

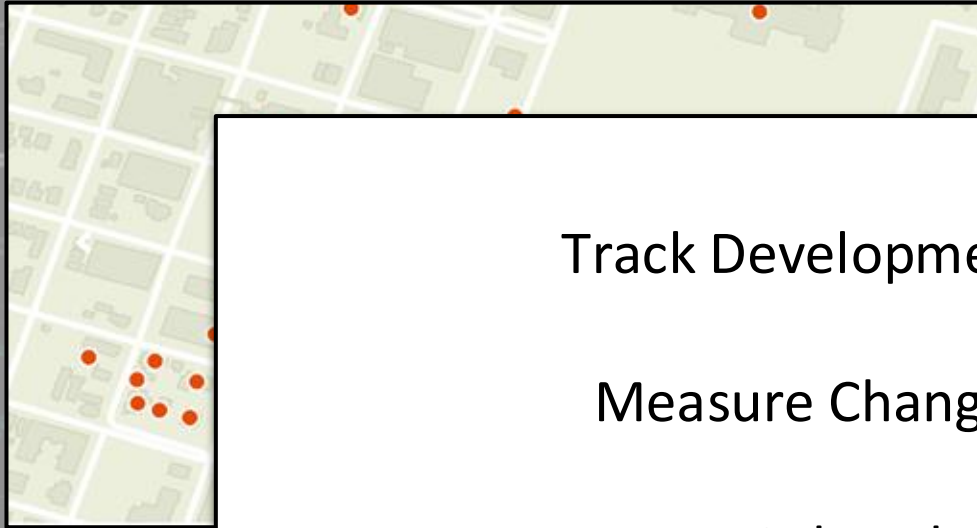
Enabling the Dissemination of Large Volumes of Historical Texas Maps Using an Automated Georeferencing Workflow

Michael Shensky, Katherine Strickland, **Alex Marden**, and Hannah Dubbe
University of Texas at Austin

2024 Texas GIS Forum

"This work was supported by Good Systems, a research grand challenge at the University of Texas at Austin."

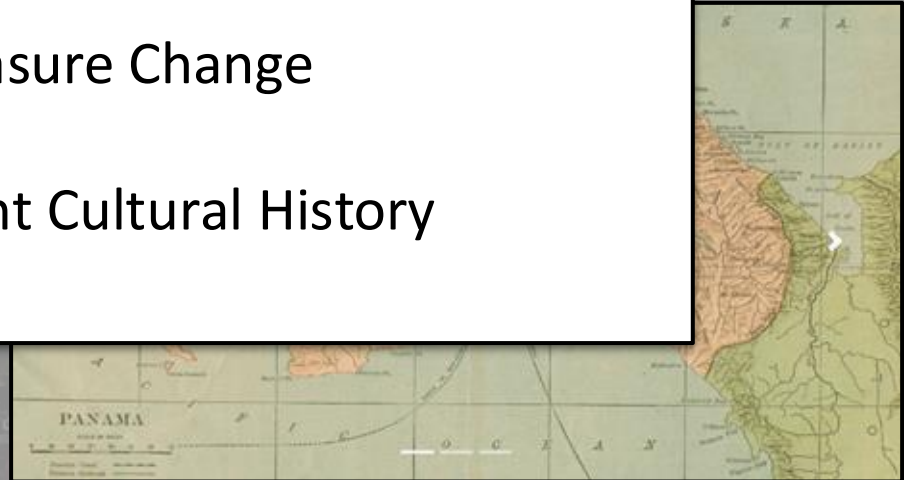
Importance of Historical Geospatial Data



Track Development

Measure Change

Document Cultural History



Texas GeoData Portal

- Developed by the UT Libraries at UT Austin
- Launched in **November 2019**
- Facilitates access to **UTL geospatial data**
- 2,675 UTL datasets currently available
- Plans to process and share **70,000+** additional datasets
- Publicly accessible, no UT affiliation required



The screenshot shows the Texas GeoData Portal website. At the top, there is an orange header with the University of Texas at Austin logo and the text "GeoData University of Texas Libraries". To the right of the header are links for "About", "Help", and "Contact". Below the header, the main content area features the text "Explore geospatial data from UT Austin and beyond" and a search bar with "Search" and "Browse All" buttons. The main content area is divided into three sections: "Scanned Maps", "Geospatial Data", and "Alexander Architectural Archives". The "Geospatial Data" section is active, showing a map of Austin, TX with red dots indicating data points. To the left of the map, there are two filters: "Place" and "Subject". The "Place" filter shows a list of locations: Austin, TX; Texas; United States; North America; Canada; Mexico; Europe; and China. The "Subject" filter shows a list of subjects: Boundaries; Transportation; Imagery & Base Maps; Topographic Maps; Society; Environment; Land Use; and Population. Below the map, the URL <https://geodata.lib.utexas.edu> is displayed.

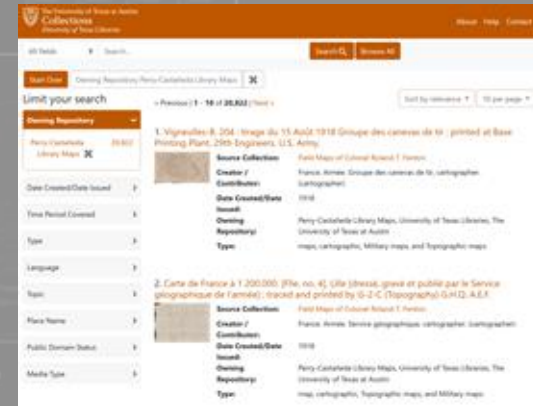
- **Capabilities:** browse, search, download data, access metadata, map preview

Automated Georeferencing Project Background

PCL Map Collection

The UT Libraries manages a growing collection of over **70,000** scanned map images that need expanded metadata and georeferencing to support discovery and use in research

Of these maps **20,822** have detailed metadata and are shared in the Collections Portal



Automated Georeferencing Project Background

Motivation

Prior to the start of this project only about **800** maps had been georeferenced and published to UT Austin's Texas GeoData Portal

Importance of georeferencing



Not georeferenced

No geographic coordinate info



Georeferenced

3+ map points with coordinate info

Size, rotation, and skew calculated

Challenges of Georeferencing a Large Volume of Maps



Photo by [Ocean Ng](#) on [Unsplash](#)

Key Challenge: Efficiently Annotating Scanned Map Images

If it takes **~10 minutes** to manually georeference one map...

It would take almost **6 years** for a single person working **40-hour weeks** to process all **70,000+ maps**

Sanborn Fire Insurance Maps

Documents urban development at a key point in time for change

Layout, construction materials, and building use is supplied in unprecedented detail

Problem: No coordinate data is available on the maps for simple text recognition



UT Sanborn Fire Insurance Map Collection

The screenshot shows the University of Texas at Austin Collections website. The search bar contains the term 'sanborn'. The search results are displayed in a list format. The first result is '1. Sanborn Fire Insurance Maps [San Antonio, Texas, 1896; sheet 48]'. The second result is '2. Sanborn Fire Insurance Maps [Elgin, Texas, 1905, Sheet 1]'. The website header includes the University of Texas at Austin logo and navigation links for 'About', 'Help', and 'Contact'. The search bar also includes a 'Search' button and a 'Browse All' button. The search results are sorted by relevance and show 10 items per page.

The University of Texas at Austin
Collections
University of Texas Libraries

About Help Contact

All fields | sanborn | Search | Browse All

Start Over | sanborn | X

Limit your search

« Previous | 1 - 10 of 14,095 | Next »

Sort by relevance | 10 per page

Owning Repository

Date Created/Date Issued

Time Period Covered

Type

Language

Topic

Place Name

Public Domain Status

1. Sanborn Fire Insurance Maps [San Antonio, Texas, 1896; sheet 48]

Source Collection: Sanborn Fire Insurance Maps [San Antonio, Texas, 1896]
Creator / Contributor: Sanborn Map Company. (cartographer)
Date Created/Date Issued: 1896
Owning Repository: Perry-Castañeda Library Maps, University of Texas Libraries, The University of Texas at Austin
Type: maps, cartographic, Fire Insurance maps, and Digital maps

2. Sanborn Fire Insurance Maps [Elgin, Texas, 1905, Sheet 1]

Source Collection: Sanborn Fire Insurance Maps [Elgin, Texas, 1905]
Creator / Contributor: Sanborn Map Company (cartographer)
Date Created/Date Issued: 1905

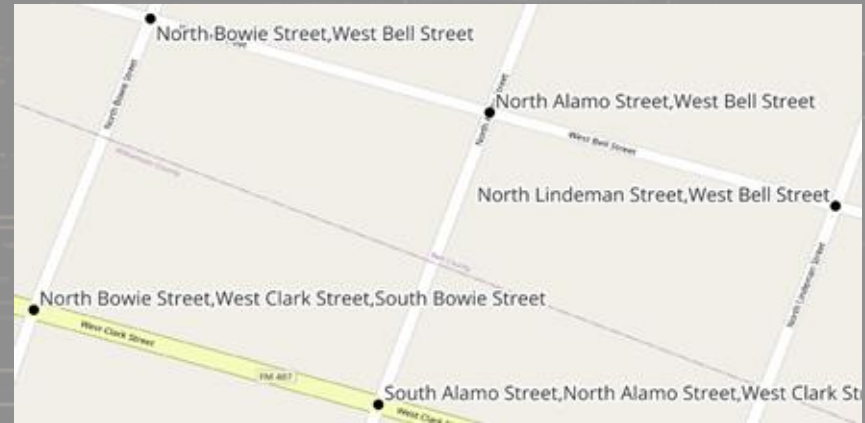
- 14,002 Sanborn Map Sheets
- 13,968 Sanborn sheets of Texas cities
- Maps range from 1877 to 1951
- Maps of Austin were manually georeferenced in 2019

Plan for Automating Georeferencing

Intersection data with
known **image**
coordinates



Intersection data with
known **geographic**
coordinates



Can we automate map georeferencing by matching objects (street intersections) with identified **image coordinates** to the same objects in an open geospatial dataset which contains their **geographic coordinates**?

Creating Map Annotations for Training Data

Custom annotation tool


- Configurable
- Support for multiple annotation types
- Easy to use
- Quick and responsive
- Support for wide variety of maps

ML4GIS Project



```
"defaultingData": {
  "src": "https://maps.lib.utexas.edu/maps/sanborn/tau-sanborn-bvltex-1907-d.tne",
  "name": "Image 1",
  "regions": [
    {
      "type": "point",
      "x": 0.3409412061769796,
      "y": 0.6672774098845257,
      "highlighted": false,
      "editingLabels": false,
      "color": "#ff0000",
      "id": "5088142262307938",
      "tags": ["street intersection"],
      "comment": "LINDEN & WALL",
    },
    {
      "type": "point",
      "x": 0.3409412061769796,
      "y": 0.426079964349943,
      "highlighted": false,
      "editingLabels": false,
      "color": "#ff0000",
      "id": "26823979387294881",
      "tags": ["street intersection"],
      "comment": "Wall & Monroe",
    },
    {
      "type": "point",
      "x": 0.8165553636707523,
      "y": 0.6248199887023542,
      "highlighted": false,
      "editingLabels": false,
      "color": "#ff0000",
      "id": "1998095312587281",
      "tags": ["street intersection"],
      "comment": "TAYLOR & MAIN",
    },
    {
      "type": "point",
      "x": 0.8188641702605279,
      "y": 0.34837369967173145,
      "highlighted": true,
      "editingLabels": false,
      "color": "#ff0000",
      "id": "5557924680961092",
      "tags": ["street intersection"],
      "comment": "DENNIS & MAIN",
    },
    {
      "type": "point",
      "x": 1.2078988886577158,
      "y": 0.47006554766866285,
      "highlighted": false,
      "editingLabels": false,
      "color": "#ff0000",
      "id": "09976136856768389",
    }
  ]
}
```

Machine Learning for Object Detection



```
custom-object-detection-TF2.ipynb ☆
File Edit View Insert Runtime Tools Help Last edited on September
+ Code + Text
2
3 # clone the tensorflow models on the colab cloud vm
4 !git clone --q https://github.com/tensorflow/models.git
5 # navigate to /models/research folder to compile protos
6 %cd models/research
7 # Compile protos.
8 !protoc object_detection/protos/*.proto --python_out=.
9 # Install TensorFlow Object Detection API.
10 !cp object_detection/packages/tf2/setup.py .
11 !python -m pip install .

[ ] 1 # 7. test model builder
2
3 !python object_detection/builders/model_builder_tf2_test.py

[ ] 1 %cd /content/drive/MyDrive/customTF2/data

/content/drive/MyDrive/customTF2/data

[ ] 1 !pwd

/content/drive/MyDrive/customTF2/data

[ ] 1 # create test and train labels
2
3
4 #creating two dir for training and testing
5 !mkdir test_labels train_labels
6
```

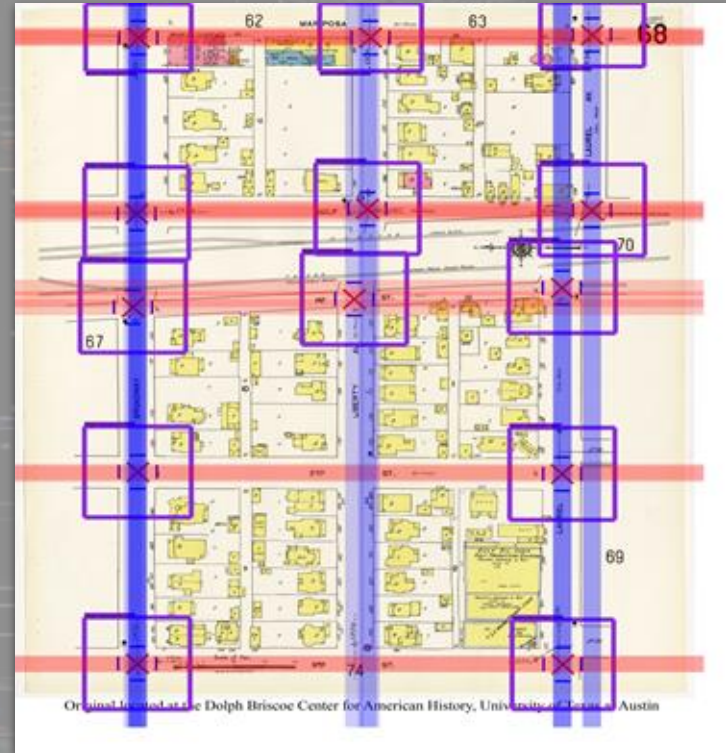
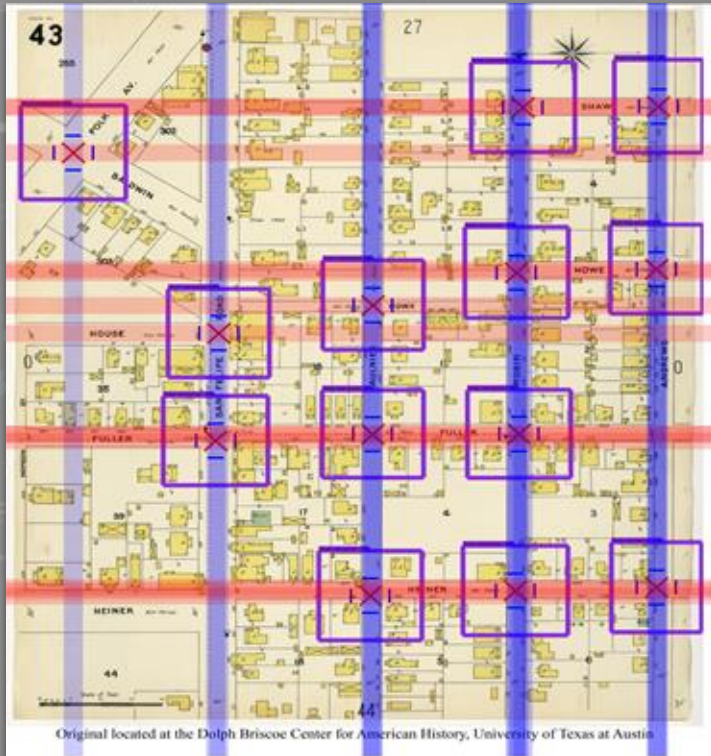
- A Google Colab notebook was created to facilitate creation and sharing of Python code for training an object detection model
- The Pro tier of Google Colab was necessary to obtain sufficient computational resources for training the model
- The TensorFlow package was used to train the model using street intersection annotation data for 200 Sanborn maps

Object Detection Model Output: Street Intersection Centroid Image Coordinates

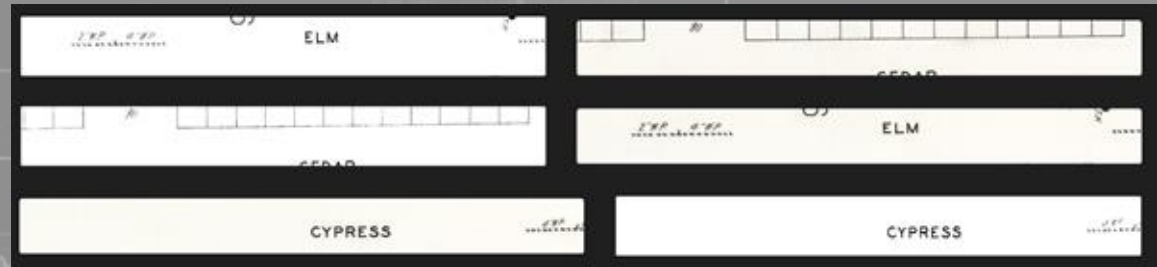
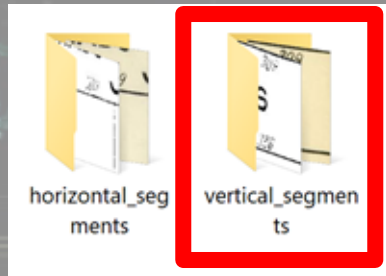
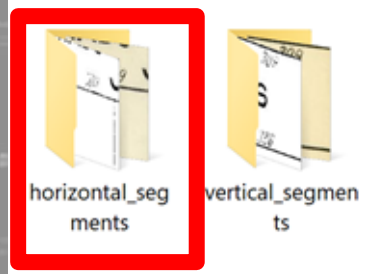
Name	Date modified	Type	Size
bartlett_1900_1_annotation_centroids	6/26/2023 5:47 PM	JSON File	1 KB
bartlett_1905_1_annotation_centroids	6/26/2023 5:47 PM	JSON File	1 KB
bartlett_1905_2_annotation_centroids	6/26/2023 5:47 PM	JSON File	1 KB
bartlett_1912_1_annotation_centroids	6/26/2023 5:47 PM	JSON File	1 KB
bartlett_1912_2_annotation_centroids	6/26/2023 5:48 PM	JSON File	1 KB
bartlett_1912_3_annotation_centroids	6/26/2023 5:48 PM	JSON File	1 KB
bartlett_1912_4_annotation_centroids	6/26/2023 5:48 PM	JSON File	1 KB
bartlett_1912_5_annotation_centroids	6/26/2023 5:48 PM	JSON File	1 KB
bartlett_1912_6_annotation_centroids	6/26/2023 5:48 PM	JSON File	1 KB
bartlett_1921_1_annotation_centroids	6/26/2023 5:48 PM	JSON File	1 KB
bartlett_1921_2_annotation_centroids	6/26/2023 5:48 PM	JSON File	1 KB
bartlett_1921_3_annotation_centroids	6/26/2023 5:48 PM	JSON File	1 KB
bartlett_1921_4_annotation_centroids	6/26/2023 5:45 PM	JSON File	1 KB
bartlett_1921_5_annotation_centroids	6/26/2023 5:49 PM	JSON File	1 KB
bartlett_1921_6_annotation_centroids	6/26/2023 5:49 PM	JSON File	1 KB
bartlett_1921_7_annotation_centroids	6/26/2023 5:49 PM	JSON File	1 KB
bartlett_1921_8_annotation_centroids	6/26/2023 5:49 PM	JSON File	1 KB
bartlett_1921_9_annotation_centroids	6/26/2023 5:49 PM	JSON File	1 KB
tru-sanborn-asherton-1951-1_annotation...	6/23/2023 12:28 AM	JSON File	1 KB
tru-sanborn-asherton-1951-2_annotation...	6/23/2023 12:28 AM	JSON File	1 KB
tru-sanborn-asherton-1951-3_annotation...	6/23/2023 12:29 AM	JSON File	1 KB
tru-sanborn-asherton-1951-4_annotation...	6/23/2023 12:29 AM	JSON File	1 KB
tru-sanborn-bandera-1950-1_annotation...	6/23/2023 12:30 AM	JSON File	1 KB
tru-sanborn-bandera-1950-2_annotation...	6/23/2023 12:30 AM	JSON File	1 KB
tru-sanborn-bandera-1950-3_annotation...	6/23/2023 12:30 AM	JSON File	1 KB
tru-sanborn-bandera-1950-4_annotation...	6/23/2023 12:31 AM	JSON File	1 KB
tru-sanborn-bandera-1950-5_annotation...	6/23/2023 12:31 AM	JSON File	1 KB

```
1 {  
2   "annotationcentroids": [ [1992, 1050],  
3     [186, 2134],  
4     [177, 1042],  
5     [1994, 2095],  
6     [1993, 148],  
7     [1996, 3367],  
8     [1487, 152],  
9     [183, 2929]  
10 ]  
11 }  
12
```

Visualization of Street Intersection Detection



Cropping and OCRing Map Images

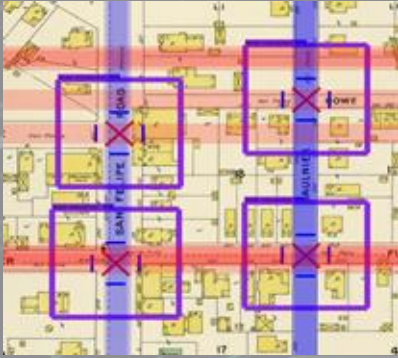


OCR carried out with Tesseract OCR Engine and Pytesseract Python package

→ <https://tesseract-ocr.github.io>

→ <https://pypi.org/project/pytesseract/>

Geolocated Intersection Data Saved as JSON



+

Intersection data with
known geographic
coordinates

=

```
51 {  
52   "mapfilename": "bartlett_1921_4.jpg",  
53   "georeferencedintersections": [  
54     {  
55       "city": "Bartlett",  
56       "streets": ["alamo", "davilla"],  
57       "image_x": 1831,  
58       "image_y": 1195,  
59       "geo_x": -97.432719999999996,  
60       "geo_y": 30.7988610000000045  
61     },  
62     {  
63       "city": "Bartlett",  
64       "streets": ["lucy", "bell"],  
65       "image_x": 2810,  
66       "image_y": 1693,  
67       "geo_x": -97.430796999999998,  
68       "geo_y": 30.7972360000000055  
69     }  
51 }
```

Map points with
known coordinates



SP	ST	SP	ST
SP	ST	SP	ST
SP	ST	SP	ST
SP	ST	SP	ST

ELM	ELM
ELM	ELM
CYPRESS	CYPRESS
CYPRESS	CYPRESS

Georeferencing with Geolocated Image Coordinates



```
import time as t
label_map_util = tf.nn.conv2d
fft2 = tf.nn.conv2d
import numpy as np
from PIL import Image
from PIL import ImageEnhance
from object_detection.utils import visualization_utils as viz_utils
from object_detection.utils import label_map_util
import os
import tensorflow
import matplotlib.pyplot as plt
import re
import geopandas as gpd
from shapely.geometry import mapping
from itertools import combinations
import time
import cv2
import json
import warnings
warnings.filterwarnings('ignore')
import time
import math
from datetime import datetime

et = time.time()
timeprlist = [time.time()]
logfile = datetime.now().strftime("%Y_%m_%d_%H_%M_%S") + "_log.txt"
logfilepath = "logs/" + logfile

def writeLogMessage_log(filepath="logs/" + datetime.now().strftime("%Y_%m_%d_%H_%M_%S") + "_log.txt"):
    processedLines = []
    ct = time.time()
    totalTime = ct-et
    stoptime = ct-timeprlist[-1]
    timeprlist.append(ct)
    timeLogProcess = [stoptime, totalTime]
```

Name	Date	Type	Size
bartlett_1900_1_georeferenced	8/7/2008 3:17 PM	TIF File	32,037 KB
bartlett_1905_1_georeferenced	8/7/2008 3:17 PM	TIF File	32,108 KB
bartlett_1912_1_georeferenced	8/7/2008 3:20 PM	TIF File	31,976 KB
bartlett_1912_2_georeferenced	8/7/2008 3:21 PM	TIF File	31,844 KB
bartlett_1912_3_georeferenced	8/7/2008 3:22 PM	TIF File	32,020 KB
bartlett_1912_4_georeferenced	8/7/2008 3:22 PM	TIF File	32,073 KB
bartlett_1921_2_georeferenced	8/19/2008 11:28 AM	TIF File	31,828 KB
bartlett_1921_3_georeferenced	8/19/2008 11:30 AM	TIF File	31,748 KB
bartlett_1921_4_georeferenced	8/19/2008 11:31 AM	TIF File	31,775 KB
bartlett_1921_7_georeferenced	8/7/2008 3:28 PM	TIF File	31,562 KB
bartlett_1921_8_georeferenced	8/7/2008 3:29 PM	TIF File	32,020 KB
txu-sanborn-goliad-1906-2_georeferenced	2/17/2010 1:13 PM	TIF File	41,437 KB
txu-sanborn-goliad-1912-5_georeferenced	2/17/2010 1:16 PM	TIF File	40,909 KB
txu-sanborn-goliad-1922-2_georeferenced	2/17/2010 1:17 PM	TIF File	40,809 KB
txu-sanborn-goliad-1922-3_georeferenced	2/17/2010 1:17 PM	TIF File	40,819 KB
txu-sanborn-goliad-1922-4_georeferenced	2/17/2010 1:17 PM	TIF File	41,019 KB
txu-sanborn-gorman-1921-2_georeferenc...	2/17/2010 1:45 PM	TIF File	41,049 KB
txu-sanborn-gorman-1921-3_georeferenc...	2/17/2010 1:45 PM	TIF File	41,059 KB
txu-sanborn-granger-1921-2_georeferenc...	2/19/2010 3:15 PM	TIF File	41,049 KB
txu-sanborn-granger-1921-6_georeferenc...	2/19/2010 3:16 PM	TIF File	41,109 KB
txu-sanborn-hico-1907-3_georeferenced	3/9/2010 11:54 AM	TIF File	41,098 KB
txu-sanborn-kosse-1899-1_georeferenced	3/15/2010 5:09 PM	TIF File	41,208 KB
txu-sanborn-kosse-1905-1_georeferenced	3/15/2010 5:10 PM	TIF File	40,859 KB
txu-sanborn-kosse-1911-1_georeferenced	3/15/2010 5:10 PM	TIF File	41,009 KB
txu-sanborn-kosse-1921-2_georeferenced	3/15/2010 5:11 PM	TIF File	41,288 KB

georeferencing-automator.py

Quality Assessment

- Convex Hull approach for measuring spatial distribution of ground control points
- Root mean square error for measuring the accuracy of ground control points



Results

- 13,830 Sanborn Fire Insurance maps of Texas processed
- Final process took about 1 week to complete
- Maps were processed in batches of several thousand at a time



Results

Accurately Georeferenced Maps (14.0%)



Distorted Georeferenced Maps (4.3%)



Maps That Could Not Be Georeferenced (81.7%)



Results

	All processed Texas Sanborn maps with generated accuracy reports	Georeferenced maps (both accurate and inaccurate)	Accurately georeferenced maps	Inaccurately georeferenced maps
Total number	13,830	2560	1962	598
Mean GCPs per map (median)	1.37 (0)	6.49 (5)	6.22 (5)	7.39 (4)
Mean GCP convex hull area to map coverage ratio (median)	N/A	.64 (.19)	.24 (.20)	1.97 (.06)
RMSE in feet (median)	N/A	18.43 (2.47)	2.77 (2.34)	69.81 (4.87)

Results

The University of Texas at Austin
GeoData
University of Texas Libraries

About Help Contact

The Texas GeoData Portal has recently been updated to GeoBlacklight 4 to provide new and improved functionality. As a result of this change, some records previously available from other institutions are not currently accessible through the Portal, but additional records will be actively added over time.

sanborn fire insurance map

sanborn fire insurance map Provider: Texas

Limit your search

« Previous | 1 - 10 of 1,856 | Next »

Sort by Relevance 10 per page

Provider

▼ Texas 1,856

Place

Format

Resource Class

Year

Language

Creator

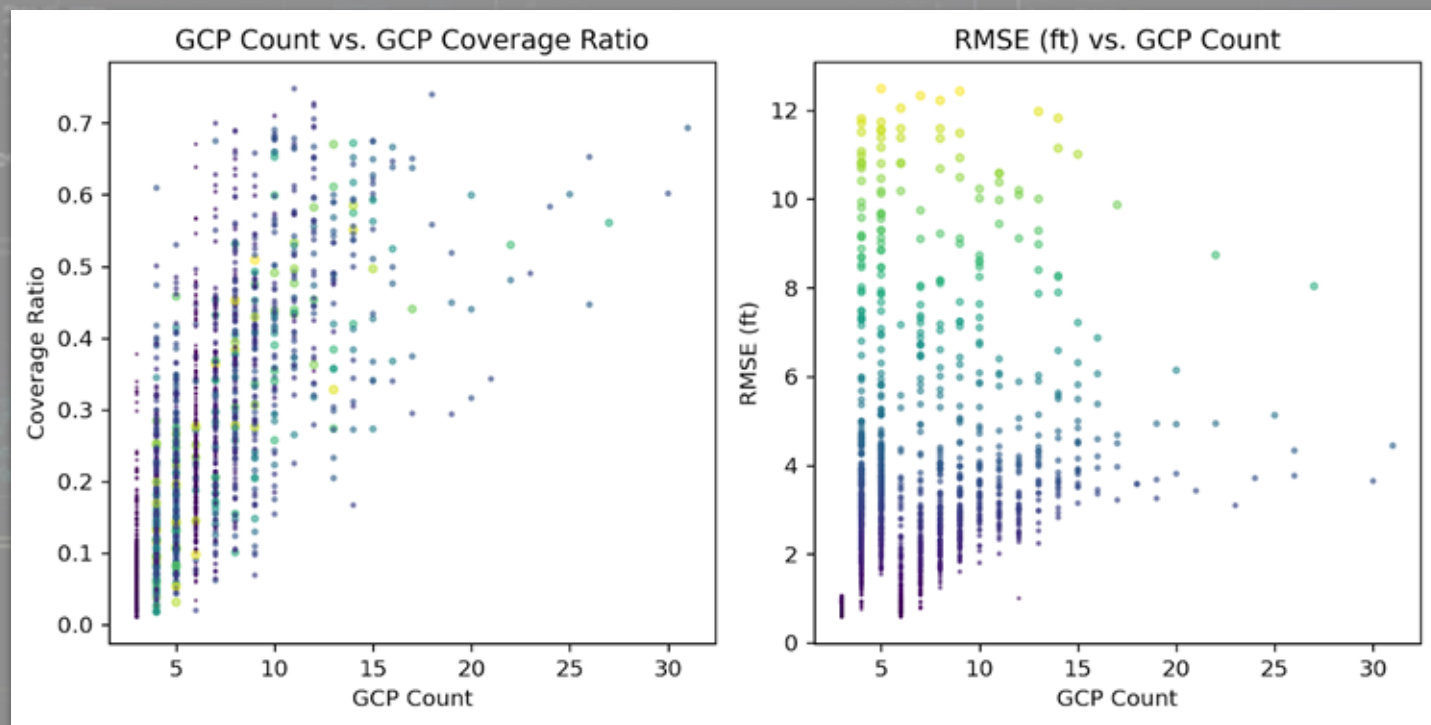
1. Sanborn Fire Insurance collection
2. Sanborn Fire Insurance Maps [Texas, Texas, 1885; Sheet Sanborn Fire Insurance Maps [Austin, Texas, 1885;]
3. Sanborn Fire Insurance Maps [Austin, Texas, 1921; Sheet 69]
4. Sanborn Fire Insurance Maps [Austin, Texas, 1921; Sheet 68]
5. Sanborn Fire Insurance Maps [Austin, Texas, 1885; Sheet 7]
6. Sanborn Fire Insurance Maps [Austin, Texas, 1889; Sheet 9]

Search when I move the map

AUSTIN

Leaflet | © OpenStreetMap contributors, © Carto

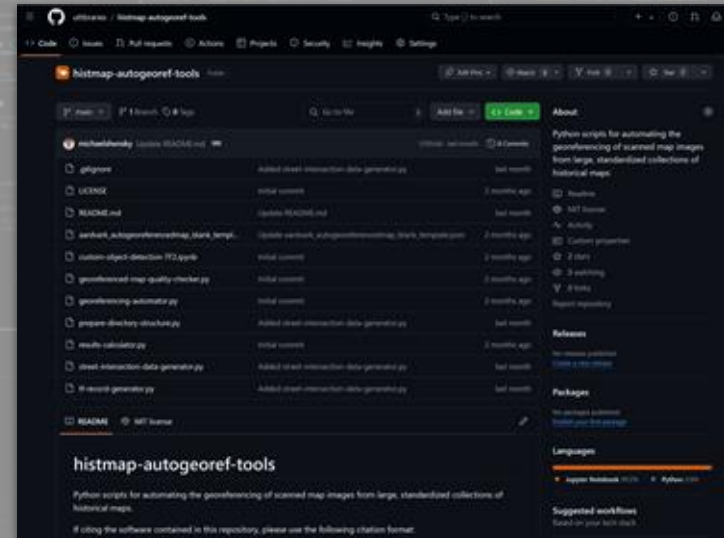
Results



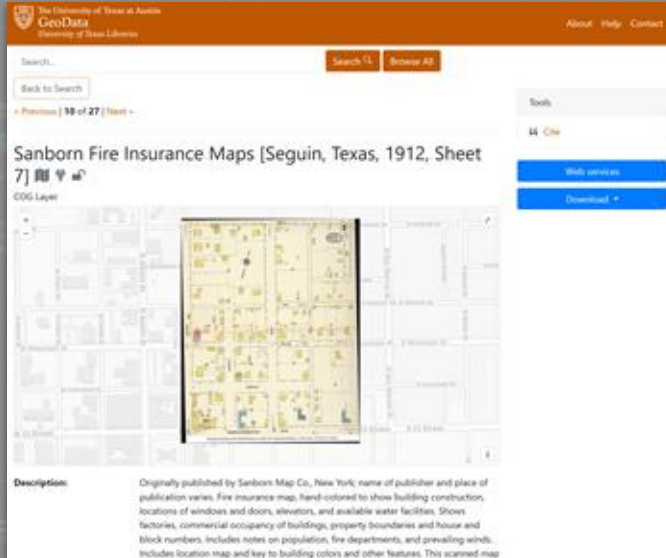
Code Shared on GitHub

- The code is available at

<https://github.com/utlibraries/histmap-autogeoref-tools>



Discussion



All processed maps added to the
Texas GeoData portal

<https://geodata.lib.utexas.edu/>

- Currently a proof of concept with limited but promising results
- Powerful tool for large collections of Sanborn Maps
 - Can be paired with other methods of efficient manual georeferencing
- Ground Control Point distribution is more impactful than number (assuming at least 3 GCPs)

Next Steps for Automated Georeferencing Workflow


1. Improve object detection model and street intersection detection rate
2. Improve OCR success rate for street labels
3. Automate detection and separation of inset maps



Next Steps for Automated Georeferencing Workflow

4. Test out related approach for georeferencing other map collections
5. Create vector datasets derived from georeferenced maps

```
1 | #display inference on All test images
2 |
3 | import glob
4 | from IPython.display import Image, display
5 |
6 | i = 0
7 | limit = 10000 # max images to print
8 | #for filepath in glob.glob('content/drive/MyDrive/maps_pilot/test_images_test*.jpg'): Assuming JPG
9 | #for filepath in glob.glob('content/yolov7/val/test_img1*.jpg'): Assuming JPG
10 | # i < limit:
11 |     display(Image(filename=filepath))
12 |     print("i")
13 |     i = i + 1
```



Moving forward

- Cross training and sharing of code essential for sustainability
- Size of collection is important for justifying investment of time
- Standardization of collection materials important for reliable processing
- New ground can still be broken using ML and AI to automate complex processes

GIS-Ready Historical Texas Maps

- The Texas GeoData Portal now has 1,798 georeferenced Sanborn Map Sheets from various locations across Texas
 - The map sheets are publicly available to download
- More sheets will be added as the automated georeferencing process is improved
- Availability of georeferenced Sanborn Map Sheets varies across locations

Demo

[About](#) [Help](#) [Contact](#)

Explore geospatial data from UT Austin and beyond

[Search](#)

[Browse All](#)

Place

Austin, TX	Canada
Texas	Mexico
United States	Europe
North America	China
more...	

Subject

Boundaries	Society
Transportation	Environment
Imagery & Base Maps	Land Use
Topographic Maps	Population
more...	

Acknowledgements

ML4GIS Annotation Tool Project Team

- Aaron Choate
- Dana Gurrari

Development of the ML4GIS annotation tool

- Alyssa Zhan
- Patrick Chao

Annotation creation

- Supreeya Karn

ML4GIS Project Partners at City of Austin and Austin History Center

- Ross Clark
- Jennifer Hecker
- Mike Miller

Project funding

- Good Systems, a research grand challenge at UT Austin

Questions?

Contact:

Michael Shensky (he/him)

The University of Texas at Austin

Head of Research Data Services | University of Texas
Libraries

m.shensky@austin.utexas.edu

Alex Marden (he/him)

The University of Texas at Austin

GIS and Geospatial Coordinator | University of Texas
Libraries

alex.marden@austin.utexas.edu



Texas GeoData Portal

