

ICPSR

SHARING DATA TO ADVANCE SCIENCE

Enabling Spatial Search and Geovisualization in Archonnex

Inter-university Consortium for Political and Social Research

Bing She

Application Programmer Staff Specialist

Presentation at Association of American Geographers

Annual Meeting 2019 in Washington, DC

Motivation

- Spatial search is more intuitive when users want studies of certain regions

The 1915 Iowa State Census Project (ICPSR 28501)

Published: Dec 14, 2010 [Cite this Study](#) | [Share this study](#)

Principal Investigator(s): [?](#)

[Claudia Goldin](#), Harvard University, and National Bureau of Economic Research

Colonial Land Grants Database, Calhoun Critical Zone Observatory, (ICPSR 37078)

Published: Jun 22, 2018 [Cite this Study](#) | [Share this study](#)

Principal Investigator(s): [?](#)

[Michael Coughlan](#), University of Georgia; [Donald Nelson](#), University of Georgia

<https://doi.org/10.3886/ICPSR37078.v1>

Version V1

Download ▾

Analyze Online (0)

Documentation Only

GIS

Other

Documentation

Variables

Publications

Export Metadata



Status quo

The Geographic Coverage field in the study metadata represents the geography of a study

Issues:

- Not matched to the spatial data
- Does not conform to a standard and contain duplicates

Related Websites

The screenshot shows the Data.Gov website interface. At the top, there is a navigation bar with 'DATA', 'TOPICS', 'IMPACT', 'APPLICATIONS', and 'DEVELOPERS'. Below this is a sub-navigation bar with 'LOCAL', 'DATA CATALOG', and 'Organizations'. The main content area features a search bar with the text 'Search datasets...' and a filter for 'Local Government'. A search result is displayed for 'Colorado Thermal Springs', including a map of the state of Colorado and a brief description of the dataset.

Data.Gov

The screenshot shows the INSPIRE Geoportal website. The header includes the European Commission logo and the text 'INSPIRE GEOPORTAL Enhancing access to European spatial data'. Below the header is a navigation bar with 'Home', 'Priority Data Sets Viewer', 'INSPIRE Thematic Viewer', and 'Harvesting status'. The main content area features a map of Europe with a legend indicating 'INSPIRE Geoportal 137410 Metadata records'. A search bar is visible at the top right.

EU Inspire

The screenshot shows the Big Ten Academic Alliance Geoportal website. The header includes the 'BIG ACADEMIC ALLIANCE' logo and the text 'Geoportal'. Below the header is a navigation bar with 'Bookmarks (0)', 'History', 'About', 'Help', and 'Login'. The main content area features a search bar with the text 'Search' and a search result for '100 Year Flood'. The search results are displayed in a table with columns for 'Place', 'Genre', and 'Subject'. A map of the United States is visible on the right side of the page.

Big Ten Academic Alliance Geoportal
(One instance of GeoBlacklight)

Why not use existing solutions?

Existing solutions such as GeoBlacklight, GeoNode, GeoNetwork:

- Not directly applicable to ICPSR studies
- Hard to incorporate curation process
- Lack of Attribute-level metadata
- Not compatible with Archonnex architecture

Spatial pilot project at ICPSR

Goal: Enable spatial search and geovisualization functions in ICPSR's Digital Asset Management System (DAMS) - Archonnex

- Starts with those studies that have GIS data files and not restricted.
- Seven studies identified:
2895, 4546, 4547, 35617, 36353, 36745, 37078

Proposed workflows

- Publish spatial files to the RDBMS with spatial extension and the OGC-compliant server
- Spatial metadata edit for curation
- Spatial search and Geovisualization for end users

Technology stack

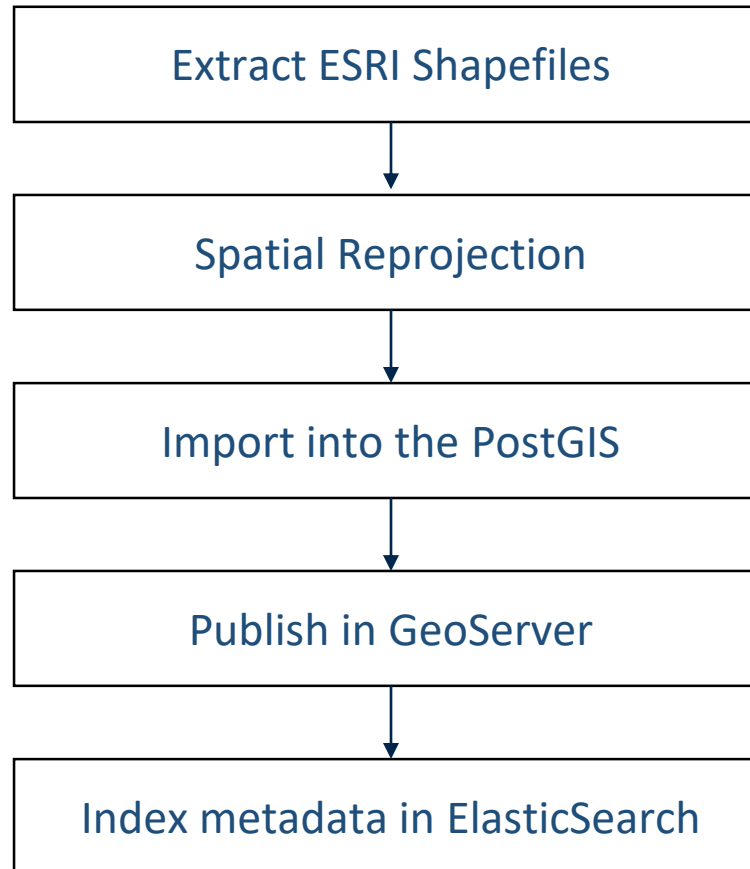
Consistent and compatible with the architecture of Archonnex

Frontend	React, Leaflet, Mapbox GL JS, Bootstrap
Backend	Spring MVC, Spring JPA with Hibernate, Spatial4J, JTS, Geotools, GeoNames
Database	PostgreSQL+PostGIS
Map server	GeoServer
Search engine	ElasticSearch

Key technical issues

- Map representation (vector tiles and raster tiles)
- Map operations (zoom, edit, envelope selection)
- Spatial data formats (shapefiles for publish, GeoJSON and WKB for transfer, WKB for database)
- Geographical coverage extraction (simple heuristic algorithm using overlapping ratio)













Publish spatial files



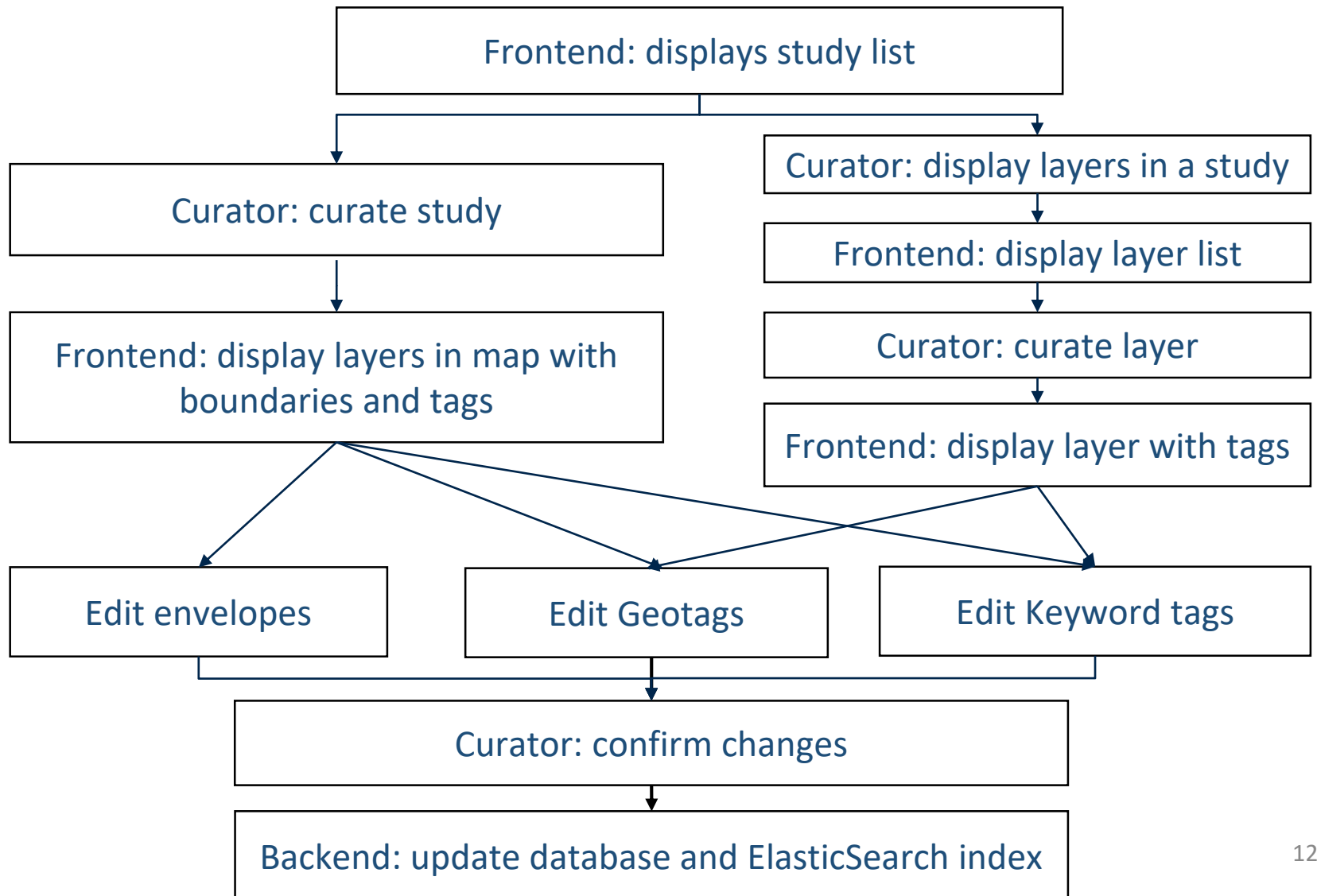
For simplicity, this pilot project assumes that the data won't change. The system will bulk import and publish the data at setup phase through API calls. The ElasticSearch index will also be prebuilt

Publish spatial files

140 layers added to PostGreSQL and Geoserver

<input type="checkbox"/>	Type	Title	Name	Store	Enabled	Native SRS
<input type="checkbox"/>		icpsr02895flgacount	icpsr:icpsr02895flgacount	postgis	✓	EPSG:4326
<input type="checkbox"/>		icpsr02895fstates	icpsr:icpsr02895fstates	postgis	✓	EPSG:4326
<input type="checkbox"/>		icpsr02895nmdvcount	icpsr:icpsr02895nmdvcount	postgis	✓	EPSG:4326
<input type="checkbox"/>		icpsr02895ohcounty	icpsr:icpsr02895ohcounty	postgis	✓	EPSG:4326
<input type="checkbox"/>		icpsr02895ohio	icpsr:icpsr02895ohio	postgis	✓	EPSG:4326
<input type="checkbox"/>		icpsr02895pacounty	icpsr:icpsr02895pacounty	postgis	✓	EPSG:4326
<input type="checkbox"/>		icpsr02895pazip	icpsr:icpsr02895pazip	postgis	✓	EPSG:4326
<input type="checkbox"/>		icpsr02895penn	icpsr:icpsr02895penn	postgis	✓	EPSG:4326
<input type="checkbox"/>		icpsr02895powstates	icpsr:icpsr02895powstates	postgis	✓	EPSG:4326
<input type="checkbox"/>		icpsr02895scnccount	icpsr:icpsr02895scnccount	postgis	✓	EPSG:4326
<input type="checkbox"/>		icpsr02895statereg	icpsr:icpsr02895statereg	postgis	✓	EPSG:4326
<input type="checkbox"/>		icpsr02895us_states	icpsr:icpsr02895us_states	postgis	✓	EPSG:4326
<input type="checkbox"/>		icpsr02895uspgh95	icpsr:icpsr02895uspgh95	postgis	✓	EPSG:4326
<input type="checkbox"/>		icpsr02895westv	icpsr:icpsr02895westv	postgis	✓	EPSG:4326
<input type="checkbox"/>		icpsr02895wvcounty	icpsr:icpsr02895wvcounty	postgis	✓	EPSG:4326
<input type="checkbox"/>		icpsr04546ucnij	icpsr:icpsr04546ucnij	postgis	✓	EPSG:4326

Spatial metadata edit



Spatial metadata edit

- Edit geographical coverage

ICPSR Curate Find Data by Studies Find Data by Layers Sign In

Spatial curation for: Examination of Crime Guns and Homicide in Pittsburgh, Pennsylvania, 1987-1998

Edit Geographical Coverage

Overlapping ratio: 0.3

The geographical coverage of this study has already been set. You may continue to edit the shapes

Pittsburgh, US (40.44062, -79.99589)

crime

Select Layers to Display

All None

- flgacount
- fstates
- nmdvcount

Spatial metadata edit

- Geotagging
- Keyword tagging

pittsburgh

Close

Pittsburgh, US (40.44062, -79.99589)

Pittsburgh International Airport, US (40.49608, -80.25547)

University of Pittsburgh, US (40.44285, -79.95922)

East Pittsburgh, US (40.39562, -79.83866)

Pittsburg, US (37.41088, -94.70496)

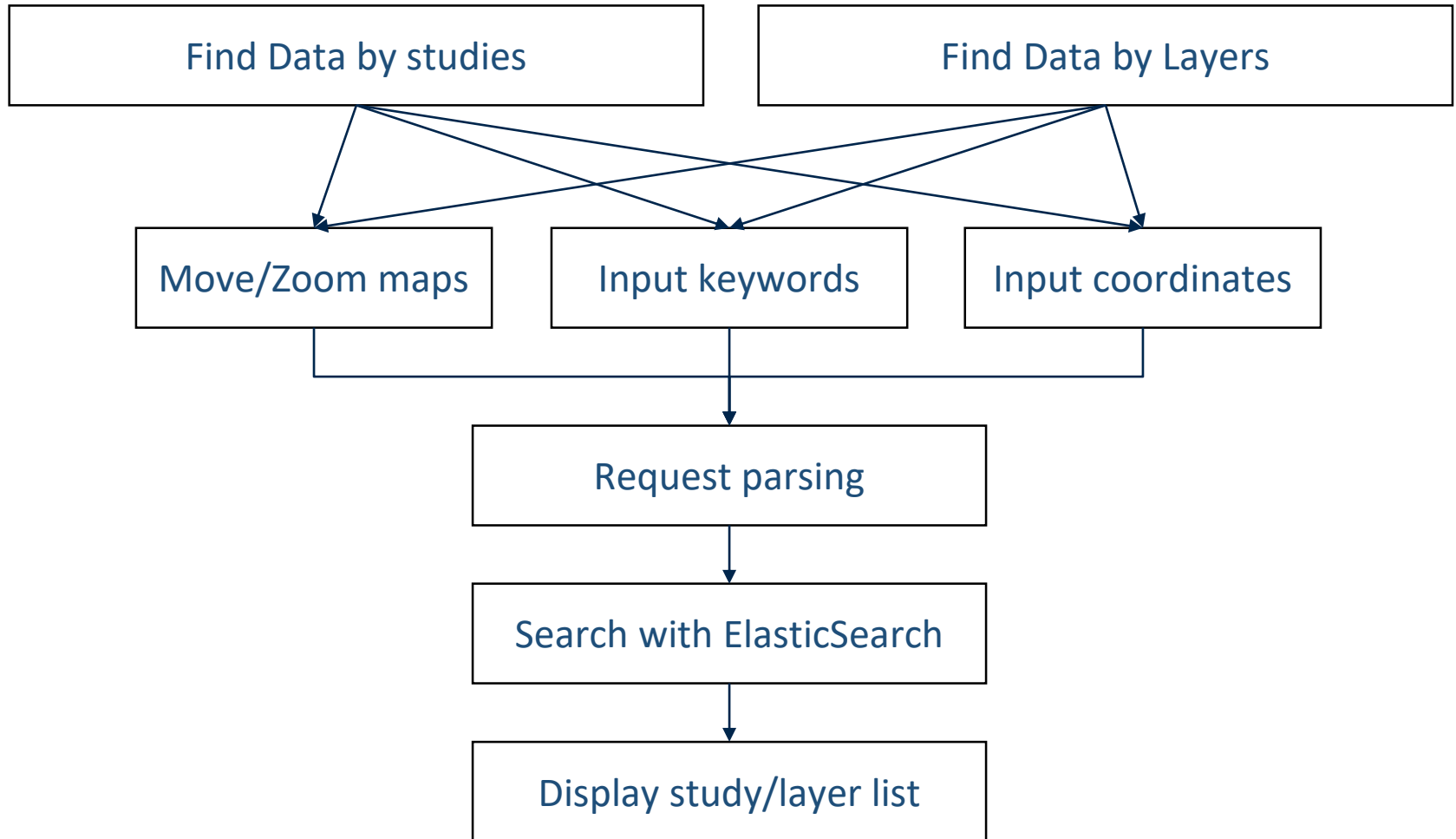
Pittsburgh, CA (44.38342, -76.33269)

ICPSR Curate Find Data by Studies Find Data by Layers

GeoNames information for Pittsburgh

Attribute	Value
GeoNames ID	5206379
Name	Pittsburgh
Alternate Names	
Continent Code	NA
Country Code	US
Country Name	
Population	304391
Feature Class Name	city, village,...
Feature Code	PPLA2
Feature Code Name	seat of a second-order administrative division
Location	(40.44062, -79.99589)
Admin Code 1	PA
Admin Name 1	Pennsylvania
Admin Code 2	003
Admin Name 2	Allegheny
Admin Code 3	61000
Admin Name 3	City of Pittsburgh

Spatial search



Spatial search

- Text + Map + Coordinate

ICPSR

Curate

Find Data by Studies

Find Data by Layers

Sign In

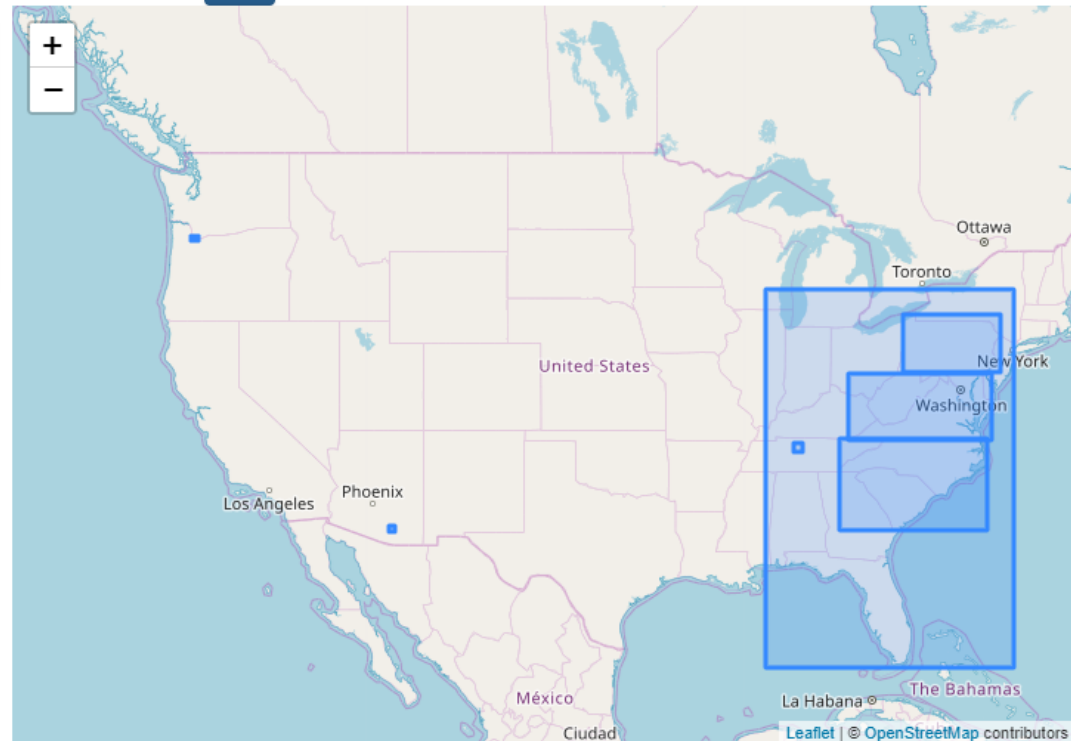
crime

GO

1. Geographies of Urban Crime in Nashville, Tennessee, Portland, Oregon, and Tucson, Arizona, 1998-2002

2. Examination of Crime Guns and Homicide in Pittsburgh, Pennsylvania, 1987-1998

3. Exploratory Spatial Data Approach to Identify the Context of Unemployment-Crime Linkages in Virginia, 1995-2000



Study view

- Vector tiles + Thematic mapping

ICPSR

Curate

Find Data by Studies

Find Data by Layers

Sign In

Examination of Crime Guns and Homicide in Pittsburgh, Pennsylvania, 1987-1998

This study examined spatial and temporal features of crime guns in Pittsburgh, Pennsylvania, in order to ascertain how gun availability affected criminal behavior among youth, whether the effects differed between young adults and juveniles, and whether that relationship changed over time. Rather than investigating the general prevalence of guns, this study focused only on those firearms used in the commission of crimes. Crime guns were defined specifically as those used in murders, assaults, robberies, weapons offenses, and drug offenses. The emphasis of the project was on the attributes of crime guns and those who possess them, the geographic sources of those guns, the distribution of crime guns over neighborhoods in a city, and the relationship between the prevalence of crime guns and the incidence of homicide.

United States, Pennsylvania, Pittsburgh

Select Layers to Display

All None

figacount

fstates

nmdvcount

ohcounty

ohio

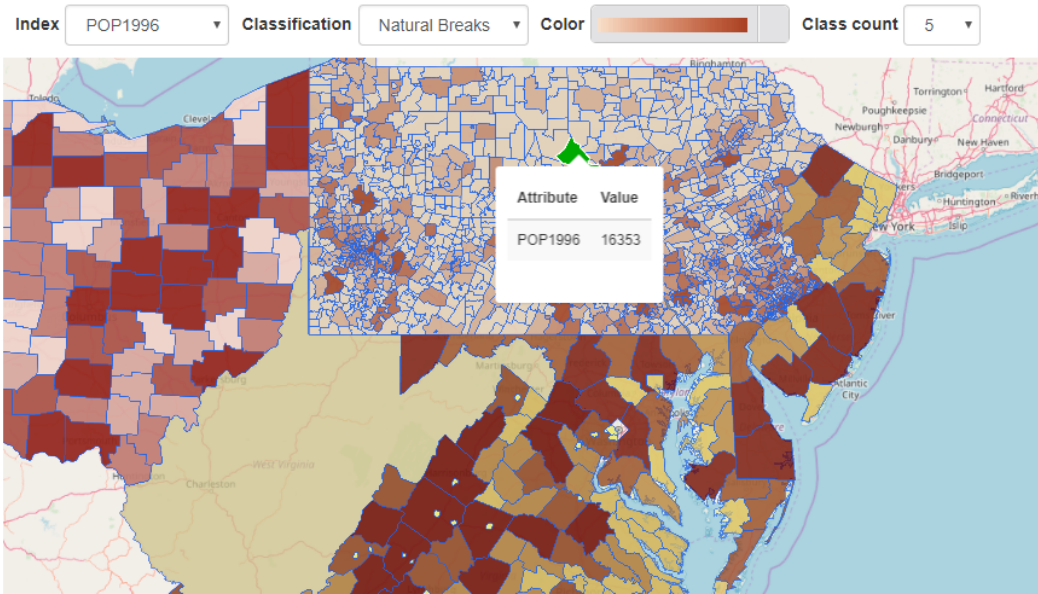
pacounty

pazip

penn

powstates

scnccount



Study view

ICPSR

Curate

Find Data by Studies

Find Data by Layers

Sign In

Historical Transportation of Navigable Rivers, Canals, and Railroads in the United States

This collection contains GIS materials which cover the spread of different modes of transportation in the lower 48 states from America's founding through (approximately) 1911. There are three transportation modes included in this collection: canals, steamboat-navigated (as opposed to simply navigable) rivers, and railroads.

The GIS materials can be downloaded by accessing the "Other" link.

United States, 1780 -- 1920

Select Layers to Display

All

None

RR1826_1911Modified050916

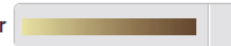
SteamboatNavigatedRiverConti

T19thC_Canals

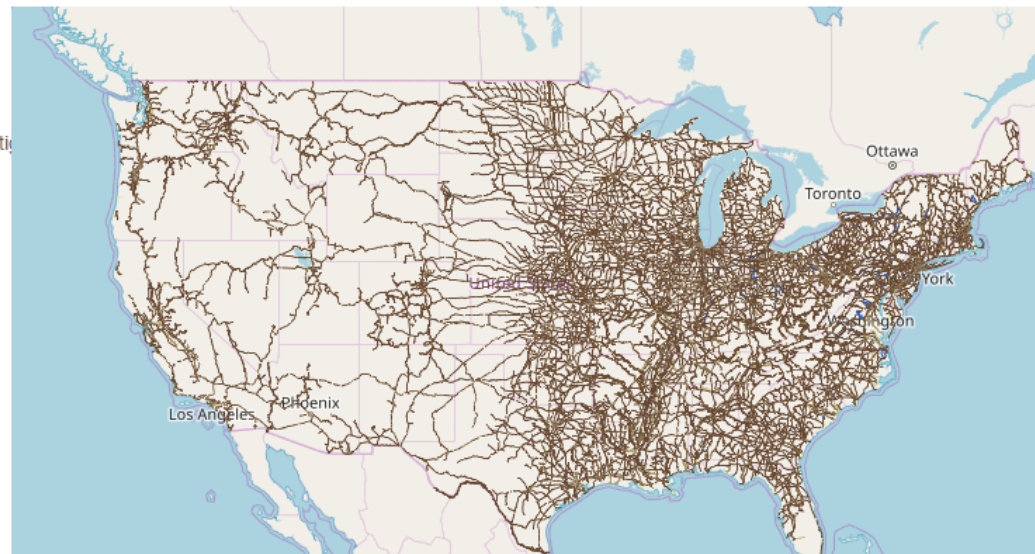
Index VxCount

Classification Quantile

Color



Class count 5



Challenges

- Data cleaning (e.g. incorrect or missing projection)
- Large spatial data
- Extract geographical coverage from text
- Standards for spatial information that fit social science data

Future works

- Variable metadata editor
- Generate spatial data from existing studies that have spatial information in textual or numerical formats
- Handle big spatial data (e.g. taxi trajectories)
- Online spatial data analysis