

# Georeferenced Control Points for Historical Maps

Aabid Shamji, John Gouwar, and Jerod Weinman\*  
Grinnell College

December 10, 2019

## Abstract

This data set consists of nearly 6,400 control points georeferencing 31 historical maps (1831–1927) drawn from the David Rumsey collection. Geographical coordinates use the NAD83 geographical coordinate system. The data is available at <https://doi.org/11084/23330>.

## Overview

This data set features manually-identified ground truth correspondences between geographical coordinates (latitude and longitude) and pixel coordinates (row and column) in historical map images. The 31 maps in the data set are originally drawn from the David Rumsey collection [2], but were curated for earlier work in toponym detection and recognition [6, 7, 8]. Other data sets involving these images include toponym annotations alone plus their unique geographic identifiers [1] and complete text annotations [5]. This data was created for the empirical evaluation of automated georeferencing techniques [4, 3]

## Geographical Support Data

The ground truth coordinates are ultimately drawn from several data sets provided by the United States Geological Survey (USGS).

**National Boundary Dataset (NBD):** Modern state and county boundaries.<sup>1</sup>

**National Transportation Dataset (NTD):** Highway paths and junctions.<sup>2</sup>

**Geographic Names Information System (GNIS):** Official locations and names (along with historical or alternate names and spellings) of geographical features along with their geographical class (e.g., *Arroyo*, *Basin*, *Cliff*, *Dam*, etc.).<sup>3</sup>

Because this data is part of the U.S. federal government, the standard geographical coordinate system (GCS) of the data is the North American Datum of 1983 (NAD83), which utilizes the Geodetic Reference System of 1980 (GRS80) as the reference ellipsoid.

## Process

As described by Howe et al. [3], the correspondences arise from two different manual matching processes, one graphical and one textual. In the first case, human annotators matched the geographical and image coordinates of well-localized map locations—e.g., junctions, corners, or other topologically distinct points. The ground truth geographical coordinates were drawn from USGS-provided GIS data described above, specifically state and county boundaries from the NBD and road junctions from the NTD.

In the second case, human annotators matched the labeled toponyms on the maps from earlier data [1] with entries from the USGS Geographic Names Information System (GNIS), subsequently allowing the

---

\*Corresponding author: [jerod@acm.org](mailto:jerod@acm.org)

<sup>1</sup><https://www.sciencebase.gov/catalog/item/4f70b219e4b058caae3f8e19>

<sup>2</sup><https://www.sciencebase.gov/catalog/item/4f70b1f4e4b058caae3f8e16>

<sup>3</sup><https://geonames.usgs.gov>

system to put the image coordinates of the toponym’s place marker (usually a city) in correspondence with the geographical coordinate reported by the GNIS.

By combining correspondences from both methods, annotators produce a sufficient number of well-distributed points so that a parametric alignment could fit reasonably well over the entire map image, given a reasonable projection.

## Format

Each map has associated with it some number of files with the name `basename.n.txt`. — where `basename` is the map’s base filename, e.g., `D0006-0285025`, and `n` is an integer (starting at 1), describing the unique coordinate system in use on the map image. The files each contain tab-separated lines with four numbers of the form

```
col row lon lat
```

or equivalently

```
ximg yimg xgcs ygcs
```

## References

- [1] Larry Boateng Asante, David Cambroner Sanchez, Ravi Chande, Dylan Gumm, and Jerod Weinman. A data set of annotated historical maps. Technical report, Grinnell College, Grinnell, Iowa, July 2017. <https://doi.org/11084/19349>.
- [2] Cartography Associates. David Rumsey map collection. <https://www.davidrumsey.com>.
- [3] Nicholas R. Howe, Jerod Weinman, John Gouwar, and Aabid Shamji. Details of deformable part models for automatically georeferencing historical map images. Technical report, Smith College, Northampton, Massachusetts.
- [4] Nicholas R. Howe, Jerod Weinman, John Gouwar, and Aabid Shamji. Deformable part models for automatically georeferencing historical map images. In *Proc. of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '19)*, Nov. 2019. <https://doi.org/10.1145/3347146.3359367>.
- [5] Archan Ray, Ziwen Chen, Ben Gafford, Nathan Gifford, Jagath Jai Kumar, Abyaya Lamsal, Liam-Niehus Staab, Jerod Weinman, and Erik-Learned-Miller. Historical map annotations for text detection and recognition. Technical report, Grinnell College, Department of Computer Science, Grinnell College, Grinnell, Iowa 50112, Oct. 2018. <https://doi.org/11084/23294>.
- [6] Jerod Weinman. Toponym recognition in historical maps by gazetteer alignment. In *Proc. IAPR International Conference on Document Analysis and Recognition (ICDAR '13)*, pages 1044–1048, Aug. 2013. <https://doi.org/10.1109/ICDAR.2013.209>.
- [7] Jerod Weinman. Geographic and style models for historical map alignment and toponym recognition. In *Proc. IAPR International Conference on Document Analysis and Recognition (ICDAR '17)*, pages 957–964, 2017. <https://doi.org/10.1109/ICDAR.2017.160>.
- [8] Jerod Weinman, Ziwen Chen, Ben Gafford, Nathan Gifford, Abyaya Lamsal, and Liam Niehus-Staab. Deep neural networks for text detection and recognition in historical maps. In *Proc. IAPR International Conference on Document Analysis and Recognition (ICDAR '19)*, pages 902–909, Sep. 2019. <https://doi.org/10.1109/ICDAR.2019.00149>.

**Acknowledgment** This material is based upon work supported by the National Science Foundation under Grant No. 1526350.