attention is being paid to the tiling approach for online map publishing to achieve compatibility with mapping web applications like Google Maps or Google Earth. Format and cartographic projection of the tiles used in such systems is described together with software projects which can be used to produce them. The article also proposes an alternate simplified workflow for scanned old maps publishing. It is targeted on libraries and other memory institutions which are digitizing historical documents but do not have the resources for their correct geo-referencing. The web-based tools necessary for this workflow are now in the development through the OldMapsOnline.org project (http://www.oldmapsonline.org/).

Introduction

There is an enormous amount of old maps and other historical documents in the collections of libraries and archives all over the world. Increasing numbers of these documents are being digitized every year to protect them from excessive use by publishing them online. Many of these documents are physically large and their digital version usually has a very high resolution and the resulting file size makes them unwieldy and unsuitable for traditional online publishing by simple direct uploading onto a webserver. Therefore we have to use an alternative means of online publication, which includes publishing in one of several formats of tiles suitable for different web viewers or use of one of the proprietary commercial image servers. The standardized way of publishing a map - using a map server with support of the OpenGIS® Web Map Service (WMS) protocol - is rarely used.

These diverse forms of publication inhibit or completely prevent us from reusing the published raster images of maps for further analysis or study by comparing maps or other documents by overlaying. Tasks such as examination of the evolution of the landscape depicted on maps from different periods are usually impossible. Ordinarily, the only available means of accessing the published document is through a simple web-viewer. It is nice to have a simple preview of the historical document on a website, but

---

* Mgr, Open Source Programmer and Consultant, Moravian Library Brno, Czech Republic - Czech Technical University in Prague - Project OldMapsOnline.org [klokan@klokan.cz]

** Ing, Moravian Library Brno, Czech Republic - Czech Technical University in Prague - Project Old-MapsOnline.org [zabak@mzk.cz]

there is much more that could be gained when the document is published using open standards.

Therefore, we researched available publishing methods, to find and recommend the best way of fulfilling the following requirements:

- *Fast and comfortable viewer* available for users through the web browser with no special extensions. Data published once should be reusable in different online viewers so the publisher/user could choose (or replace) the technology without the need to republish the raster data itself.
- *Maximum integration of published maps* with existing systems (maps should be available under the WMS standard to allow cooperation with existing GIS software). The workflow should follow standards as much as possible.
- *Presentation in 3D viewers* should be possible. Users should be able to find and see documents in applications like Google Earth, Microsoft Virtual Earth, NASA WorldWind, etc.
- *Easy for data providers.* Libraries and archives should be able to publish raster maps without the need to install and configure new software or hardware systems. No need for an IT professional on their side is one of the main requirements.
- *Expandable and open solution.* Maximum usage of existing Open-Source tools. The whole publishing process should not depend on one product or one company. There should be a possibility of choice.
- *Reasonable price.* The necessary software tools need to be available for a reasonable price or preferably for free.

This paper suggests a workflow which is trying to fulfill all of these requirements. Some pieces of software used in the suggested workflow do not yet have a user-friendly interface, but every described step is technically possible with available tools or at least with prototypes. This paper assumes that the descriptive metadata (name of the author, document title, time of creation, etc.) are stored separately from the raster documents (in formats like MARC21, ISO-19115, FGDC etc.) in a library or other cataloguing system. In such systems it is always possible to specify a link (URL) to the raster data itself and we assume that this kind of interconnection is available to any institution using our workflow to facilitate a basic keyword based search or browsing. Published raster data can later be embedded in a digital library (e.g. Fedora Commons, Greenstone, ...) or a map cataloguing system (e.g. GeoNetwork) to make the effort of efficient metadata exchange through standardized protocols (OAI-PMH, Z39.50 or SRU/SRW) more rewarding.

The Moravian Library in Brno received a R&D grant from Czech Ministry of Culture (DC08P02OUK006 - Old Maps Online) to define and implement this workflow and to develop the missing software tools or their functions. The project is scheduled to end in 2011. Our aim is to make the process not only simple but also more efficient so we welcome any cooperation from other projects and institutions.

The traditional method for publishing scanned raster maps

Let us start with a description of the current standards based workflow for publishing of digitized maps. Unfortunately this workflow includes many steps requiring expert carto-

graphic and IT knowledge. Later on we will compare that with the new workflow suggested for libraries and archives.

For the purpose of publishing any given map it is necessary to know its correct cartographic parameters:

1. **Coordinate system** - old maps differ especially in the position of the Prime Meridians, e.g. it is very common to find Ferro Island or another place as the coordinate system origin on an old map, instead of Greenwich as standardized in 1884.
2. **Ellipsoid** - used for approximation of the shape of the Earth. Old maps do not use today's most common WGS84 ellipsoid (default for consumer GPS satellite receivers), but depend on a different one - e.g. Bessel, Hayford or Krassowski.
3. **The projection method** used for transformation from a round surface to a plane.
4. **Coordinates specifying the area**, which is covered by the map. They are based on the previously defined attributes.

Librarians very often wrongly assume that recording the coordinates which they see on the maps is sufficient for georeferencing. Without specifying the prime meridian, and without conversion of coordinates for different ellipsoids, such a georeferencing is almost useless. Similarly, the raster picture of the map must be precisely warped by an algorithm which uses the original projection of the map, otherwise we would be degrading the precision of the map.

On the other hand, even the approximate bounding box coordinate specification can help to determine the relevance of search results so we have developed a simple on-line application which nicely visualizes the coordinate shift caused by different Prime Meridians and helps to find the bounding box coordinates on top of Google Maps. It also allows drag & drop editing of the specified area. The coordinates are then formatted as MARC21 fields that can be copied directly in a library cataloguing system. This tool is available at http://www.oldmapsonline.org/coordinates/. The newly suggested method described later will allow generation of MARC21 fields as well.

In the standard workflow, the publisher of the map should have GIS software, which he would use to specify the above mentioned cartographic attributes of the map together with the collection of matching ground control points (GCPs) in the image. It is expected that he uses an existing map (which usually has to be bought as well) to specify the ground control point coordinates. The specification of correct and precise GCPs is the most important and most time-consuming part of the work on the georeferencing of an old map. Sometimes it could be really hard to choose the right points so the historical knowledge of how the map was made always helps.

With the knowledge of GCPs applied on the scanned raster image the whole map can be transformed (rectified) into the form compatible with other existing maps. Geometrical accuracy and deformation of a map can also be analyzed and visualized, i.e. by using the MapAnalyst (http://mapanalyst.cartography.ch/) open-source tool which can generate a distortion grid, displacement vectors or scale isolines from a set of GCPs.

To perform the rectification of the georeferenced map it is usually necessary to use GIS software (e.g. the freely available GDAL library which contains the gdalwarp command line tool in FWTools package http://fwtools.maptools.org/). The rectified scanned map is then saved into a standardized format, which has support for geometadata - usually Geo-TIFF, JPEG2000 or proprietary ECW or MrSID.

The next step of the publishing process is the installation of WMS server (i.e. the open-source project UNM MapServer http://mapserver.gis.umn.edu/,also included in FWTools) which allows access to the map through a standardized protocol. A WMS server can return a section of the map based on the request with specified geographic coordinates. It can even warp the image into the different map projection. The OpenGIS® Web Map Service (WMS) standard is implemented in almost all GIS applications and there are several web viewers available for this protocol (i.e. open-source projects like msCross http://sf.net/projects/mscross/, OpenLayers http://www.openlayers.org/ or ka-Map http://ka-map.maptools.org/).

Publishing of scanned rasters as pre-generated tiles

With the increasing popularity of global on-line mapping web applications (e.g. Google Maps, Microsoft Virtual Earth[1], Yahoo Maps) and with the quality of freely available base maps in those applications it seems very practical to publish almost[2] any old maps as a compatible layer on top of these applications.

Luckily, all the well-known web applications use the same spatial reference system for raster maps - a Mercator projection on a spherical approximation of the Earth[3]. Use of the same spatial reference system is also the reason why it is possible to produce a raster map which is compatible with all three services. When the publication process is finished, users can compare old maps with other available map layers including satellite images. Maps can be compared with each other as well.
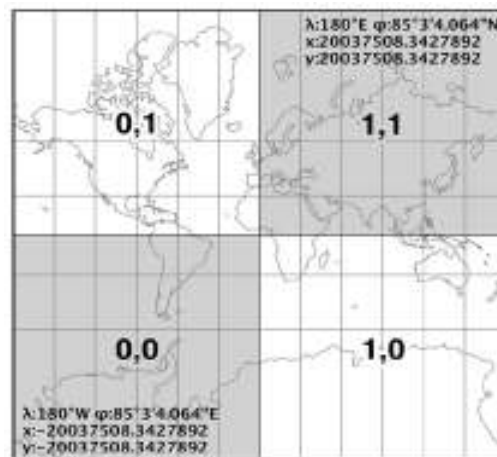
Global Mercator popular on the web: rasters are projected using EPSG:3785 / EPSG:900913, but coordinates displayed to users are longitude/latitude in WGS84.

For all those three systems we can provide our maps in the form of small image tiles. Only the tiles needed to display the area currently on the screen are transported over the network. It is fast, because the tiles are prepared in advance for a fixed area in

---

[1] The Virtual Earth Tile System is described at http://msdn2.microsoft.com/en-us/library/bb25968.aspx

[2] Unfortunately polar areas are extremely deformed by a global Mercator projection

[3] Detailed description of the projection used for tiles is available as EPSG:900913. http://www.spatialreference.org/ref/user/google-projection/

fixed zoom levels, so they are available immediately (no waiting as in the case of dynamically generated WMS requests). A tile is a raster image, usually a JPEG or PNG file. All tiles have the same size, usually 256 x 256 pixels. Tiles form a "zoom pyramid" with the whole globe covered by one tile, in the next zoom level with 4 tiles, then 16 tiles, etc.
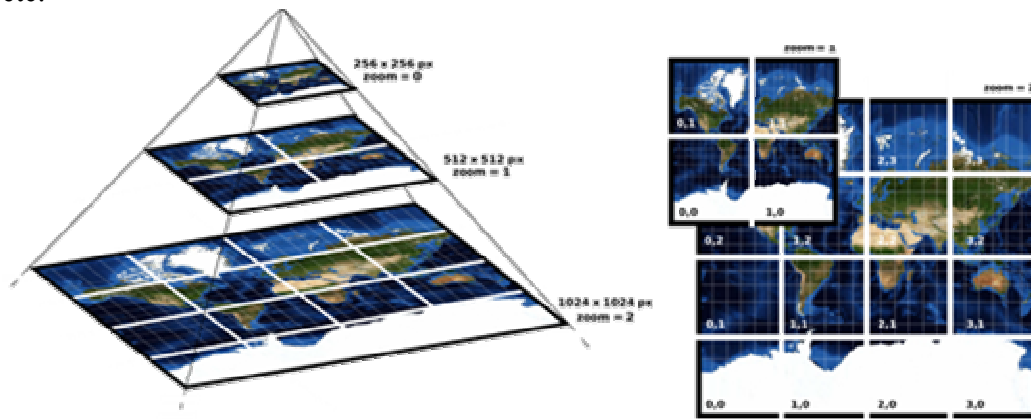
The tile pyramid as defined in TMS, using a global-Mercator profile.

With the WMS server installed and configured to provide raster maps, the easiest way of publishing map tiles in a form compatible with the above mentioned systems is to set up a tile caching system. Open-Source projects like TileCache (http://tilecache.org/) or GeoWebCache (http://geowebcache.org/) do exactly this.

Small institutions or individuals who do not have an online WMS server available but have georeferenced raster files with all geo-metadata can use the GDAL2Tiles application (also available as a part of the FWTools) to pre-generate the tiles[4] into separate files. After uploading these files on a web server, the map is published - without a need of extensive server configuration. GDAL2Tiles can also generate a simple webviewer for the tiles. If the input raster has the native projection of Google Earth (EPSG:4326, known as Plate Carré, or global geodetic http://www.spatialreference.org/ref/epsg/4326/) then the KML metadata for displaying tiles in 3D are generated by the GDAL2Tiles as well.

The principle of publishing tiles together with a simple web viewer is similar to the Microsoft Research MapCruncher application http://research.microsoft.com/ MapCruncher/ and is described in the article http://research.microsoft.com/ ~johndo/OSR2007-4b.pdf. With MapCruncher a user can register (by GCPs) and reproject an image. The program generates warped tiles so that the image can be placed on top of existing imagery of the Virtual Earth base maps. The functionality of GDAL2Tiles is similar, but it expects an already georeferenced image as an input. In contrast to MapCruncher, GDAL2Tiles does not have problems with very large images and can be used for batch processing. GDAL2Tiles was successfully tested on map raster files with size exceeding 30 gigabytes.

---

[4] Some of the functionality of the GDAL2Tiles utility is still in development. The final version is expected at the end of summer 2008. Right now support for the generation of Global Mercator profile tiles is missing. Reprojection of the raster has to be done in advance into a temporary file by gdalwarp utility.

The standard publishing process which is utilizing MapServer (as described above) is quite complicated and the setup of all the systems and their configuration is not an easy procedure. The raster data publisher (usually a library, an archive or a museum) would need a person with all the necessary IT and GIS knowledge. Also, many institutions do not have enough human resources to create all the geographic metadata, but they are scanning maps nevertheless and are willing to make them publicly available. Although part of the work can be done by an external subject, this might be prohibitively expensive for the map owner and it would make sense only for really valuable maps from the given collection. In cases where all maps are to be published, especially when the map collection numbers thousands of items, we need to simplify the process of publishing and geo-referencing.

The process of publishing of digital copy of a historical document should be as simple as possible and with the tiling approach it really is (a simple step-by-step tutorial for this process is available on the project website - http://www.oldmapsonline.org/publish/). The data provider simply runs an application which processes the digital master image produced by the scanner (usually in TIFF format) and generates a directory with tiles and a simple web viewer. Publishing of the raster means just copying of the generated directory onto the institutional web server. Any existing web-server can be used, there is no need to setup anything new on the server side. The only requirement is the disk space needed for tiles on the webserver. Visitors to the website where the tiles are published can then very comfortably zoom and pan the digital copy of the document and look at the details directly from the window of the web browser. The standard web-viewer uses the OpenLayers project, a JavaScript/AJAX based OpenSource project with support for all standard versions of webrowsers and computer platforms.

OpenLayers based web-viewer generated automatically by GDAL2Tiles.

The most important fact is that the same published tiles can also be used in different viewers (e.g. based on Google Maps, or WorldKit.org). Any large raster image or digital

copy of all kinds of (historical) documents can be published this way. The tiles can even be used as a map overlay[5] in Google Earth, just by adding geo-metadata in the KML format.
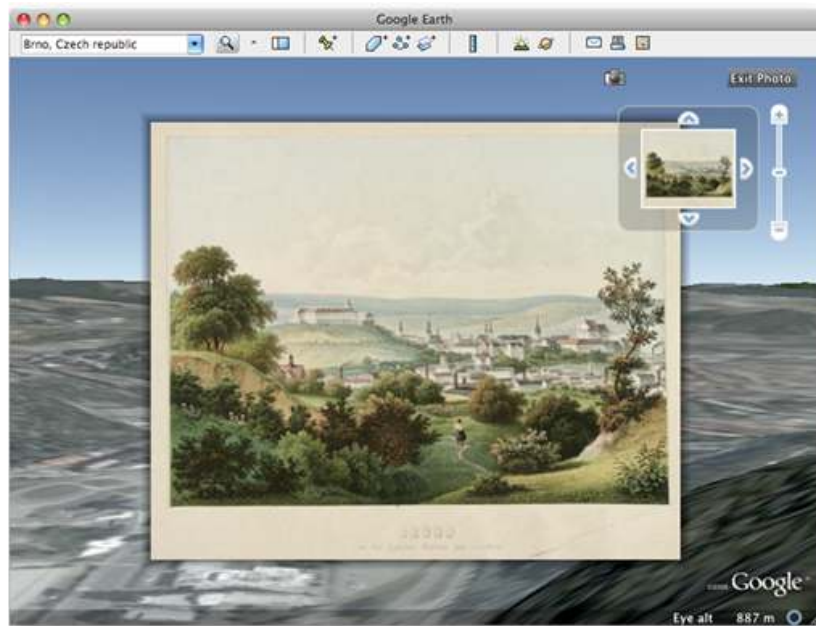
"Veduta" of the city of Brno in Google Earth.

If the document is not a map, but a "veduta" (a highly detailed painting of a cityscape or a landscape), a digitized historical photograph or a postcard, the tiles can be used to display it as a file in a 3D virtual world on the location where the given work originated. In Google Earth, this can be done with PhotoOverlay[6]. Visitors can then explore the details of the scan with zoom and pan functions even in the context of 3D environment. The key here is to use standardized or at least well-known formats of tiles.
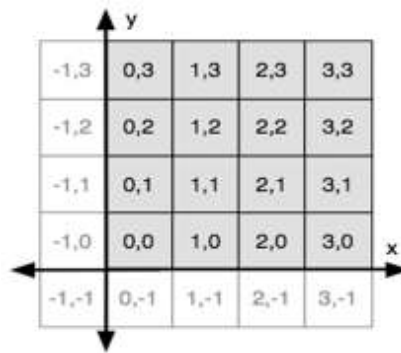
Tile numbering conventions of Tile Map Service (TMS). Tiles are available from the server in URLs such as http://server/.../map/zoom/x/y.png.

For the purposes of our project, we have chosen OSGeo specification TMS (Tile Map Service - http://wiki.osgeo.org/ wiki/Tile_Map_Service_Specification), a protocol defined for map tile exchange which is also used by tile cashing servers (tiles produced by Tile-Cache.org for example).

---

[5] For a high quality map overlay you have to first warp the raster data

[6] http://code.google.com/ apis/kml/documentation/photos.html

GDAL2Tiles, which has already been mentioned in this article is an application which generates such tiles from several raster image formats (TIFF/GeoTIFF, MrSID, ECW, JPEG2000, JPEG, PNG,...) and which supports very large raster files, up to hundreds of gigabytes in size. GDAL2Tiles version with a graphical user interface and installer for MS Windows will be soon available from our project website http://www.oldmapsonline.org/publish/.
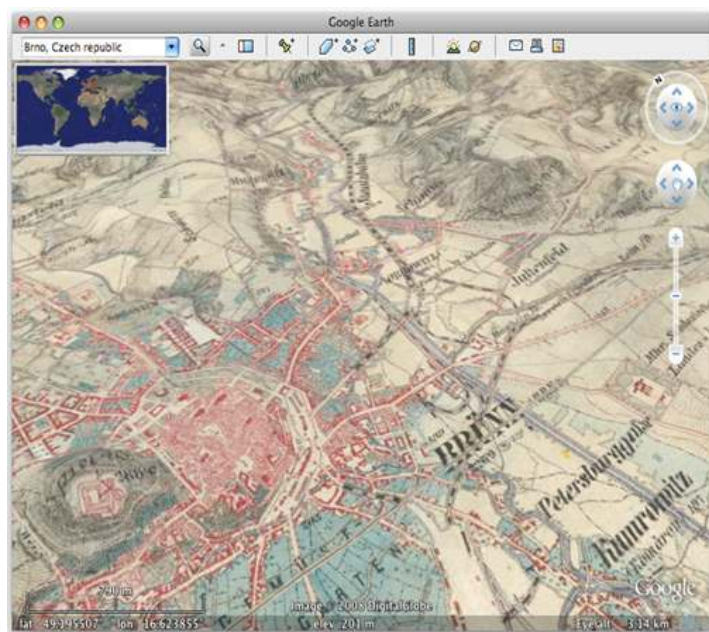
Third Military Survey of Austro-Hungarian Monarchy (1869-1887) in Google Earth 3D viewer.

Another tile format and viewer that anyone can use is Zoomify (http://www.zoomify.com/). Zoomify is based on proprietary Adobe Flash technology and allows to smoothly zoom and pan images in a web-viewer. There are several applications for generating the tiles on the Zoomify project website. As an alternative, one can also use open-source ZoomifyImage http://sf.net/projects/zoomifyimage/, which prepares the tiles and supports batch processing. Zoomify is currently quite popular for publishing large images on the web so we are supporting it in our workflow as an alternative way to TMS tiles.

In our suggested workflow a publisher always provide his digital documents from his own website - whether he chooses TMS tiles or Zoomify. The appearance of the online map presentation can be changed to match design of publisher's website. Both TMS (with OpenLayers viewer) and Zoomify use an attractive interface for presenting the document and both allow detailed online examination of the published digital document.

On-line georeference tool

Large map collections may contain thousands of maps and to specify the correct spatial reference and GCPs for each document is impossible in a reasonable time.

On the other hand, waiting for publication of the scanned maps until all the metadata are correct is often not acceptable. Why not specify geo-metadata by an on-line tool after the scanned map is published as an image? It is technically possible to create a web application for online map georeference (registration and also rectification). One prototype of

such application is MetaCarta Rectifier (http://labs.metacarta.com/rectifier/). Unfortunately, it is of limited use for our purposes because it does not support remote tiled image formats and large image files and its source code is not publicly available. Another similar tool is Yahoo MapMixer (http://maps.yahoo.com/mapmixer/).
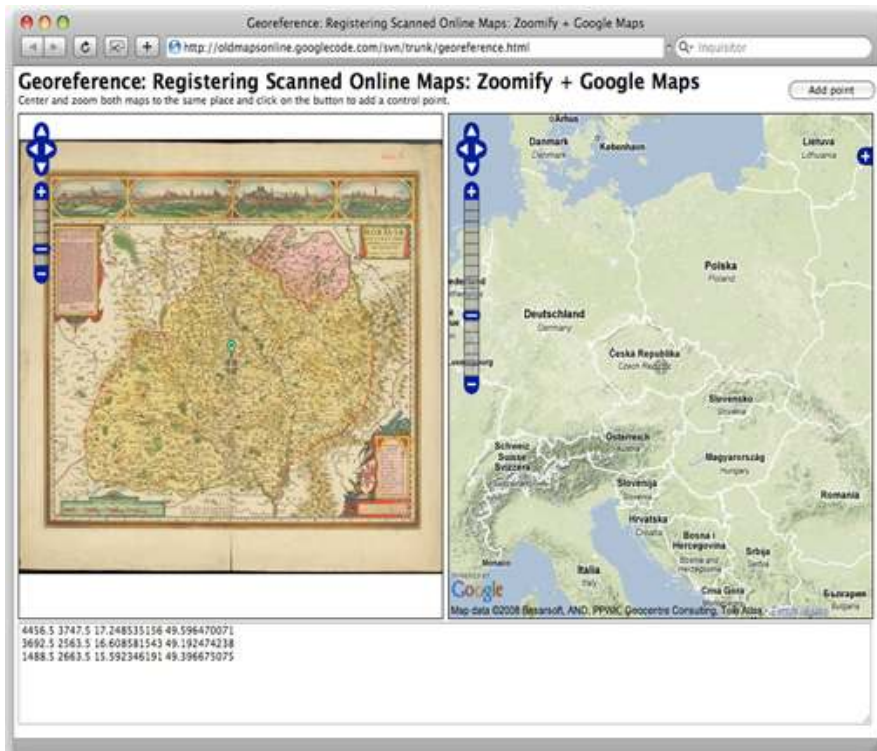
First GUI prototype of the online georeference tool. It allows registration of a scanned and already published map by specifying several pairs of Ground Control Points (GCPs).

The online georeference application should allow opening an old map published using tiles (TMS or Zoomify) from the remote server and it should also allow specifying the GCPs. The GCPs counterparts (points located on an existing map and used to learn the real coordinates of the GCPs) can be specified on a map loaded from a WMS server, or even on existing base maps of public map applications (Google Maps, Yahoo Maps, Microsoft Virtual Earth).

For the initial guess of location of the map we can use a web-service developed in the European project DigMap.eu http://www.digmap.eu/. This service positions an old map on the globe automatically by examining existing descriptive metadata from the libraries and utilizing a geographic thesaurus.

The user of the georeference application should also be able to specify the exact coordinate system information for the old map, probably from the on-line SpatialReference.org database or directly using the exact definition in the form of EPSG code, WKT or PROJ.4.

The main idea of this approach is that the quality of georeference for a map could improve over time, as more precise geo-metadata are available. In the beginning we can specify just a couple of points based on the Google Maps base map, later on we can add coordinate system information (ellipsoid, coordinate system and projection) and more

precise ground points based for example on a high quality WMS service provided by a government agency.

OpenLayers project provides the perfect technical base for such a tool on the client side. The GDAL library and tools like gdalwarp will perform necessary raster processing on the server side. Coordinates for GCPs must be saved in a database on a server with all the necessary information about their origin, including who specified them, when and against what base map and zoom level they were selected.
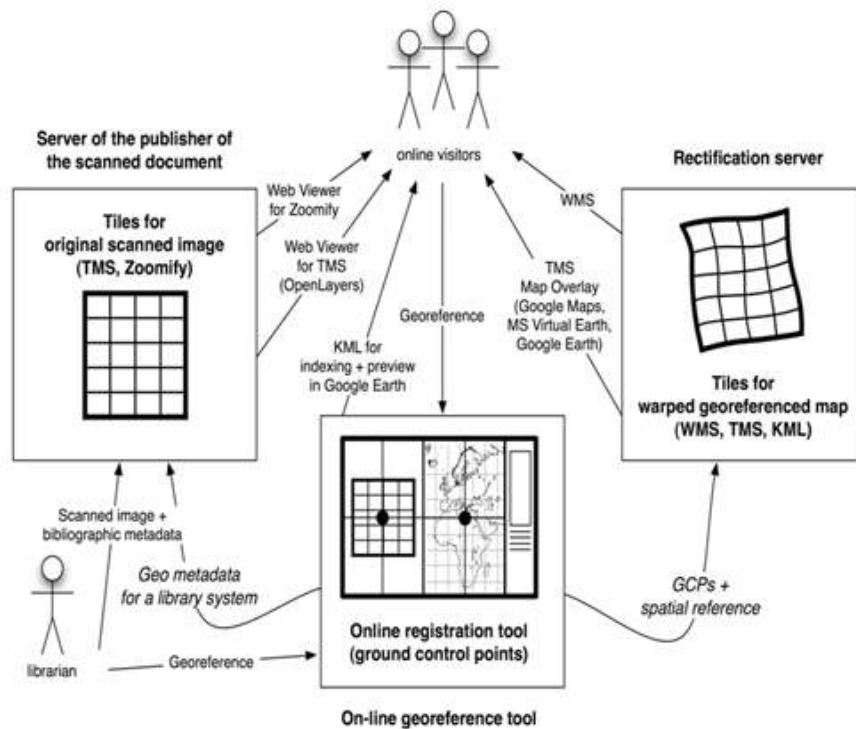
Scheme of the map publishing, online rectification and a server with WMS/TMS/KML.

The key functionality of the system is to do all important processing of scanned maps online, possibly on a different server located anywhere in the world. Therefore, memory institutions do not need to buy or set up specialized GIS software, yet their maps could be available in an interchangeable tile formats as well as trough WMS protocol. They only need to publish their scanned maps as Zoomify or TMS tiles.

When the rectification server performs the warping of remotely published raster tiles it can provide a map in a form of:

- *OpenGIS® Web Map Service (WMS)* - for standardized interchange with other maps in the GIS world.
- *Tiles with a Global Mercator projection (TMS, WMS-C)* for presentation of maps as an overlay on top of popular on-line map services a la Google Maps.

Even without warping of scanned images the registration points can be used for generation of important geo-metadata:

- *World file* for scanned maps for usage in GIS software.
- *Keyhole Markup Language (KML) Metadata* - for non-warped map presentation in 3D using Google Earth and soon probably also with other programs, as KML is now an OpenGIS standard.

[19]

- *SiteMap.xml with records for all maps*. Documents then appear in the search results of fulltext services which are supporting this protocol (like search functionality in Google Maps or Google Earth).

When the online georeference tool is used on published scanned maps, it will automatically provide all necessary geo-metadata for any geographic information retrieval (GIR) system. Generated metadata will be very useful for geo-temporal search of maps: the user can select an area of interest on a map, choose a time interval on a timeline and the search system will return the list of relevant documents or maps. The idea of such a search engine is the cornerstone of Alexandria Digital Library project (http://www.alexandria.ucsb.edu/) or of Electronic Cultural Atlas Initiative (http://www.ecai.org/) with the TimeMap system (http://www.timemap.net/).

Conclusion

The online georeference tool and rectification server described above will be developed and the results will be published as an open-source project. Results of the development will appear on the http://www.oldmapsonline.org/ website . The prototypes will run on the same website as well. We are coordinating our work with other grants and projects, the closest one of which is led by Jan Hartmann from the University of Amsterdam (http://mapserver.sara.nl/papers/2007/lisbon_digmap/).

This paper presents efficient procedures that should lead to online publication of scanned maps and shows the problems facing memory institutions with large collections and limited resources. A solution to these problems has been proposed through the use of open source tools on whose development we are currently working. We recommended free software tools for online publishing of scanned historical documents, which can be directly used by anybody who is willing to publish these documents.

References

Documentation for mentioned Open Source projects is available on-line. The list with short description of several free GIS projects is at http://www.maptools.org/.

Microsoft Virtual Earth Tile System; Virtual Earth Developer Center; http://msdn.microsoft.com/en-us/library/bb259689.aspx

Google Maps API Documetation; Map Overlays: Custom Map Types; http://code.google.com/apis/maps/documentation/overlays.html#Custom_Map_Types

OpenGIS(R) Web Map Service (WMS) Implementation Specification, Version 1.3.0, 2006-Mar-15, 06-042; http://portal.opengeospatial.org/files/?artifact_id=14416 and http://en.wikipedia.org/wiki/Web_Map_Service

Elson, J. (et al) (2007), MapCruncher: Integrating the World's Geographic Information; Microsoft Research; SIGOPS Operating Systems Review 41 (2); http://research.microsoft.com/~johndo/OSR2007-4b.pdf

Jenny, B, (2006), MapAnalyst - A digital tool for the analysis of the planimetric accuracy of historical maps; e-Perimetron, Vol.1, No. 3, ISSN:1790-3769; http://www.e-perimetron.org/Vol_1_3/Jenny.pdf

Smith, T. R. et al. (2001), The Alexandria Digital Library Project; Global Digital Library Development in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration (Ching-chih Chen, editor) (Beijing, China: Tsinghua University Press, 2001): 287-298; ISBN 7-3020-4473-2; http://www.alexandria.ucsb.edu/~gjanee/archive/2001/china-paper.pdf

Janée, G. et al. (2004), Issues in Georeferenced Digital Libraries; Alexandria Digital Library; D-Lib Magazine, Volume 10 Number 5, ISSN 1082-9873; http://www.dlib.org/dlib/may04/janee/05janee.html

Johnson, I. (2005), Indexing and delivering Historical Maps Online Using TimeMap; National Library of Australia Magazine, Jan 2005; http://www.timemap.net/tm/documents/publications/2004_08_nla_magazine.pdf

Hartmann, J. (2007), First DigMap WorkShop, Lisbon, Portugal; http://mapserver.sara.nl/papers/2007/lisbon_digmap/