CONFERENCE PAPER

ONTARIO'S HISTORICAL TOPOGRAPHIC MAP DIGITIZATION PROJECT

Colleen Beard, Brock University
Jay Brodeur, McMaster University
Sharon Janzen, Brock University
Amber Leahey, Scholars Portal, OCUL
Sarah Simpkin, University of Ottawa

INTRODUCTION

This conference report summarizes the Ontario Council of University Libraries (OCUL) Geo Community's project to digitize early 20th century historical topographic maps. The presentation was made to the CARTO Conference 2015, held in June in Ottawa, Canada.

The project, distributed across several Ontario university libraries, will add approximately 800 maps to our collective digital holdings, including ~615 map sheets from the 1:63,360 national topographic map series, all of which are in the public domain, and ~177 map sheets from the 1:25,000 national topographic map series, representing the subset that is currently in the public domain (a total of approximately 621 Ontario sheets exist in this series). Topographic maps at these scales are heavily used by researchers interested in examining changes over time (urban sprawl, transportation patterns, diminishing woodlots, shoreline erosion, etc.)

Access to the older series is uneven across institutions, but our ability to leverage the OCUL Scholars GeoPortal platform (http://geo.scholarsportal.info) and existing equipment at our institutions will allow us to share our digitized and georeferenced maps with the public at large. Funding from OCUL ($32,000 - January 2015 to April 2017) is allowing us to cover student staffing costs and those associated with wear and tear on existing equipment. This distributed project takes advantage of online collaboration tools such as Google Sheets, which allows us to manage a master list of the known maps across institutions, as well as the current status of each item. Institutions will be able to contribute by scanning maps, adding their holdings to the inventory, providing georeferencing support, and creating metadata for the records. Overall, our goal is to create and provide access to a high quality, consistent digital collection that preserves historical topographic information and meets the needs of current and future users. It is our understanding that a national strategy for preservation of these maps is required and we hope that this will assist with continued efforts in this area.

WORKFLOWS, SPECIFICATIONS AND STANDARDS

The current project workflow and its associated specifications and standards reflect efforts to achieve the aforementioned goals in a setting where groups from differing institutions will be working collaboratively on various processing steps. The project workflow (Figure 1) outlines the processes required to develop high quality digital products for end-users; these include steps to digitize, describe, georeference and transform the material, as well as those to coordinate processes and provide quality control and assurance.

Given the distributed and collaborative nature of this project, it is critically important that clear standards and common processes be developed to ensure that products are consistent and of a high quality. As a part of this work, a thorough scanner comparison and georeferencing investigation was undertaken, with the aim of identifying and characterizing sources of variation and errors, and subsequently develop standards, workflows and QA procedures to mitigate such issues.

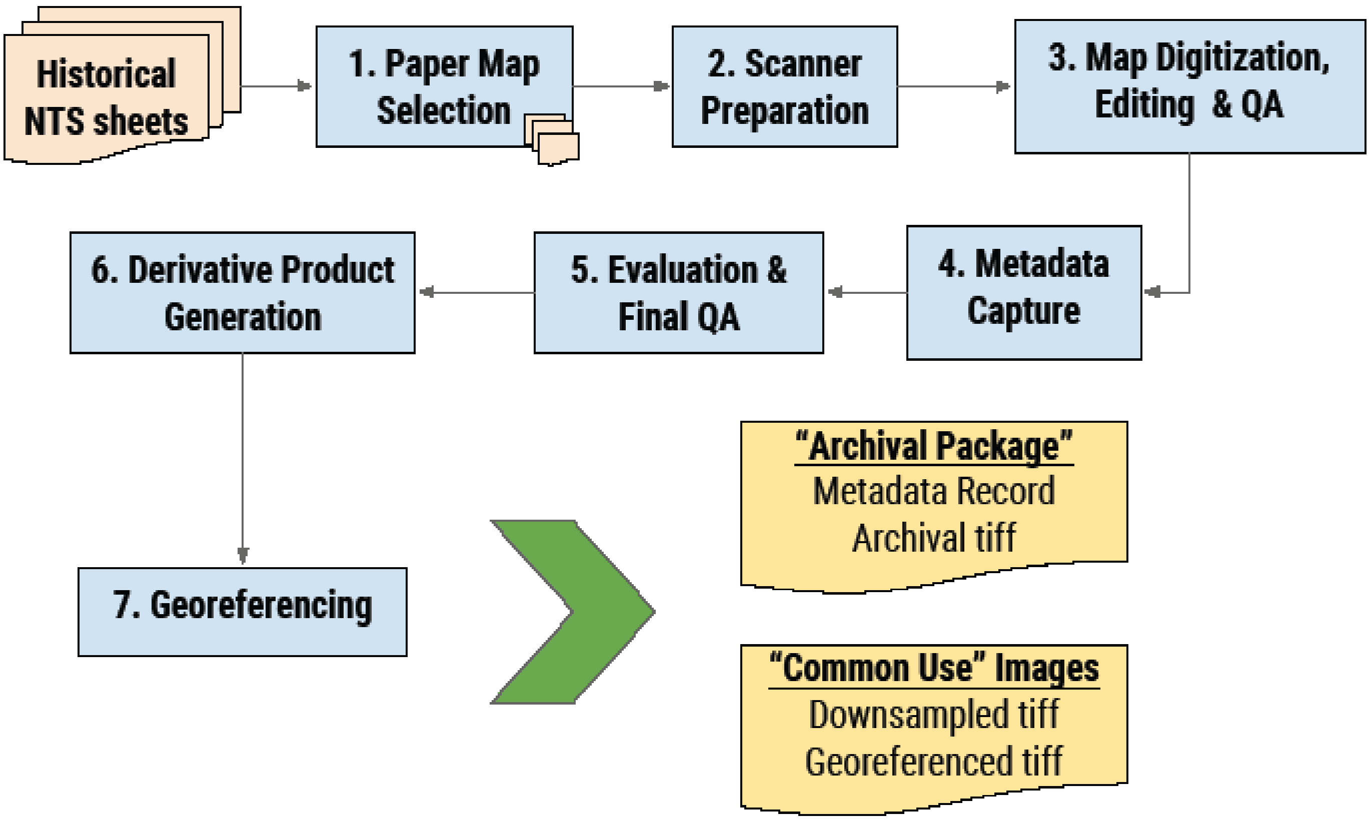

Figure 1. The Project Workflow

In the scanner comparison test, five institutions provided a digitized version of the same map sheet (identical sheet number and year of publication; different physical copy) for observation. The submitted images were digitized using a variety of sheet-fed and overhead photography scanners, collected at resolutions varying between 300 and 600 points per inch (ppi). Results of the comparison (https://goo.gl/f7R1zu) revealed substantial colour variation between the scanned images, which were attributed to one or a combination of the following influences: a) variation in the condition of the maps; b) inherent differences in the methods' colour output; and, c) post-scan filters that were applied by the scanning software. Additionally, considerable differences in image clarity at high zoom were apparent (https://goo.gl/laOhWF), which reflected the varying scanning resolution.

In order to minimize the sources of variation in scanned images, a number of recommendations were developed:

- Good quality paper maps should be selected for digitization, wherever possible;
- Colour calibration and QA procedures will be standardized, using a common calibration target and well-defined processes;
- Maps will be digitized at 600 ppi resolution, with 24-bit colour depth;

The purpose of the georeferencing tests was to establish appropriate input requirements for georeferencing by balancing the need for accuracy with the associated time commitments and staff resources at the libraries. In this exploration, the number of ground control points (GCPs) and the transformation model type (1st, 2nd, 3rd order polynomial) were varied to understand their effect on accuracy for a number of georeferenced maps. Results showed that a second order polynomial provided considerable benefits over a first-order model, and that generally, 8 to 12 GCPs were required to maximize accuracy (https://goo.gl/Fta7ly). When time requirements were considered, results showed that this approach would require approximately 15 minutes of work per sheet (https://goo.gl/192ZA2) -- an acceptable commitment given the resources that are available to the project.

Moving forward, it is envisioned that collaborators at different institutions will participate by downloading the high quality digitized maps from servers at Scholars GeoPortal, and then building and uploading the GCP file for each image. This approach minimizes data transfer requirements, and allows for transformations to be carried out (and re-applied, if necessary) at a single location using a "batch" process.

METADATA CONSIDERATIONS

As we embark on this project to digitize historical topographic paper maps held in our collections, we are providing standard metadata to describe and preserve the inherent descriptive and historical information about the paper maps, as well as describe the processes and steps taken to transform and digitize these resources for use in GIS. Describing the data so that users have a greater sense of the provenance, that is - where these data come from and how they were created, is important for our libraries. Additionally, we are especially interested in improving access to these data online, and the creation of standards-based granular metadata enables that.

Today, the most common standard for the description of geospatial data is the ISO - 19115 standard for geographic information. It is heavily used by government and data producers in Canada and internationally, with the Government of Canada formally adopting the standard in 2012.

The OCUL Scholars GeoPortal is a project that aims to bring together all digital data holdings into a shared portal. A shared metadata repository and metadata editor enables the description of data for the collection, using the ISO - 19115 North American Profile (NAP) metadata standard. Today, this NAP editor and metadata repository host over 2000 metadata records consisting of over 1000 shared metadata records, and over 1000 metadata records from the University of Toronto's local Map and Data Library (MDL) collections. To facilitate the integration of metadata in various formats from data producers and other collections over time, the use of metadata crosswalks including FGDC to ISO 19115, and a local crosswalk, have been developed.

Currently, the digitization project members are collecting metadata in a shared google spreadsheet and these data will be the basis for a metadata migration project. We will be developing the metadata crosswalk and determining best practises for the description of these data throughout the project lifecycle, utilizing current infrastructure and practises that have been established by the community to date. With the creation of rich, descriptive, and granular metadata we will hopefully improve access to these valuable library resources.

ACCESS (…and then they were seamless – in quest of perfecting discovery!)

Although various tools are being explored for the project, we presented the options for georeferencing and image analysis using ArcMap.

Setting the ArcMap workspace for georeferencing was presented as being an important first step to any georeferencing project. This includes setting the datum and coordinate system to reflect the original map; selection of an accurate control; using a high quality 600 ppi raster image; and utilizing the Image Viewer to view raster and control layers in separate windows.

The ideal number of control points required to georeference the historical topographic maps is dependent on the transformation that is desired. Concepts about residual error and how these relate to the different transformation types were discussed. It was suggested that at least the four known coordinates at the map corners be used. Although the spline transformation requiring 10 control points seems to be a slightly better fit to the Earth's surface, it is difficult to justify the added effort for overall modest difference.

The rectification specifications for creating a GeoTiff were also outlined. ArcMap provides several options for adjusting or enhancing the image quality. Achieving the best results concurred with georeferencing test findings (see Standards), by applying: the cubic convolution resample type, which sharpens the image; jpg compression at 100, which reduces the image file size up to ¼ of the original but at the expense of some quality. It is recommended that reducing the file size for downloading purposes is best achieved using the resampling tool offered in the suite of ArcMap tools. As well, applying enhancements to image quality can also be achieved using the Image Analysis options within ArcMap.

Creating an image display of georeferenced maps is one of the ultimate objectives in any digitization project with serial parts. The criteria to be considered in such a display is a platform; seamless display; but providing the map in its entirety (i.e. including the map collar); providing map overlay with transparency option; and image download.

Two website examples were discussed that display the U.S. topo series with varying options. The USGS Historical Topographic Map Explorer (http://historicalmaps.arcgis.com/usgs) designed by Esri provides a seamless map overlay with transparency. However, image download (with collar) is offered as geopdf only – not ideal for reuse in a GIS platform. Whereas the USGS TopoView (http://ngmdb.usgs.gov/maps/TopoView/viewer) offers multiple file formats for download (jpeg, kmz, geopdf, geotiff) and enhanced search options, but no seamless map display. A portal using a combination of both these sites is ideal.

Brock has experimented with image display options using ArcMap, achieving some success. Georeferenced topographic maps are stored in a mosaic dataset - a data model within the geodatabase used to manage a collection of raster images. In brief, the steps to creating a seamless display in ArcMap involve creating footprints – a manual method for virtually cropping the map to its neatline. This method preserves the map in its entirety (with collar) but displays only the content defined by the footprint. The new map document, which appears seamless, can be shared in ArcGIS Online (AGOL) as an image service – a publishing process that takes place within the desktop platform. Although this process consumes Esri organizational credits for caching and storage, the cost is minimal. Global sharing and convenient access via ArcMap desktop are key advantages of AGOL (Brock has used this procedure for many of its digitized map and air photo collections).

Brock has also added seamless displays to its local data listing, including a download option for individual geotiff map images (http://www.brocku.ca/maplibrary/digital/Niagara-NTS-63k.php).

CONCLUSION

The research and standards presented here highlight the ongoing considerations of OCUL libraries in undertaking a large map digitization project.

Get in touch!

Sarah Simpkin (sarah.simpkin@uottawa.ca), Jay Brodeur (brodeujj@mcmaster.ca), Amber Leahey (amber.leahey@utoronto.ca), Colleen Beard (cbeard@brocku.ca), Sharon Janzen (sjanzen@brocku.ca)

Special thanks to the OCUL Geo Community especially Cheryl Woods and Eva Dodsworth, who are actively involved in this project, as well as McMaster University summer students Victoria Balkwill Tweedie and Katie Maloney for their hard work and assistance.