

Winter 2016

Historical map digitization in libraries: Collaborative approaches for large map series

Cheryl Woods

Western University, cawoods@uwo.ca

Sarah Simpkin

University of Ottawa, sarah.simpkin@uottawa.ca

Jason Brodeur

McMaster University, brodeujj@mcmaster.ca

Amber Leahey

Scholars Portal, amber.leahey@utoronto.ca

Colleen Beard

Brock University, cbeard@brocku.ca

See next page for additional authors

Follow this and additional works at: <https://ir.lib.uwo.ca/wlpub>

 Part of the [Library and Information Science Commons](#)

Citation of this paper:

Woods, Cheryl; Simpkin, Sarah; Brodeur, Jason; Leahey, Amber; Beard, Colleen; and Janzen, Sharon, "Historical map digitization in libraries: Collaborative approaches for large map series" (2016). *Western Libraries Publications*. 48.

<https://ir.lib.uwo.ca/wlpub/48>

Authors

Cheryl Woods, Sarah Simpkin, Jason Brodeur, Amber Leahey, Colleen Beard, and Sharon Janzen

Historical map digitization in libraries: Collaborative approaches for large map series

Sarah Simpkin, University of Ottawa (sarah.simpkin@uottawa.ca)

Jason Brodeur, McMaster University (brodeujj@mcmaster.ca)

Cheryl Woods, Western University (cawoods@uwo.ca)

Amber Leahey, Scholars Portal (amber.leahey@utoronto.ca)

Colleen Beard (cbeard@brocku.ca) & Sharon Janzen (sjanzen@brocku.ca), Brock University

1 INTRODUCTION

Academic libraries are playing a role in the digitization of Canadian government documents^{1,2,3}, but maps tend to be excluded from these activities due to their unique dimensions and display requirements. Using a topographic map digitization project as a case study, this paper presents a collaborative approach to map scanning, georeferencing, and metadata creation across several Ontario universities.

Collectively, the 21 institutions making up the Ontario Council of University Libraries (OCUL) possess and maintain large volumes of Canadian topographic maps. However, few OCUL universities hold complete sets of these map series. While the Canadian government's most recent topographic maps are now available online, older editions of these maps have not been digitized. This project, currently underway at several participating universities, will enable us to share digital versions of some of our most-requested historical map series with the public at large.

Topographic maps are commonly used by researchers interested in examining changes over time (urban sprawl, transportation patterns, diminishing woodlots, shoreline erosion, etc.) and we believe that digitizing, georeferencing, and publishing the maps online will augment their use in teaching, research, and public use applications. In addition, since many of the maps were published prior to 1966, a majority are considered to be in the public domain (meaning that the copyright term of protection has expired) and they may be reproduced without permission; hence, they are shareable with the wider public. Providing online access to these historical map collections will be a valuable addition to the Canadian historical GIS resources that are currently available on the web.

¹ Accessed at <https://archive.org/details/governmentpublications>

² Accessed at <https://agiig.wordpress.com/noteable-digitization-projects>

³ Accessed at <http://govinforegistry.blogspot.ca>

In this paper, we describe and discuss the collaborative approach that OCUL member institutions are currently undertaking to digitize all of Ontario's public-domain historical topographic maps at the 1:25,000 and 1:63,360 scales. Additionally, we document our process for establishing scanning and georeferencing guidelines for the project. It is our hope that this paper will serve as a useful reference for other institutions undertaking map digitization projects.

1.1 PROJECT ORIGINS

The OCUL Geo Community (formerly the OCUL Map Group) is a forum for the exchange of information and ideas pertaining to maps, geospatial data, and other cartographic resources, both print and digital, within the wider Ontario Council of University Libraries. For a number of years, the community has discussed priorities for map digitization (Trimble et al., 2015). In August 2012, community members identified the Canadian topographic series holdings within their collections, and after further discussion, two series emerged as a priority for digitization: 1:25,000 and 1:63,360. By spring 2014, members agreed to proceed with a formal budget proposal to OCUL for a three-year funded project (January 2015-April 2017). A draft proposal was completed in early September 2014 by the project managers (Cheryl Woods, Western University; Jason Brodeur, McMaster University; Sarah Simpkin, University of Ottawa) and community members were asked to provide feedback.

One of the top priorities for the group was to align the project's objectives with OCUL's commitment to enhancing information services in Ontario. Specifically, the project supports the following OCUL strategic plan goals:

- Ensuring maximum discoverability of digital library resources;
- Contributing to building world-class digital library services for Ontario students; and,
- Providing and preserving academic resources essential for teaching, learning and research.

The community's proposal was successful and the team received CAD \$32,000.00 to focus on digitizing and georeferencing topographic maps of Ontario that were not already digitized and publicly available. Two primary examples meet these criteria and are included in this project: the Ontario sheets in the 1:63,360 national topographic series (published between 1904 and 1949), and those in the 1:25,000 series that are greater than 50 years old (published between 1956 and 1967). Over 800 maps are included: 627 map sheets from the 1:63,360 national topographic map series and 233 map sheets from the 1:25,000 national topographic map series.

The OCUL Geo Community is well-positioned to leverage the expertise and equipment already available at member institutions. The project managers made an open call to community members, who volunteered to participate in various capacities, such as supplying maps from their collections, scanning, testing, creating metadata, and georeferencing.

Original maps are currently being scanned to specifications that are standardized for format and resolution (600 ppi). Georeferencing and metadata creation are then carried out to enhance the usability of the digital files, in accordance with a digitization plan that was developed for the project. The project also involves hiring student staff members at a number of OCUL schools, who are tasked with digitizing, georeferencing, and metadata creation under the supervision of community members at participating institutions. A majority of the project's funding supports the hiring of student employees, and supervisors contribute their time in-kind.

1.2 THE 1:25,000 AND 1:63,360 MAP SERIES

As mentioned above, the Government of Canada's 1:25,000 and 1:63,360 scale map series were prioritized for this project as they are both commonly requested from our users and had not been digitized by other parties. The 1:25,000 maps produced by the Army Survey Establishment (and then in 1966 by the Mapping and Charting Establishment) are the most detailed of any federally-produced series. These maps were originally intended for military use and were produced for military training areas (camps) during the First World War. In 1959, the Government of Canada's attention turned to the protection of Canadian cities in the event of an atomic attack. Canada's 17 largest cities were mapped immediately, with the resulting series becoming known as the military city plans. By 1970, there was enough public demand for a civilian version of these maps. A redesign of the military city plans was undertaken, and work on the NTS (National Topographic System) 1:25,000 series continued until 1978, when it was stopped for two reasons: first, demand for new maps of the Canadian Arctic shifted attention toward map production using the emerging 1:50,000 standard; second, some provincial mapping agencies had begun publishing their own map series at 1:10,000 and 1:20,000 scales (Nicholson & Sebert, 1981, p. 119).

L.M. Sebert, the former Head of the Mapping Programme, Topographical Survey Directorate, Surveys and Mapping Branch, Ottawa, stated that "[t]he 1:63,360 series and its successor the 1:50,000 are the most important series in Canadian mapping. The 1:50,000 scale is the largest scale at which large areas of Canada have been mapped, and it is the largest scale for which complete coverage of the country has been programmed" (Nicholson & Sebert, 1981). The 1:63,360 series, first produced by the Department of Militia and Defence (representing at a scale of one inch to one mile), began in 1904 and ended in 1949 when it was converted to the 1:50,000 series. According to Sebert (1976):

"The detail shown on these early sheets was remarkable. In addition to the differentiation of the construction materials for buildings (red for stone and brick), there was a similar differentiation of bridges into stone, iron and wooden construction. Rural industries were depicted by symbol and initials, and these included saw mills, grist mills and flour mills, factories, blacksmith shops, hotels and taverns, all being further defined as being of stone or wood construction. Woods were depicted by coniferous or deciduous tree symbols, and as these were drawn by hand the density of the woods was indicated, by the density of the symbols, into open, medium or close growth. Fenced roads were differentiated from unfenced; telephone lines were shown; telephone offices were identified."

This amount of detail for cultural features (Figure 1) was not maintained in the 1:50,000 series.

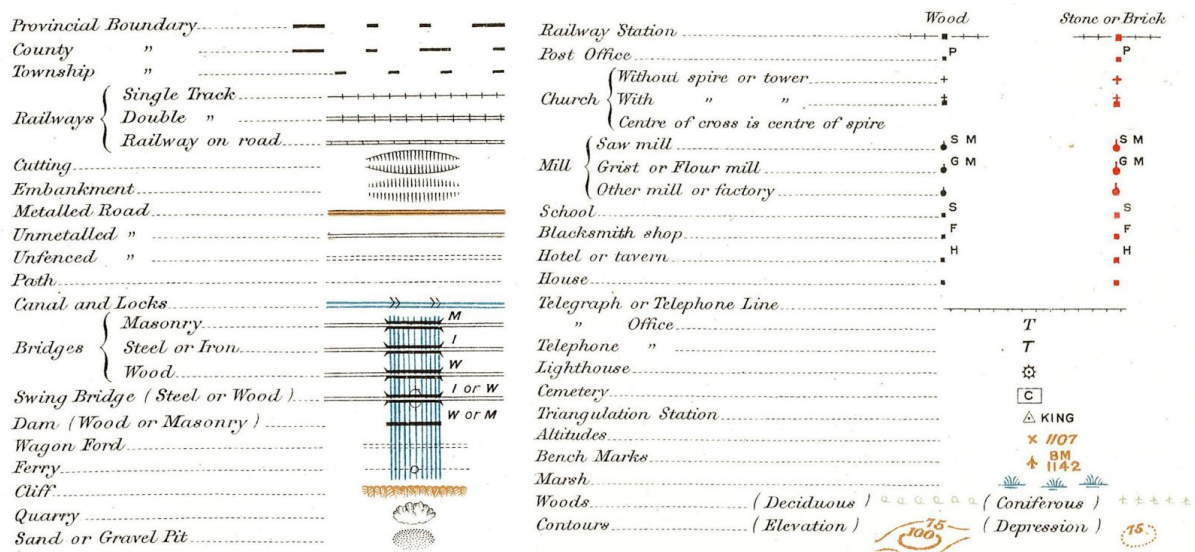


Figure 1: Cropped legend of Port Burwell sheet (040 110) from 1922 showing symbology details typical of the 1:63,360 series maps.

2 ORGANIZING AND ENABLING COLLABORATION AMONG DISTRIBUTED PARTNERS

With participating universities located across southern Ontario, the topographic map digitization project is by nature a distributed one and requires considerable coordination. Map collections, digitization equipment, and project staff are not typically located under the same roof. For these reasons, the project has relied on collaborative tools and shared storage for the scanned images, which at high resolutions can each require as many as 900 megabytes of storage.

2.1 INVENTORY CREATION

The team started by building an inventory of our collective holdings and the status of each map. Between April 2014 and August 2015, community members reported the maps that were held at their respective institutions, while project managers worked to identify any gaps. As the team could not locate a complete inventory of published maps in the two series, project managers assembled a working inventory of known maps from various sources, incorporating holdings lists from Canadian map libraries, the Archives of Ontario, Library and Archives Canada, Natural Resources Canada, and the Toronto Reference Library.

Projects of this nature typically rely on spreadsheets for managing inventories, updating statuses, and inputting metadata about each map sheet. To overcome the duplication and versioning challenges of sharing local copies of files across multiple institutions, the group chose to use Google Sheets (with local backup copies).

The team built and standardized the initial map inventory using one worksheet for each series and one row for each map and map edition. Maps that will be fully in the public domain by the end of the project were then identified. Individual maps are being tracked throughout the process, from “Have - not scanned” to “Scanned” to “Georeferenced”. Project leaders monitor the worksheets

to keep track of progress and send notices to individuals at participating libraries when it is time to send maps to be scanned.

2.2 WORKFLOW

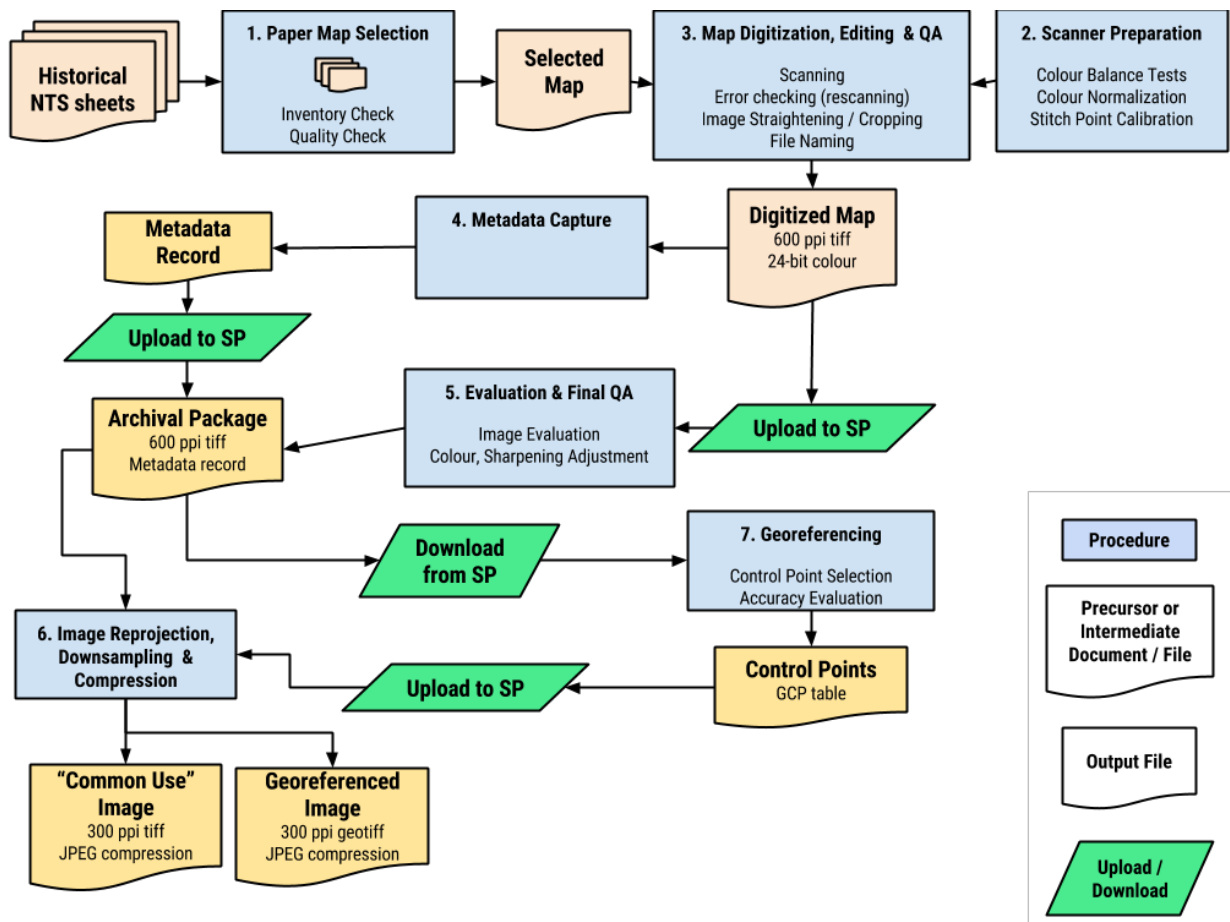


Figure 2: Project workflow schematic showing procedures, files, and file transfer steps.

The project workflow (Figure 2) begins with the map selection phase. Participating institutions are invited to record their holdings of 1:25,000 and 1:63,360 scale maps into a shared Google spreadsheet. Map sheets included in this master list should be of acceptable quality for digitization purposes. Notes about the condition of each map and other variables (for example, whether the map has any overprinting and whether or not it has been encapsulated) may be included in this phase. For this project, we determined that where multiple copies of a map exist, preference should be given to non-encapsulated maps to improve the quality of the scans.

Once maps have been selected, they move through the digitization, metadata capture, and georeferencing portions of the workflow, which are described in more detail in later sections of the paper. Our final goal, once quality checks have been completed, is to display the georeferenced maps on OCUL's Scholars GeoPortal platform, where they will be available for public viewing and downloading.

As of November 2015, 650 of 860 known maps have been scanned. The majority of the scanning is being completed at three (McMaster University, University of Waterloo, Western University) of the institutions because they have the same make and model of large format scanner, which

ensures consistency. One terabyte of FTP storage has been supplied by Scholars Portal, the service arm of OCUL, to store the files that will then be retrieved for georeferencing. We are currently investigating options for digitizing the remaining known maps in these series that are not held at OCUL schools. This may include contracting out the scanning work to organizations such as Library and Archives Canada, whose map collections cannot be loaned to other institutions.

3 STANDARDS FOR MAP DIGITIZATION AND GEOREFERENCING

The process of establishing standards for map digitization and georeferencing activities reflects the project's overarching goal to produce consistent, complete, and high-quality products, while also using a collaborative model that enables contributions from as many interested OCUL institutions as possible. Acknowledging that satisfying these objectives would require careful coordination and perhaps strategic compromises, we sought to develop standards by synthesizing documented general best practices with findings from a number of preliminary digitization and georeferencing tests.

Map digitization tests were conducted to assess the performance of various map digitization systems available at OCUL institutions, compare the quality of previously-digitized maps at others, and develop standards and procedures to guide project digitization activities. Georeferencing tests were conducted in order to establish methodologies and standards that would balance our need for accurate georeferenced outputs and time efficiency to keep within the project's resources.

3.1 MAP SCANNING: COMPARING SCANNER OUTPUT AND DEVELOPING STANDARDS

Stated generally, the process of digitization involves the conversion of information from analog to digital formats. In the case of cartographic material such as topographic map sheets, digitizing these resources allows their information to be shared more broadly, and enables new forms of analyses and knowledge dissemination. Due to the typically large dimensions of topographic map sheets, digitization requires specialized equipment; this requirement is usually met through the use of an overhead camera system or large-format sheetfed scanner.

In overhead camera systems, the large-format item is kept stationary on a copy stand, while an elevated camera captures an image from a fixed position, or a series of overlapping images from multiple positions above the item. In sheetfed scanner systems, the item is guided across a stationary camera array, which recursively samples narrow strips of the item. Each system type possesses advantages and disadvantages relative to the other, and the most appropriate equipment for a task is often determined by both the characteristics of the items to be digitized, and the requirements for the resulting images. For example, light levels can be more easily controlled by sheetfed scanners, but passing items through their rollers can be damaging to sensitive or fragile materials.

Regardless of method, comparable standards for map digitization have been developed by various organizations and project groups (Table 1). These published values suggest that map sheets should be digitized in TIFF format, using 24-bit colour depth, and at a minimum resolution of 300 points per inch (ppi), with a preference for higher resolution images where achievable.

Table 1: Image specification requirements for digitized maps, as published for comprehensive digitization projects.

Source	Required image specifications for digitized maps
Federal Agencies Digitization Initiative - Still Image Working Group (2016)	TIFF format; 400 ppi resolution (minimum); 24-bit RGB colour
Banach, Shelburne, Shepherd, & Rubenstein (2011)	TIFF format; 300 - 400 ppi resolution; 24-bit RGB colour
Dale, Leech, Bogus, Mathews, & Blood (2013)	TIFF format; 300 - 600 ppi resolution (minimum); 24-bit RGB colour
Allord, Fishburn, & Walter (2014)	TIFF format; 400 ppi resolution (minimum); 600 ppi resolution (recommended); 24-bit RGB colour

Prior to beginning map scanning activities, we conducted a comparison of scanned maps from a number of OCUL institutions that were identified as potential contributors to the digitization stage of the project, since they either had facilities available for map scanning, or possessed previously-digitized historical topographic map sheets as a result of contracted work from a commercial provider. The objectives of this comparison were to 1) evaluate variation in scan quality and appearance (e.g., sharpness, colouration, and consistency) across digitized products available from various OCUL institutions, and 2) use these findings to develop map scanning standards and procedures that would achieve the project goals of producing high quality products while enabling collaboration across partner institutions.

3.1.1 Methods

For the comparison test, groups from six institutions were asked to submit digitized images of a selected map sheet (sheet 30L/13, Dunnville, 1938); some of the images were created at the time of request, while others had been previously digitized. The submitted map images were generated using a variety of equipment and methods, as both sheetfed scanners and overhead camera systems were used to create images that varied in resolution between 300 and 600 ppi (Table 2). Given the variety of equipment being compared, each group was required to use equipment-specific methods for image colour and quality calibration. The submitted images were inspected at zoom levels ranging from full-extent to 1:1 scales, in order to assess their brightness, contrast, saturation, internal consistency, and sharpness, as well as to identify artifacts and errors introduced to the images as a result of the digitization process.

Table 2. Digitization equipment type and scanning resolution for each scanned map that was compared. ¹ Indicates equivalent equipment

Label	Digitization Equipment Type	Image Resolution (ppi)
--------------	------------------------------------	-------------------------------

A	feed-through scanner	300
B	overhead photography (fixed)	400
C	overhead photography (fixed)	300
D	feed-through scanner ¹	600
E	feed-through scanner ¹	600
F	feed-through scanner ¹	600

3.1.2 Results

A comparison of the map images obtained from the six participating groups (Figure 3) showed substantial colour variation between the digitized sheets in terms of their contrast, brightness, and colour saturation. Although some variation was attributed to condition of the source maps used at each institution, further tests and analyses revealed inherent differences in the colour characteristics of scanned images produced by the different methods. In the case of the photographed maps (Figure 3, panels B and C), such variation is likely to have resulted from the use of different camera equipment and settings, as well as differing ambient lighting conditions in the imaging environment.

Considerable colour differences were also observed across images that were collected using comparable feed-through scanning equipment (Figure 3, panels D, E, F). A small amount of the variation was attributed to differing scanner calibration coefficients, which arises from the equipment's lack of automated colour calibration procedures (equipment is normalized against a white target only). A more substantial contrast variation in the case of one of the images (Figure 3, panel E), was found to result from the application of post-scan contrast adjustment filters, which were applied latently within a 'map-specific' preset in the scanning software.

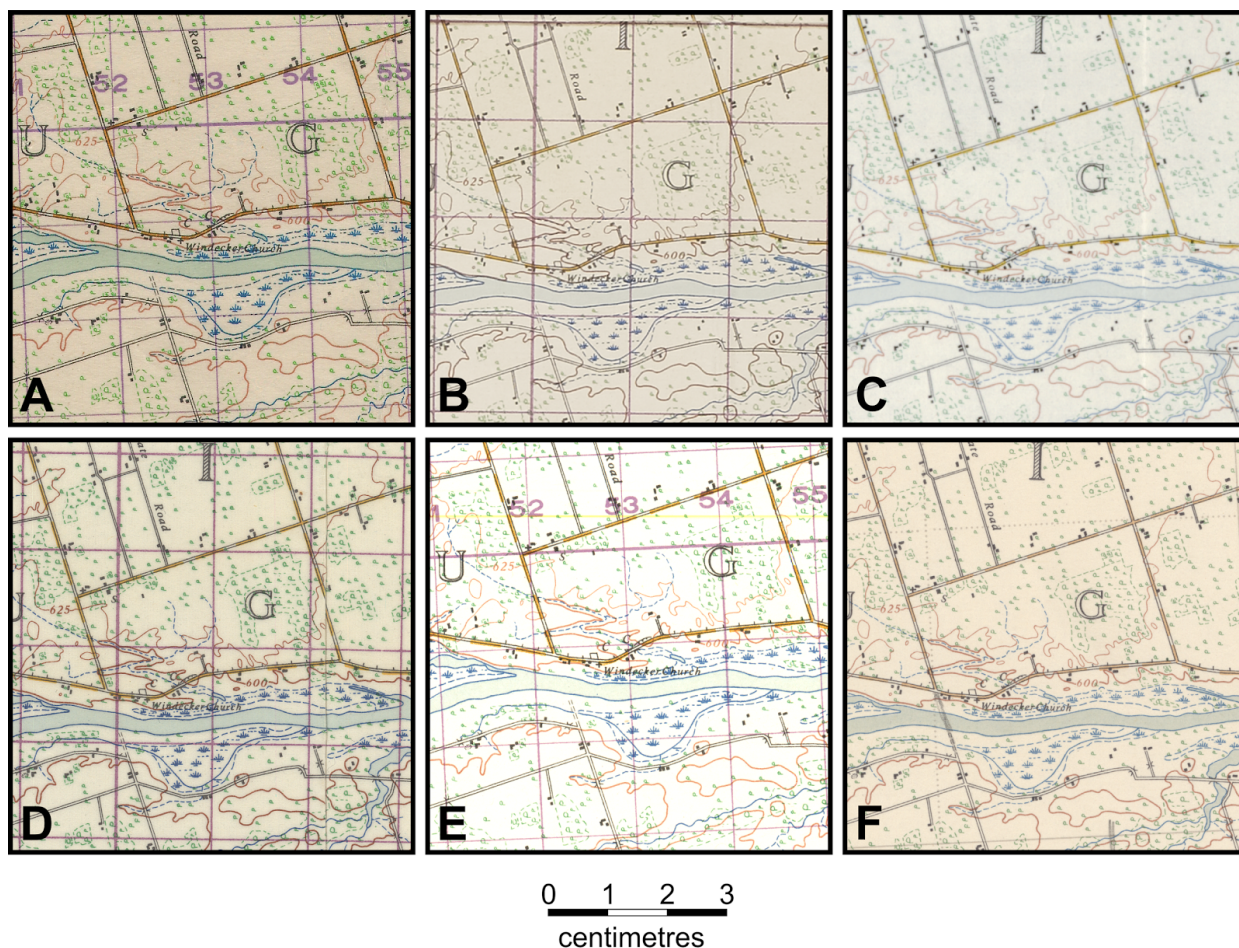


Figure 3. Colouration comparison of equivalent regions for digitized maps created at six different OCUL institutions. Panel labels correspond to information given in Table 2.

Investigating each digitized sheet at a larger scale (Figure 4) revealed considerable discrepancies in image sharpness. While much of this variability was explained by image resolution, the improving effect of post-scan image sharpening filters applied to images was also noted. While an image sharpening filter was not applied to the images shown in panels B and C of Figure 4 (400 and 300 ppi resolution, respectively), its use in the 300 ppi image in panel A demonstrates substantial sharpness improvement. This sharpness improvement, however, is achieved at the cost of increased image noise and a resulting ‘grainy’ appearance. Results of this comparison indicated that while high (600 ppi) resolution images (such as in panels D, E, and F of Figure 4) are most desirable for this project, it may be acceptable to use sharpened images from lower resolution (300 - 400 ppi) scans, in isolated cases where higher resolution products cannot be created. Such cases may arise when rare map sheets are too fragile to be used in a feed-through scanner, and instead must be photographed.

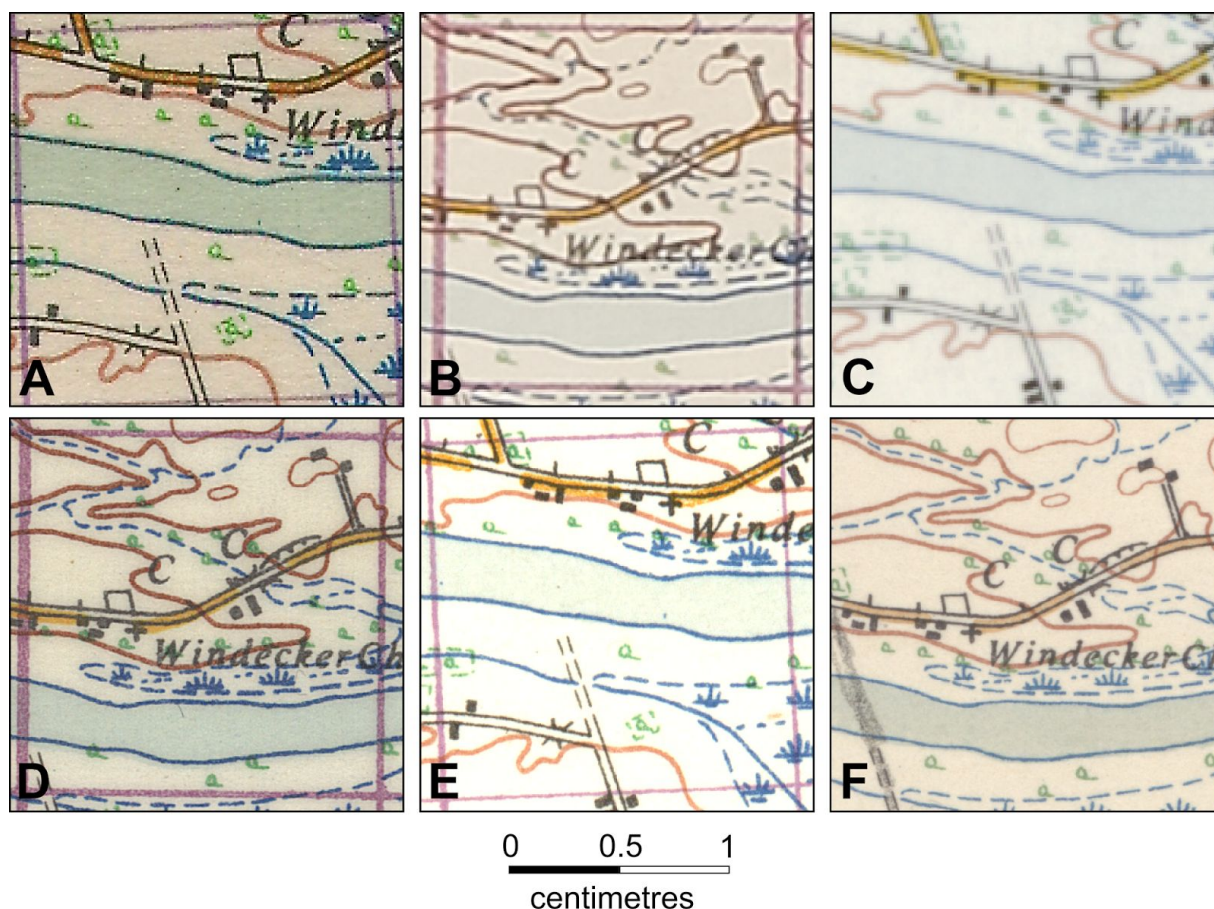


Figure 4. Sharpness comparison of equivalent areas for digitized maps produced at six different OCU institutions. Panel labels correspond to information given in Table 2. Image resolution is 300 ppi in panels A and C, 400 ppi in panel B, and 600 ppi in panels D, E, and F.

For some of the digitized sheets, a close investigation revealed the presence of artifacts and errors that were inserted into the images by the scanning equipment. Most notably, the images created using feed-through scanners commonly contained ‘stitching’ errors—image offsets that are produced as a result of mis- or poorly-aligned camera arrays on the scanning equipment (Figure 5). In most cases, calibration software and procedures can be applied to mitigate these offsets completely, or at least diminish them to the internal precision limits of the device. In other cases, physical irregularities of the map itself—such as creases and lamination ridges—create sporadic, localized artifacts, which may or may not be avoidable by careful rescanning of the material.

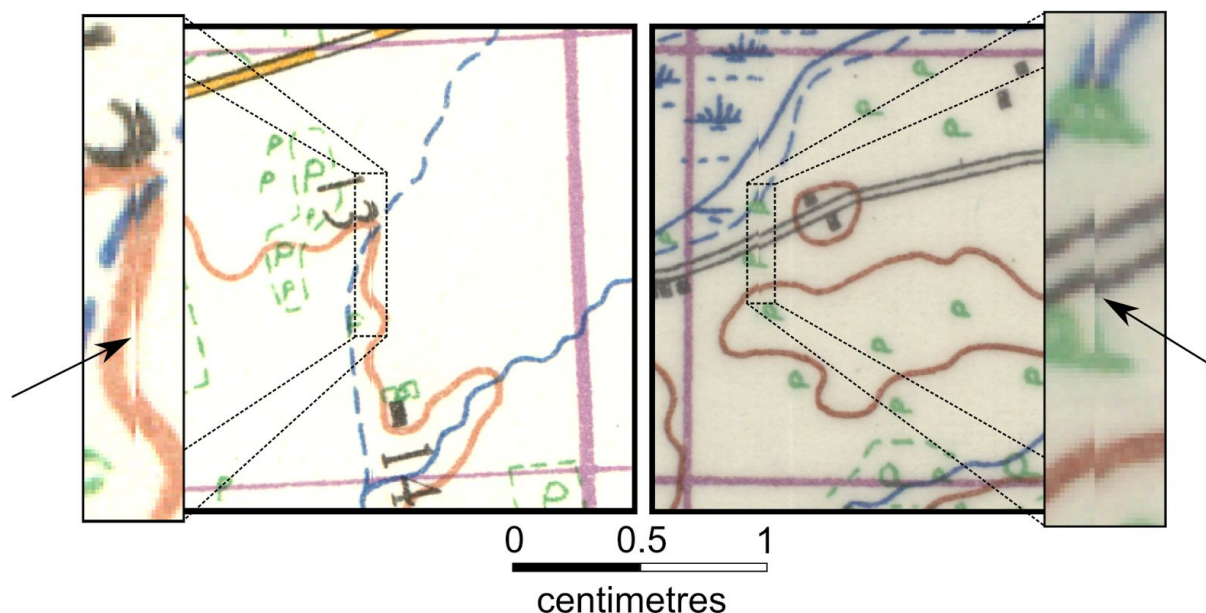


Figure 5. Stitching errors observed in map sheets that were digitized using feed-through scanners. A 'right-left' stitching error is illustrated in the left pane, while the right pane demonstrates a 'front-back' stitching error.

3.1.3 Recommended standards for map digitization

Project standards for map digitization were developed in consideration of the digitized image comparison results, as well as the project's stated objectives of producing high quality, consistent materials while enabling community-wide participation. Generally, the digitization standards set for this project meet or exceed those set for comparable historical map digitization projects (e.g., Allord, Fishburn, & Walter, 2014). Recommended activities and standards for digitization are as follows:

- In cases where a given map sheet is held by multiple institutions, every effort is made to digitize the sheet that is in the best physical condition—both in terms of its physical integrity and its appearance.
- Map sheets are digitized at 600 ppi resolution and 24-bit colour. Lower resolution (300 or 400 ppi) images may be accepted, in special situations.
- Standard colour calibration and quality control processes are implemented, including the use of a common colour calibration target to assess the colour characteristics of output from digitization equipment, and the implementation of standardized equipment settings where possible.
- Common procedures are used with feed-through scanners to minimize stitching errors through calibration, and to detect such errors during post-scan quality assurance methods.

3.2 MAP IMAGE GEOREFERENCING: DEVELOPING STANDARDS

Georeferencing is a procedure whereby phenomena and information are associated with geographic locations, as specified within a defined spatial reference system (Hill, 2009). In the context of digitized maps, this process involves associating pixels of the raster image with geographic coordinates (Hackeloer, Klasing, Krisp, & Meng, 2014). Georeferencing digitized geographic material, such as topographic maps, enables their information to be transformed into

any desired projection, and displayed alongside other georeferenced digital data, whether in a Geographic Information System or other analysis and dissemination software. Therefore, the production and provision of georeferenced digital collections builds upon digitization efforts to promote new and interesting uses of geospatial information for analysis, visualization, and dissemination (Knowles & Hillier, 2008).

In common practice, an image is georeferenced by an operator, who identifies in the image a number of ground control points (GCPs) where coordinates (in a desired reference system) are known to a reasonable precision (Hackeloeer, Klasing, Krisp, & Meng, 2014). When an appropriate number of GCPs have been inserted, a transformation model is then applied to the points in order to ‘warp’ the image to fit the projection of the target (desired) reference system. Depending on accuracy requirements and the nature of image warping that is necessary, a variety of transformation models may be used. Polynomial transformation models use a least-squares minimization procedure to fit GCPs to coordinates in the target reference system using polynomial expressions of varying order (e.g., first-, second-, third-order, etc.). Using such models, the image is warped globally to find a general best-fit for all GCPs, and the degree to which the image may be ‘warped’ increases with order. In comparison, spline and adjust transformation models warp the image locally to minimize error around all GCPs—a desirable approach when local accuracy is important or when spatial accuracy varies throughout an image.

In the case of digitized maps, the accuracy of the resulting georeferenced and transformed image (i.e., the alignment of features in the image with their true location on the earth) depends upon a number of factors, including:

- the accuracy with which the map sheet is digitized;
- the quality and quantity of inserted GCPs;
- the appropriateness of the transformation model used;
- the accuracy of information contained within the map; and
- the validity of reference data used in the georeferencing process.

Though georeferencing may be accomplished with a minimum of three orthogonal GCPs, the accuracy of the resulting image typically increases as more are inserted. Determining an appropriate number of GCPs requires compromising between the benefits of increased accuracy and the detriments of additional time and effort requirements. Therefore, the desired number of GCPs for georeferencing operations will depend upon the accuracy needed for the resulting georeferenced image, the type of transformation function that is used to reproject it, the ability of the operator to identify high-accuracy GCPs in the image, and the time, resources, and expertise available. Guidelines established by a similar map digitization project undertaken by the U.S. Geological Survey Historical Topographic Map Collection have suggested that 16 GCPs (and an absolute minimum of 7) should be used to georeference images of both 1:25,000 and 1:63,360 scale maps (Allord, Walter, Fishburn, & Shea, 2014).

Recently, a number of automated georeferencing methods have been developed, where image detection algorithms and additional contextual information are used to automatically register GCPs into the target raster. For example, QUAD-G software (Burt, White, & Allord, 2012; 2014) was used to automatically georeference maps in the U.S. Geological Survey Historical Topographic Map Collection (Allord, Walter, Fishburn, & Shea, 2014). While such an approach is promising for cases where digitized maps are highly standardized, the lack of such standardization in our

collection—particularly in early edition maps—precluded the use of this software for our purposes.

Alternatively, we sought to develop requirements, documentation, and standards that would allow georeferencing to be carried out by members of various OCUL institutions, in order to develop georeferenced products of an acceptable accuracy and quality. To do so, we conducted a series of georeferencing tests to better characterize the uncertainty and time requirements associated with georeferencing operations on our digitized topographic maps. The objectives of these tests were to:

1. Determine a reasonable expectation for georeferencing accuracy by establishing a corresponding tolerance level for error.
2. Identify the sources of georeferencing errors and develop methods for detecting and mitigating these errors.
3. Develop recommendations for the georeferencing process, which define the requirements for GCP quantity and distribution, as well the transformation model used in the reprojection process.

3.2.1 Methods

Tests were conducted using four digitized map sheets (1:63,360 scale; 600 ppi resolution), which covered disparate areas of the province and spanned a range of publication years. To improve the consistency of test results, sheets were selected for areas without significant bodies of water in their coverage, as they allowed GCPs to be spread evenly throughout the entire map area.

In each test, the first four GCPs were placed at the neatline corners, since the precise coordinates of these points were known for all sheets. Considering that many of the 1:63,360 scale maps lack a graticule that is usable for inserting GCPs, we decided to add additional GCPs exclusively through the identification of corresponding landmarks on the target map and modern reference geospatial data layers, which included a vector road layer (2014 Routefile, DMTI Spatial Inc.) and tiled raster base maps. Given the temporal discrepancy between the target and reference information, care was taken to select only landmarks (intersections, railway crossings, etc.) where confidence was high that the location was consistent between the periods. GCPs were manually added to the map image in a uniformly distributed pattern, in order to maximize the spacing between all points and the GCP coverage throughout the map (Figure 6). Preliminary tests indicated that a total number of 24 GCPs was sufficient to provide accurate error estimates for each map sheet.

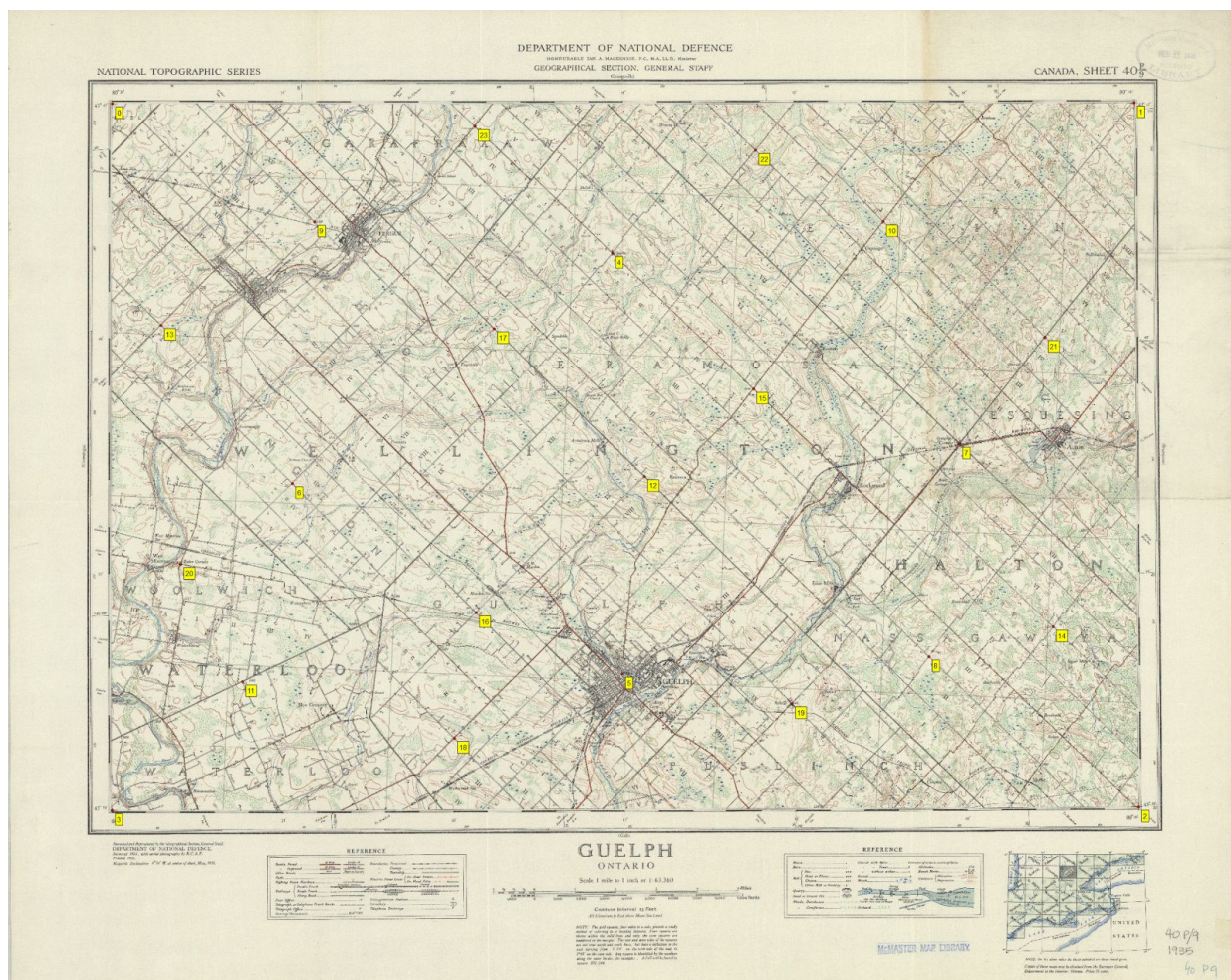


Figure 6. Distribution of 24 ground control points used for georeferencing error tests conducted on 1:63360 map sheet 40P/9 (1935).

All tests were carried out using the Georeferencer tool in QGIS (QGIS Development Team, 2015), which uses the Geospatial Data Abstraction Library (GDAL). The accuracy of each GCP was evaluated using the residual error reported by the Georeferencer tool, which measures the degree to which the location of a given GCP on the transformed image deviates from the original reference coordinates entered for this point during the manual georeferencing process. For a given test, total map georeferencing error was summarized and reported as the mean absolute error (MAE) for all GCPs used to parameterize the transformation model as:

$$MAE = \sqrt{\left[\frac{1}{N} \sum_{i=1}^N |\hat{x} - x| \right]^2 + \left[\frac{1}{N} \sum_{i=1}^N |\hat{y} - y| \right]^2},$$

where \hat{x} and \hat{y} denote the coordinates of the GCP in the *transformed* image, and x and y indicate coordinates of the original, reference GCP. MAE was calculated in units of image pixels, as well as metres on the Earth's surface. While root mean squared error (RMSE) may be used as an alternative measure of georeferencing error to MAE, internal tests showed MAE and RMSE to be very highly correlated ($r = 0.93$), indicating that the two measures were essentially equivalent for the purposes of our investigations.

We investigated the effects of GCP quantity and transformation model type on georeferencing error through the following procedure: All 24 GCPs were added to the original map image and a first-order polynomial transformation model was selected. Initially, only the GCPs at the four corners of the image were 'enabled', meaning that the transformation model was parameterized using just these four selected GCPs. All other 'disabled' GCPs were hidden from the transformation model when creating the 'best-fit'; however, their residuals were still measured within the software and were recorded to assess total MAE throughout the image. Following this, we systematically increased the number of 'enabled' GCPs and recorded residuals for all GCPs. The process described above was repeated using second- and third-order polynomial transformation models, though a greater initial number of minimum 'enabled' GCPs were required—6 for the second-order polynomial and 10 for the third-order.

Given that residual error will regularly be smaller for 'enabled' GCPs than for those that are disabled (since the transformation model is fit to only the 'enabled' points), we assessed georeferencing error for the entire map sheet using two different measures: a) MAE calculated using all 24 GCPs (enabled and disabled; denoted as MAE_A), and b) MAE calculated using only the disabled GCPs (MAE_D). Analyses of these two measures indicated that MAE_D provided a less-biased estimate of georeferencing error, which was more representative of areas in the images where 'enabled' GCPs were not present (see Figure 7, below).

3.2.2 Results and Recommendations for Georeferencing

Results from georeferencing tests showed that, generally for all polynomial transformation models tested, MAE_D decreased with increasing numbers of GCPs, until 12 GCPs were added (Figure 7). Beyond this point, adding more GCPs resulted in no further reduction in MAE_D and, in many cases, error increased slightly. While a rise in error beyond 12 GCPs was not anticipated, such an increase might occur from the selection of less ideal points as more GCPs are inserted—a consequence of georeferencing these images using reference data layers (and not graticule).

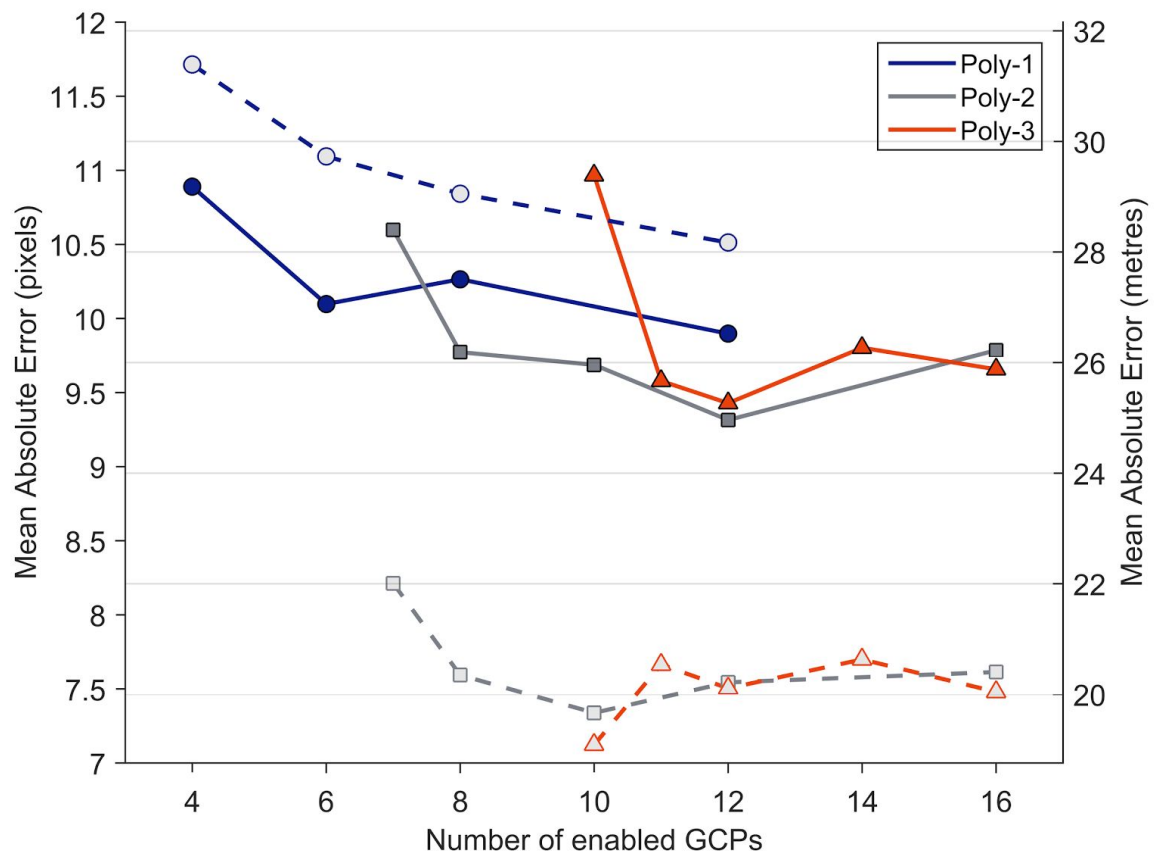


Figure 7. Averaged mean absolute error (MAE)—presented in units of map pixels and metres on the Earth’s surface—for varying quantities of enabled GCPs and orders of polynomial transformation models used. Dashed lines indicate values obtained when error from all GCPs was considered (MAE_A), while solid lines indicate values obtained when only error from non-enabled GCPs was considered (MAE_D).

Among the three polynomial models tested, the second-order polynomial consistently produced the lowest overall value of georeferencing error, though its improvement over the third-order polynomial was insignificant beyond 10 GCPs. Given that the 1:63,360 map sheets use a polyconic projection (with a central meridian at the centre of the quadrangle), the increased degree of freedom offered by the second-order polynomial transformation demonstrated clear advantages to the first-order polynomial. While generalized results suggested that error was effectively minimized with the insertion of 12 GCPs, examination of errors for individual digitized sheets (not shown) indicated that the specific number of GCPs required to achieve this varied across sheets between 8 and 12 GCPs. The resulting recommendation of these tests (to insert 8 to 12 GCPs and use a second-order polynomial) is consistent with findings and recommendations from the U.S. Geological Survey Historical Topographic Map Collection digitization project (Allord, Walter, Fishburn, & Shea, 2014), where 7 to 16 GCPs are inserted and a second-order polynomial used.

Given that georeferencing for this project will be carried out manually, the amount of time required to complete this process for the entire collection was important to consider prior to issuing georeferencing standards for the project. For this purpose, the time to complete each GCP was recorded throughout the entirety of the tests. Evaluation showed that the time required to

georeference a map image increased linearly with the number of GCPs added, at a rate of approximately 110 seconds per GCP. Assuming that an average of 10 GCPs would be placed per map image (as per requirements to minimize error), we estimated that the entire digitized collection could be appropriately georeferenced in around 250 hours—a value that is manageable within the project’s budget.

4 CATALOGUING AND METADATA PRACTICES

Early topographic maps of Canada, especially map series produced before the introduction of the NTS numbering, are not well standardized (Dubreuil, 1992). The early topographic map series do not cover the whole country, nor do they adhere to similar production standards (Dubreuil, 1992; Sebert, 1976), making library collections of early map series scarce and incomplete across Canada. Access to these series in libraries is not always available, and users often rely on digital catalogue records, found online through the library’s catalogue, and paper (or scanned) indexes to discover whether maps exist for a given geography. Typically, web access to digital map collections varies in complexity, ranging from lists of links (sometimes with thumbnails) to interactive web maps. Enhancements, such as the ability to zoom, pan, and view scanned maps as overlays on top of an existing basemap, offer a more desirable user experience for researchers.

Libraries and map collectors have often struggled with maintaining efficient storage and retrieval methods for large map series such as the national topographic maps (Andrew & Lamont, 1998). Organizing and accessing a series of hundreds or even thousands of maps, which often include historic editions and multiple versions of individual sheets, can easily become a monumental task (Andrew & Lamont, 1998). The early and modern Canadian NTS map series are considered large national map series and few libraries (if any) in Canada catalogue individual sheets from the NTS.

Some libraries in the United States, including Penn State Libraries, have undertaken individual map sheet cataloguing for select national mapping series, including the U.S. Geological Survey's 7.5-minute topographic quadrangles (Andrew & Lamont, 1998). There are also some individual sheets catalogued from the early NTS, including the set of Canada’s Militia and Defence Maps, 1905-1931, that have been published by Lorraine Dubreuil at McGill Libraries (Dubreuil, 1992). However, by no means is there a complete set of catalogue records or metadata describing these early topographic map series in Ontario libraries, which represents quite a challenge for those trying to find and access these maps online.

Cataloguing at this level of detail is costly and difficult to maintain. Most libraries typically have series-level catalogue records which direct the user to the map library’s physical map cabinets. Sometimes, large map series are catalogued at the state or provincial level, to assist with identification and retrieval by geography, but this varies across libraries (Andrew & Lamont, 1998). Typically, a paper or digital index is provided to users, and this acts as the primary finding aid and retrieval mechanism for large map series.

With the OCUL historical map digitization project, the project inventory of maps covering Ontario will become instrumental for understanding the totality of the series as well as the distribution of map holdings across the province. The inventory is also helpful for data collection, in that member libraries that are participating in digitization work can also collect consistent descriptive information from the individual map sheets that can be used to generate standardized metadata

for the digital images. The inventory to date has tracked information such as map sheet title, subtitle, edition (which is very important because there are several editions of the same sheets), year published, source library collection (whether or not the sheet was held by a particular library, and if it was scanned, georeferenced, etc.), survey date, publisher, grid presence (some of the maps contain grid lines), and more.

Overall, the transition from cataloguing to metadata represents a change for how libraries manage, describe, and provide access to digital maps and data online. In addition to a metadata format and standards change, technological advances in web-based GIS offer libraries the option of adopting new techniques for storage, enhanced identification, and access, through online catalogues and portals designed specifically for geographic information. The creation of standard metadata for the digital images and data being created by the libraries is critical to digitization work, since it has a direct impact on the granularity of access and the preservation of original map content in digital form. Metadata for digital objects, including geospatial data and digital maps, is essential for understanding the scope and content of the digitized work. Metadata provide descriptive information about the resource in a structured, standard, and transferable format (e.g., XML). GIS tools require metadata in machine-readable formats in order to read and process data and information that is useful for end users.

Today, the most common metadata standard for the description of digital geospatial data, including georeferenced map images, is the International Standards Organization (ISO) 19115 for Geographic Information. The North American Profile (NAP) of the ISO 19115, which is a set of fields specific to North America, is heavily used by government and data producers, with the Government of Canada formally adopting the standard in 2012.

The creation of standard metadata for this collection will largely be accomplished through the development of a metadata crosswalk, which will map structured information contained in the project's inventory spreadsheet to fields in the metadata standard that we are using to describe these maps. It was decided early on that individual map sheets will be described as datasets using the ISO 19115 NAP standard, since this is the standard that is used on OCU's Scholars GeoPortal platform. The structure of the mapping and information is still under consideration; however, a significant amount of the metadata mapping work has been completed already.

In the development of the mapping, we consulted some major map libraries and online geospatial data repositories, including the Harvard Map Collection, Harvard College Library⁴, and the University of Ottawa's digital map collection⁵. These collections describe similar digital geospatial data and scanned historical maps. As part of the project, an example of a sheet-level metadata record and mapping is provided for the Map of Peterborough, Ontario, 1932 [1:63,630]⁶.

⁴ Accessed at Geodata at Tufts <http://geodata.tufts.edu/>

⁵ Accessed at http://gsg.uottawa.ca/geo/indexes/hist_topo_maps.htm

⁶ Map of Peterborough, Sheet No. 031D08, 1932 [1:63,360]
<http://gsg.uottawa.ca/geo/ocul/test.xml>

5 ACCESSING THE COLLECTION ONLINE

One of the options the team is considering for online access to the collection is to present the images as one seamless image mosaic. Brock University's Map, Data & GIS Library has achieved some success using this approach. Georeferenced topographic maps are stored in a mosaic dataset, which is a data model within the geodatabase used to manage a collection of raster images. More precisely, a mosaic is a collection of raster datasets stored as a catalogue of individual maps and viewed as a mosaicked image that is dynamic, where the properties of the original imagery are maintained (Childs, 2010). The processes involved in creating a mosaic dataset using ArcMap are briefly described below.

First, a mosaic dataset is created within a new or targeted geodatabase. Prior to adding the raster images, the coordinate system is set to 'Web Mercator Auxiliary Sphere', which is compatible with other web map interfaces. A mosaic dataset consists of a footprint feature class that acts as a type of index. The process of building footprints involves defining a boundary on each map to the extent of the displayed image, such as the neatline. This is considered a 'virtual' crop. It is not necessary to permanently remove the margins of the map, or map collar, to create a seamless display. This footprint method preserves the map in its entirety, but displays only the content defined by the footprint (in this case, all information within the neatline). Although this process can be done manually, automating the footprint creation using a general-purpose language (like Python) is preferred⁷. Another option, although not yet tested on topographic maps, involves the use of the Image Boundary tool in QGIS. These processes require further exploration.



Figure 8. Mosaicked dataset of NTS maps of the Niagara Region produced by Brock University.

⁷ Accessed at gis.stackexchange.com, or the ArcGIS Online Group: ArcGIS Image Management Workflows www.arcgis.com/home/item.html?id=bdb45bdcf20f4b7d83f2975727858b33

A recommended procedure for working with mosaic datasets is the creation of overviews. This process is similar to the building of pyramids, where a set of reduced-resolution datasets are generated in order to optimize the performance of a mosaic dataset and increase display speeds at various scales. Unlike pyramids, overviews are not produced for individual rasters. Instead, they are derived by mosaicking multiple rasters (Woo, 2012). A seamless display can be achieved by creating a mosaic of images that are fused together to create one large image. We are currently testing this technique for the display and visualization of the digitized historical topographic maps produced for the OCUL project.

Using a platform that provides an underlying basemap on which to overlay historical images is a useful enhancement, enabling users to situate themselves using both contemporary and historical features as reference points. Platforms such as Google Earth, ArcGIS Online, and OCUL's own Scholars GeoPortal (built on ArcGIS Server) provide these options. Yet, not all users want to use a georeferenced image in a GIS environment. For this reason, we are exploring options for making the images available for download at a high resolution in both georeferenced and non-georeferenced versions. Digital map files can be produced in several formats, including GeoTIFF, JPEG, KMZ, and GeoPDF. Each vary in file size and resolution depending on the scanning and georeferencing processes applied in the preparation of the images.

In reviewing current practices for the display of these large map series online, two website examples that display U.S. topographic maps are of particular interest and differ in display and download options. The *USGS Historical Topographic Map Explorer*⁸ designed by Esri provides a seamless map overlay with transparency. Another example, the *USGS TopoView*⁹, offers multiple file formats for download (JPEG, KMZ, GeoPDF, and GeoTIFF) and enhanced search options, but no seamless map display or transparency functionality. It functions more as an online index to all USGS topographic maps, rather than as an interactive viewer. A portal that integrates a combination of options from both these sites, while respecting web accessibility guidelines, is ideal, and this kind of functionality is being explored for our current project.

Our intent is that Scholars GeoPortal, mentioned earlier, will house the digital images, mosaic services, and metadata, and provide access to this collection openly for anyone to discover and use. The GeoPortal will provide access to the lower resolution (300 ppi) scans (originally stored in TIFF format) and georeferenced data, while the original high quality (600 ppi) scans and georeferenced points will be archived for long-term preservation and reuse should the need arise.

6 LOOKING FORWARD

At the time of writing (12 months into the 28 months allotted), the project is well underway. A majority of the maps have been scanned and are currently being georeferenced. Team members have been turning their attention toward how materials will be displayed on Scholars GeoPortal and the logistics of making the files available for download. Projects undertaken at OCUL member institutions, such as the topographic map mosaicking at Brock University, have also provided us with insight to carry forward into the next phase of the project.

⁸ Accessed at <http://historicalmaps.arcgis.com/usgs>.

⁹ Accessed at <http://ngmdb.usgs.gov/maps/TopoView/viewer>.

The group is closely watching the work of the Canadian Historical GIS Partnership initiative, whose forthcoming white paper on data and information visualization for online Historical GIS (HGIS) applications should offer insight into the current state of these technologies (Roy, 2015). We anticipate that the digitization of historical maps in libraries will facilitate historical GIS applications in research across a variety of disciplines.

Our process for achieving a web display for the Ontario historical topographic maps that meets our desired criteria is still being explored. However, providing access to the georeferenced images, in addition to the non-georeferenced scans, at a manageable file size on the Scholars GeoPortal platform is our current intention.

7 CONCLUSIONS

Managing a map digitization project across several institutions can be a daunting task. However, this distributed configuration has allowed us to expand our access to map collections, staff expertise, and existing equipment, as well as provide job opportunities for students at multiple universities. Collaborative platforms such as Google Sheets have enabled us to manage inventories and status updates from any location, which has greatly assisted the coordination of activities around this distributed project.

Canada's historical topographic map series provide unparalleled detail about the past. These early map series are not being digitized and archived elsewhere, and are therefore the focus of digitization efforts in OCUL libraries today. We hope that by digitizing and providing access to the 1:25,000 and 1:63,360 maps of Ontario, other provinces will follow suit and we will begin to see a national collection of these digital maps emerge and evolve.

Another significant outcome of this project has been the development of best practices and specifications that can be reused by libraries across Ontario, and even throughout Canada. Recognizing the importance of standards and best practices in libraries for map digitization, georeferencing, and metadata provides a foundation for ensuring access to these collections well into the future.

The authors would like to thank Michel Castagné for his copy editing assistance.

7 REFERENCES

- Allord, G. J., Fishburn, K. A., & Walter, J. L. (2014). *Standard for the U.S. Geological Survey Historical Topographic Map Collection (ver. 2)*. U.S. Geological Survey Techniques and Methods, book 3, chap. B11, 11 p., <http://dx.doi.org/10.3133/tm11B03>.
- Allord, G. J., Walter, J. L., Fishburn, K. A., & Shea, G. A. (2014). Specification for the US Geological Survey Historical Topographic Map Collection (No. 11-B6). US Geological Survey. <http://dx.doi.org/10.3133/tm11B6>.
- Andrew, P. G., & Lamont, M. (1998). Bending the rules. *Technical Services Quarterly*, 15(3), 35-48. doi:10.1300/J124v15n03_03

Banach, M., Shelburne, B., Shepherd, K., & Rubenstein, A. (2011). *Guidelines for Digitization*. UMass Amherst Libraries.

Burt, J. E., White, J., and Allord, G. J., (2012), QUAD-G—*Automated georeferencing project: University of Wisconsin–Madison, Department of Geography*. Retrieved from <http://www.geography.wisc.edu/research/projects/QUAD-G/>.

Burt, J.E., White, J., and Allord, G. J. (2014). QUAD-G--*Automated georeferencing of scanned map images, user manual version 2.10*: University of Wisconsin–Madison, Department of Geography. Retrieved from <http://www.geography.wisc.edu/research/projects/QUAD-G/files/QUAD-GUserManualver2.10.pdf>.

Childs, C. (2010). *On-the-Fly Processing and Dynamic Raster Mosaicking*. Esri Education Services. Retrieved from <http://www.esri.com/news/arcuser/0610/files/mosaicdataset.pdf>.

Dale, R. L., Leech, R., Bogus, I., Mathews, D., & Blood, G. (2013). *Minimum Digitization Capture Recommendations*. The Association for Library Collections and Technical Services Preservation and Reformatting Section. Retrieved from http://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations#static_media.

Dubreuil, L. (1992). Canada's Militia and Defence Maps, 1905-1931. ACMLA Occasional Publication #4, 2-4, 11.

Federal Agencies Digitization Initiative - Still Image Working Group. (2016). *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files*. Retrieved from http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image_Tech_Guidelines_2015-09-02_v4.pdf

Hackeloeer, A., Klasing, K., Krisp, J. M., & Meng, L. (2014). Georeferencing: a review of methods and applications. *Annals of GIS*, 20(1), 61-69.

Hill, L. L. (2009). *Georeferencing: The geographic associations of information*. Cambridge, MA: MIT Press.

Image Management Workflows. (n.d.). Retrieved January 14, 2016, from <http://resources.arcgis.com/en/communities/imagery/01850000000v000000.htm>

Knowles, A. K., & Hillier, A. (2008). *Placing history: how maps, spatial data, and GIS are changing historical scholarship*. Redlands, CA: ESRI, Inc.

Nicholson, N. L., & Sebert, L. M. 1981. *The Maps of Canada*. Hamden: Archon Books and Folkestone: Wm Dawson & Sons Ltd.

Roy, K. (2015, November 9). Some thoughts on the state of HGIS visualization. Retrieved January 14, 2016, from <http://geohist.ca/2015/11/some-thoughts-on-the-state-of-hgis-visualization/>

Sebert, L. M. (1976). The One Inch to One Mile Series of the National Mapping Program. *The Canadian Cartographer*, 13(2), 123-131.

Trimble, L., Woods, C., Berish, F., Jakubek, D., & Simpkin, S. (2015). Collaborative Approaches to the Management of Geospatial Data Collections in Canadian Academic Libraries: A Historical Case Study. *Journal of Map & Geography Libraries*, 11(3), 330-358.
doi:10.1080/15420353.2015.1043067

Woo, S. (2012). Should I build pyramids or overviews? Retrieved March 7, 2016, from <https://blogs.esri.com/esri/arcgis/2012/11/14/should-i-build-pyramids-or-overviews>