Austro-Hungarian Maps Dataset README

Version information

Version 1.1 | Last updated: 2024-04-17

- *1.1 (2024-04-17)* LC Labs consolidated coversheet and README content with minor formatting updates
- *1.0 (2022-09-22)* First version

CONTENT ADVISORY

Maps, like other archival materials, are produced from the particular socio-political perspective of their creators, expressed through the use of labels, choice of language, drawing of borders, and other visual symbology. Maps can contain contested land claims, offensive place names, and partial perspectives on political and cultural conflict. The historical maps contained in this dataset were initially prepared and issued by the Austro-Hungarian Monarchy's Milita■rgeographisches Institut beginning around 1875. After the dissolution of the Austro-Hungarian Empire in 1918, parts of the set were continued by successive governments. Some portions of this set were produced by Third Reich Germany in the 1930s and 1940s during and after annexing areas to its south and east. Most of this set is produced by various governmental military mapping agencies.

The regions covered by this set of maps are linguistically and culturally diverse. Generally, place names in the set are identified by their German cognates, but they are also known to sometimes be in Hungarian or Slavic languages.

About the source data or collection

Brief description & background of collection

This set of maps is part of the [multi-sheet map set collection](#) held in the Library of Congress Geography and Map Division. It depicts the Austro-Hungarian Empire (officially the Austro-Hungarian Monarchy) and its successor states at a scale of 1:75,000. The Library of Congress's set spans the 1870s through the 1940s, with the bulk from the 1880s through the early 20th century. The Library considers it a composite set, because it was created by multiple entities over time. The set was initially prepared and issued by the Austro-Hungarian Monarchy's Milita■rgeographisches Institut beginning around 1875, as part of its Third Military Survey. The survey included 1:25,000 and 1:12,500 scale maps ("Aufnahmeblätter") and 1:75,000 scale maps ("Spezialkarte")—with the current set being the latter. It covered the whole of

Austria-Hungary and sometimes extended beyond its borders. After the dissolution of the Austro-Hungarian empire in 1918, parts of the set were continued by successive governments. Despite its multiple creators over time, the set is cataloged and managed by the Library of Congress as a single composite map set, under the title from its early editions: *Spezialkarte der o■■sterreichisch-ungarischen Monarchie* (or *Special map of the Austro-Hungarian Monarchy*). It is often colloquially referred to as the "Austro-Hungarian map set."

This is one of the most sought-after map sets in the Geography and Map Division, particularly by genealogists. Austria-Hungary's constituent nations served as major sources of emigration to the United States in the late nineteenth and early twentieth centuries. Searching for an ancestor's birthplace in Austria-Hungary can be challenging, especially given the variant spellings in place names and dialects, which owe to the changing nature of rule and consequent shifting of borders after the First World War. The region was extremely diverse, both linguistically and culturally, and covered parts or all of modern-day Austria, Bosnia, Croatia, Czech Republic, Hungary, Italy, Moldova, Montenegro, Poland, Romania, Serbia, Slovakia, Slovenia, and Ukraine.

The original set is divided into well over 1,000 sheet tiles. An index map of the set, from 1916, can be found online at https://hdl.loc.gov/loc.gmd/g6481a.ct011968, showing the system of sheet tiles and their 4-digit coordinate IDs. Early editions used a zone system for labelling sheet tiles, later superseded by the 4-digit system used in the accompanying dataset. The Library of Congress's collection of maps from this set is extensive but not complete. There are some sheet tiles for which the library has no editions, particularly those along border areas. The Library of Congress's collection was acquired from multiple, diverse sources over several decades.

More details about how to use this set can be found online on the guide *Cartographic Resources for Genealogical Research: Eastern Europe and Russia > Austria-Hungary*.

Original format

Generally 38 x 52 cm paper maps, either in black and white or color. Some sheets mounted on cloth. Some sheets are photostatic negatives.

Library of Congress reading room

Geography and Map Reading Room, https://www.loc.gov/rr/geogmap/

Contact

For more information please contact the Geography and Map reference specialists at https://ask.loc.gov/map-geography.

Scale of description

The entire map set is described at https://lccn.loc.gov/2016431073.

Rights information

The Library of Congress is unaware of any copyright or other restrictions in this map set. Absent any such restrictions, these materials are free to use and reuse. The determination of the status of an item ultimately rests with the person desiring to reproduce or use the item.

Digitization information

The images in this dataset were imaged in 2015 by GIS Research Fellows, using a large format sheetfeed scanner to 300ppi.

About this exploratory data package

This set is cataloged and managed by the Library of Congress as a single map set, under the title from its early editions: *Spezialkarte der o██sterreichisch-ungarischen Monarchie.* It is often colloquially referred to as the "Austro-Hungarian map set*."* More details about the content and provenance of the set can be found in this dataset's coversheet.

This dataset contains scans and georeferenced images produced as part of an experimental 6-month project in 2015, under the Library's former GIS Research Fellows program. The original project was known as the *Geographic Hot Spot Dynamic Indexing Project.* It was managed by John Hessler, Specialist in Computational Geography & Geographic Information Science in the Geography and Map Division, and included the work of four fellows: Amanda Brioshe, Erin Kelly, Evan Neuwirth, and Michael Schoelen. The goal of the project was to develop and test enhanced workflows for georeferencing map sets at scale. This georeferenced set was one of the many produced during the course of that project, and has not previously been made public.

The Geography and Map Division is currently revisiting the products of this and other experimental georeferencing workflow projects.

What's included?

The data package includes:

- Several subdirectories:
  - "sheets_img" containing 4,998 non-georeferenced map sheet images. (TIFF image format)
  - "sheets_geo" containing 4,877 georeferenced map sheets, with the map collars removed. (GeoTIFF image format)
  - "index_img" containing 3 index map sheet images. (TIFF image format)

- - "index_shapefile" containing 4 files making up a single "shapefile" that represents the georeferenced tiffs. (shapefile format)
  - "legend_img" containing 3 legend sheet images. (TIFF image format)
- `metadata.csv`: a CSV with basic metadata
- `manifest.txt`: a text file listing the image id, MD5 hash, and location of the images in the data set
- `README` (this document): An overview of the source data or collection provenance, the contents of the data package, and how the data package was created. Available as .md, .html, and .pdf.
- Data cover sheet: a more substantive overview of the data and the collection from which it is derived
- 95 randomly selected items from the *9,885* set with their corresponding image files have been provided as sample data. Included with this are a `metadata.csv` and `manifest.txt`.

Filenaming of sheets

The bulk of the dataset is composed of the 4,998 non-georeferenced scans (in the "sheets_img" directory) and their 4,877 georeferenced counterparts (in the "sheets_geo" directory). Their filenames will appear like:

- 5062_000_img.tif
- 5062_000_geo.tif

The first set of numbers (four digits) is the sheet ID from the original set of sheets, typically printed on the upper right corner of the sheets themselves. For example, "5062" is the numerical ID for the tile covering the area around and to the south of Budapest. The set includes 1003 distinct tiles, all identified by a four-digit ID. Early sheets in the set used a different ID system and did not have this four-digit ID printed on the sheet at the time of original publication. For many of these sheets, Library of Congress staff or previous custodians have hand-written the four-digit sheet number at the top right of the sheets.

The second set of number (three digits) is the sequential number of the editions within each sheet set, beginning with 000. The Budapest tile ("5062") has six editions, and so its files range from 5062_ **000** to 5062_ **005.** The most common number of editions for any given tile in the collection is six, similar to the Budapest tile. The highest number is 14, and the lowest is one (see the chart below). Editions within the same tile will generally share the same bounding coordinates over time.

Edition count	Number of tiles at that edition count
1	103

2	86
3	79
4	133
5	167
6	185
7	118
8	63
9	39
10	16
11	5
12	3
13	4
14	2

Editions were filenamed in the sequential order in which they were found in map drawers at the time of the original project. Generally, they will be ordered from most recent to earliest. So, in the example of the Budapest tile, 5062_000 is the most recent edition (published 1914) and 5062_005 is the earliest edition (published 1884).

The last portion of the filename refers to whether the image is either a georeferenced GeoTIFF with the map collar cropped out ("geo"), or is an original scan of the map sheets, cropped and rotated for clarity but including the map collar ("img"). This mirrors the organization of the files into the "sheets_geo" and "sheets_img" directories. Note that 121 of the original non-georeferenced scans do not have georeferenced counterparts. For these, GeoTIFFs were not produced in the original project.

GeoTIFF files and extracting geotransform coordinates

GeoTIFF is a format extension of the TIFF file format. GeoTIFF files are TIFF files that also contain additional spatial metadata that ties the image to a real-world location using a geographic coordinate system. Specific raster points on the image (x,y coordinates) are typically linked to coordinates (latitude and longitude) in a spatial reference system. The geographic coordinate system is also specified (such as WGS 1984), so that mapping environments (often referred to as Geographic Information Systems) know how to interpret the coordinates. When

GeoTIFF files are opened in standard image viewers, they can look warped or stretched. When opened in a GIS or mapping environment, they will display correctly.

It is possible for GeoTIFFs to have multiple control points distributed throughout the interior of the raster image, often at well-known landmarks or at specifically designed points designed for georeferencing. In the case of this dataset, a relatively simple form of georefencing called "geotransforming" was used, wherein only the corner coordinates are specified in the GeoTIFF. The top left coordinate is stored directly in the GeoTIFF, and the remaining coordinates can be easily calculated. A common library used for working with GeoTIFF files is GDAL, which also has Python bindings. GDAL documentation of how to calculate geotransformed GeoTIFF coordinates is available at https://gdal.org/tutorials/geotransforms_tut.html. Based on that documentation, the following Python code can be used to calculate the bounding coordinates of a geotransformed GeoTIFF:

```
 from osgeo import gdal
src = gdal.Open(filename.tiff)
ulx, xscale, xrot, uly, yrot, yscale = src.GetGeoTransform()
lrx = ulx + (src.RasterXSize * xscale)
lry = uly + (src.RasterYSize * yscale)
top_left = [ulx,uly]
top_right = [lrx, uly]
bottom_left = [ulx, lry]
bottom_right = [lrx,lry]
```

Computational readiness and possible uses

This dataset is the product of an experimental GIS Research Fellows project. Users could approach it from an institutional perspective, with the goal of helping to develop automated workflows for performing quality control on past projects, reviewing their readiness for reuse and public distribution, and extracting additional metadata helpful for discovery and reuse. Or, users could approach it as a research dataset ready for reuse.

Collection readiness and quality control tasks might include: analysis of image quality (e.g., are TIFFs uniformly uncompressed, do they have a uniform PPI, are there signs of lossy compression artifacts), checks on whether the GeoTIFFs are valid and well-formed, checks on whether GeoTIFF spatial information is uniformly embedded in files as geotranforms, analysis of the spatial reference transformations, and filenaming checks (e.g., do sheets appear to be named with the wrong sheet ID).

The map sheets themselves contain a rich array of additional information that has not yet been datafied, such as publication dates, place names, and publishers. The Library would typically provide sheet-level date information when creating a public digital collection for a map set.

Possible research uses might include spatial visualization, analysis of change over time, feature or text extraction, or automation of cropping historical map collars.

The GeoTIFFs in this collection contain coordinate and projection information that allows for spatial analysis, web mapping, and use in GIS environments. The shapefile included in this dataset may be helpful as a quick reference tool to access pre-computed spatial coordinates for the GeoTIFF images. The metadata.csv file may be helpful in segmenting the collection and downloading targeted sections.

How was it created?

This experimental dataset was produced in 2015 under the GIS Research Fellows program's *Geographic Hot Spot Dynamic Indexing Project*. It is one of many map sets that were digitized and georeferenced as part of that project. The general workflow for the project consisted of:

1. Imaging the physical sheets using commercial sheetfeed scanner, to 300 ppi TIFF files (likely using image editing auto-settings on the scanner).
2. Manually transcribing the sheet coordinates and translating them to the Greenwich-meridian system (from the Ferro-Island-meridian system used on the original sheets).
3. Straightening the image and cropping the map collar. (At the start of the project this was accomplished manually in Photoshop and later using an automated script developed as part of the project. It is not known which method was used on this particular set.)
4. Adding the coordinate information to the TIFF file to convert it to a GeoTIFF (using an application called Quad-G).
5. The GeoTIFF images were then loaded into ArcMap and used to generate a mosaic dataset footprint file. The footprint file was exported in shapefile format to create the shapefile found in this dataset.

The dataset was further modified in 2022:

1. Filenames were standardized for clarity.
2. The shapefile attribute table was updated with updated filenames and re-exported using Esri's arcgis Python library.
3. The "sheet_img" files were cropped and deskewed (rotated) for clarity.
4. 104 images in the "geo" directory that used approximate transformations stored in Esri .tfwx and .aux files rather than GeoTIFFs were removed.
5. A metadata.csv file was generated, based on metadata in filenames.

Spatial Reference Information

The map sheets were originally mapped under various mapping parameters. Older sheets use a Bessel 1841 reference ellipsoid, which was at some point switched to the 1892 Austrian datum.

Additionally, there are at least six different map projections used among sheets in the set, varying by both time period and area within the set. In order to account for this, the Geography and Map Research Fellows followed techniques outlined in Gabor & Timar's "Mosaicking of the 1:75,000 Sheets of the Third Military Survey of the Habsburg Empire" (2009). The location parameters of the geodetic datum used for transformation to modern projection systems are the followings: dX = +600 m; dY = +205 m; dZ = +437 m, giving exact fit at the fundamental point of Hermannskogel. This is estimated to result in a maximum error of 220 meters.

Dataset field descriptions

Below is a description of the metadata fields found in the metadata.csv file. These metadata values are derived from the information embedded in the original filepaths from 2015. They can be used to quickly slice the collection into specific portions for download.

filepath	relative path to the file, from the dataset's parent directory
filename	name of the file, with the extension
parent_dir	parent directory of the file
file_format	TIFF - ungeoreferenced TIFF file GeoTIFF - georeferenced TIFF file shapefile - the component files that together constitute a shapefile use the value "shapefile".
object_type	index map - a shapefile index map or digitized historical index map legend - a digitized historical sheet that included only a legend full map sheet - original scans of the map sheets, cropped and rotated for clarity but including the map collar georeferenced map - map sheets cropped down just to the map, without the map collar, and georeferenced
tile_id	4-digit id of the tile
edition	3-digit sequential number of the edition within the tile

Rights Statement

The Library of Congress is unaware of any copyright or other restrictions in this map set. Absent any such restrictions, these materials are free to use and reuse. The determination of the status of an item ultimately rests with the person desiring to reproduce or use the item.

Creator and contributor information

Scans and georeferenced images created by John Hessler, Amanda Brioshe, Erin Kelly, Evan Neuwirth, and Michael Schoelen

README creators: Rachel Trent, Meagan Snow

README contributors: Diane Schug-O'Neill, John Hessler, Sundeep Mahendra, Eileen J. Manchester, Chase Dooley