

Innovations in historical topographic map geoprocessing with open-source technologies

Wesley M. Gillman ^{1*}

¹ U.S. Geological Survey

*wgillman@usgs.gov

Keywords: georeferencing, historical maps, automation

The historical quadrangle scanning project was started nearly 20 years ago by the National Geospatial Program (NGP) to digitally preserve, catalog, and disseminate traditionally printed quadrangle map sheets. In this way we are effectively documenting the history of topographic map production in the United States as led by the U.S. Geological Survey (USGS) along with cooperating federal agencies, since the late 1800s. High-resolution digital scans were contributed from map libraries across the country, including universities, state and federal agencies. The National Map (TNM) currently offers more than 185,000 downloadable map products in the Historical Topographic Map Collection (HTMC) covering the United States and territories.

Completion of this task was made possible by the automated georeferencing techniques of a program known as Quad-G. Developed in coordination with the University of Wisconsin-Madison Geography Department, Quad-G applies an ingenious approach to locating ground control points (GCP) through image pattern recognition. The program automatically distinguishes the various types of printed hash marks and applies a least-squares fitting to yield a set of high-accuracy transformation coordinates. Operators can take control when needed and manually place the marks for those maps that present more of a challenge for automated processing. Original spatial reference systems (SRS) and unique GCP data comprise the spatial metadata that is applied in the geoprocessing workflow with the command-line tools in the Geospatial Data Abstraction Library (GDAL). The first-generation product was a TerraGo GeoPDF, with a projected coordinate system and custom neatline coordinate boundary, intended for use with the TerraGo Toolbar in Adobe Reader.

This year at the National Geospatial Technical Operations Center (NGTOC) we are revisiting the product specification with our sights set on a new open-source geospatial PDF. Revised operating procedures will enable us to offer an original Cloud Optimized GeoTIFF (COG), as a higher quality alternative that naturally suits the raster data format of the scanned maps. Refined data management helps keep Section 508 compliance a priority right from the start. In accordance with the Geospatial Data Act of 2018 (GDA), deliverables will now be accompanied by an ISO 19115 metadata XML. These supplementary files provide automated systems and humans alike with immediate access to expanded product information. The transition to a new public metadata standard was facilitated through adoption of an ISO-Ready JavaScript Object Notation (JSON) schema as our internal data exchange format.

With this new edition we aim to apply modern solutions to enhance the public's use of historical USGS map products. We anticipate releasing the first new products within 2026 with replacements continuing into the following year. Newly generated copies will supersede all previous editions in public facing services as they are released. All

original and derived materials will continue to be retained for archival purposes. Revised editions of the project standards and specifications are also under development. In a significant shift to our publication outlet for TNM datasets, future product releases will not be submitted individually to the new Science Data Portal, which will be reserved for aggregate collections. All products will continue to be available to the public through the TNM Downloader and TNM Access API.