ICPSR

SHARING DATA TO ADVANCE SCIENCE

Enabling Spatial Search and Geovisualization in Archonnex

Inter-university Consortium for Political and Social Research

Bing She

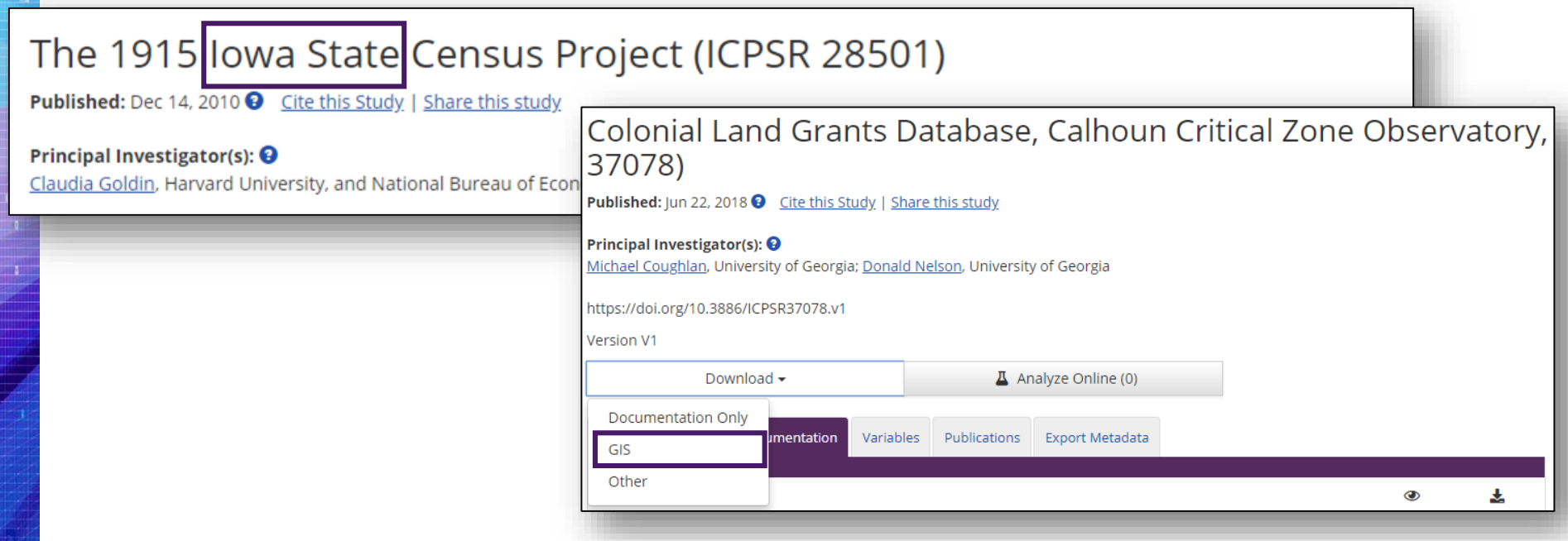
Application Programmer Staff Specialist

Presentation at Association of American Geographers Annual Meeting 2019 in Washington, DC

ICPSR INSTITUTE FOR SOCIAL RESEARCH UNIVERSITY OF MICHIGAN

Motivation

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



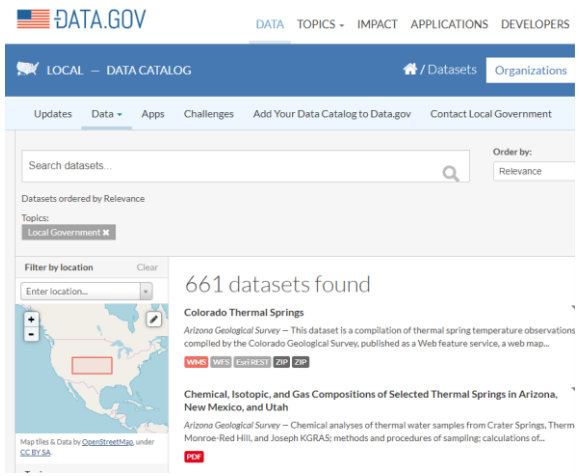
Status quo

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

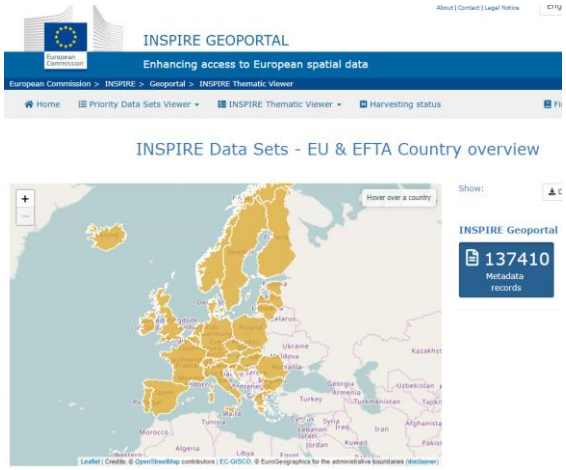
***Issues**:*

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

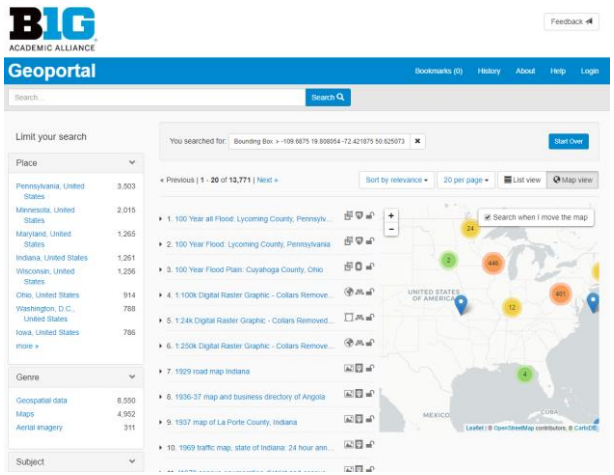
Related Websites



Data.Gov



EU Inspire



Big Ten Academic Alliance Geoportal
(One instance of GeoBlacklight)

Why not use existing solutions?

Existing solutions such as GeoBlacklight,

GeoNode, GeoNetwork:

- Not directly appliable to ICPSR studies
- Hard to incorprate 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

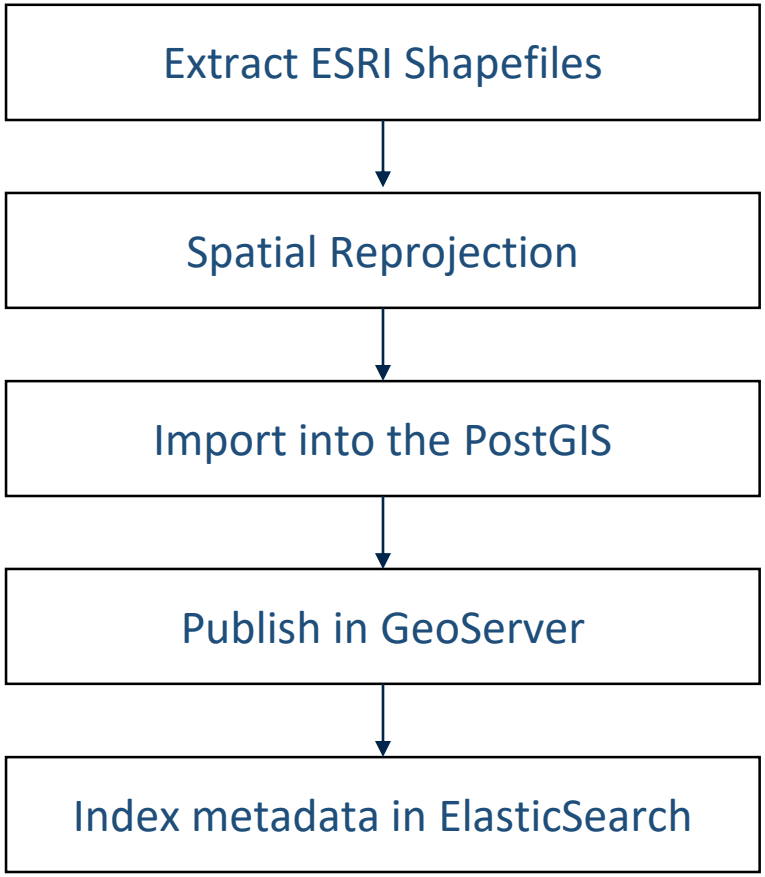
Consistent and compatible with the architecture of Archonnex

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 ElaticSearch index will also be prebuilt

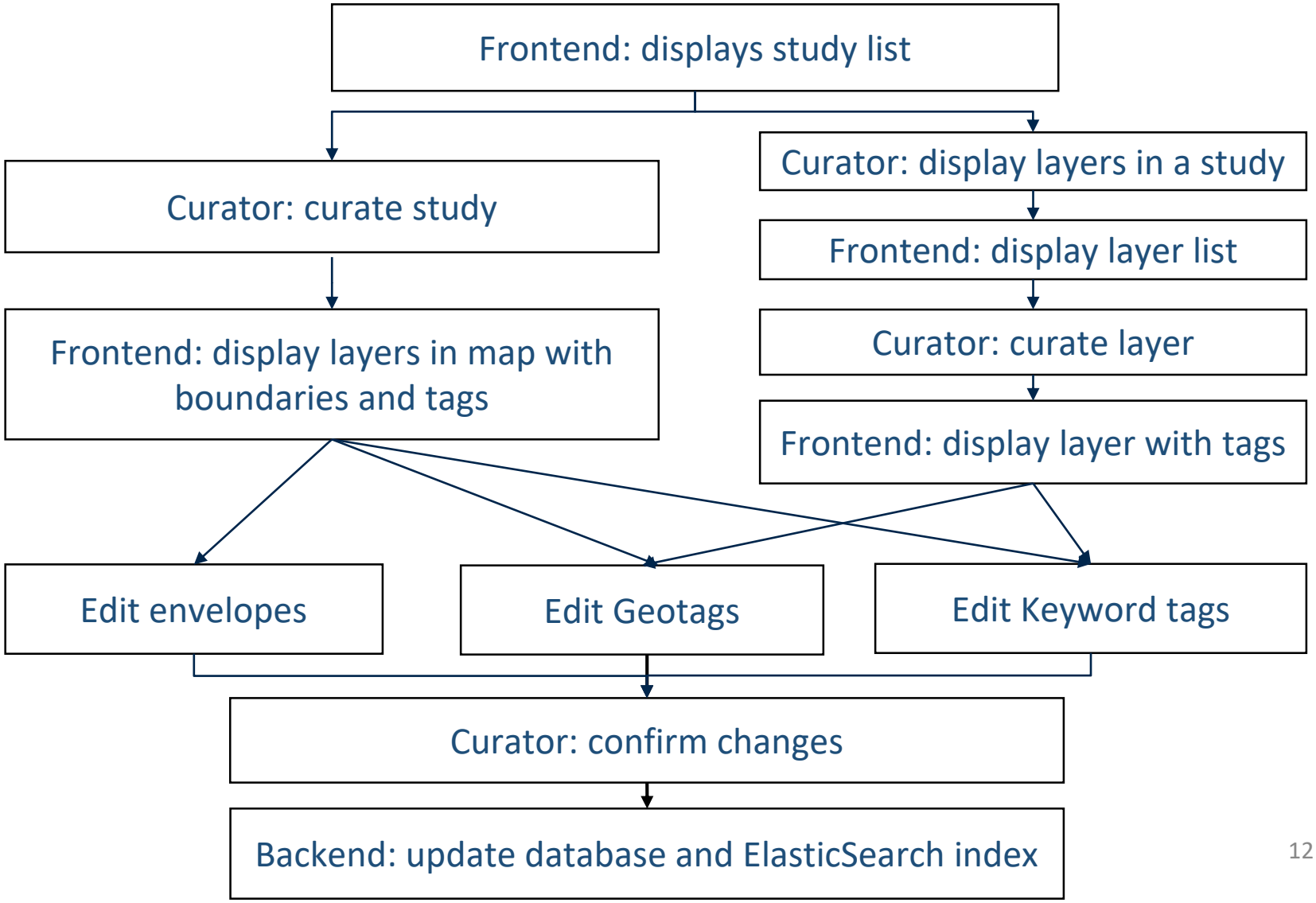
Publish spatial files

140 layers added to PostGreSQL and Geoserver

icpsr

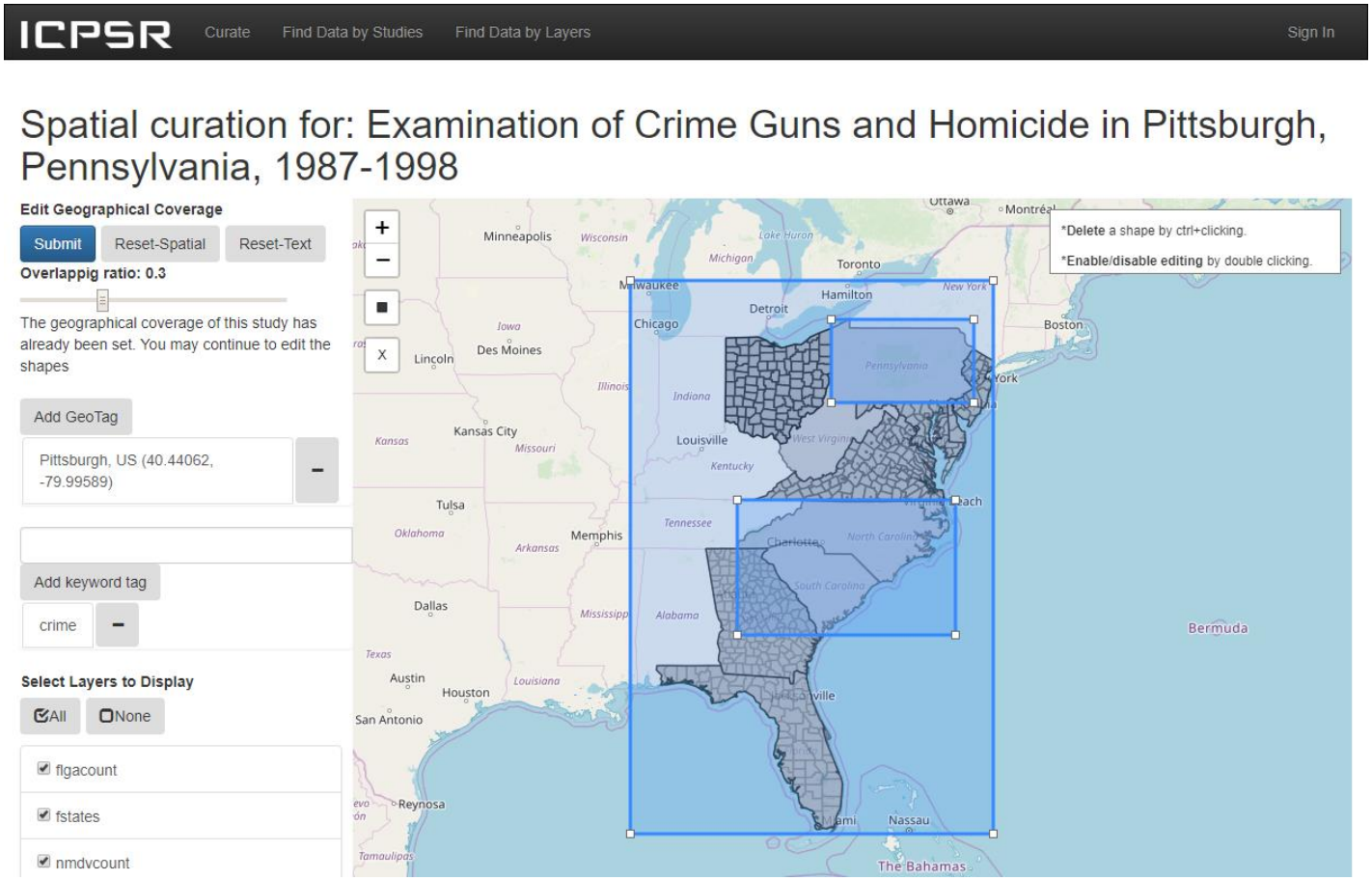
Type	Title	Name	Store	Enabled	Native SRS
	icpsr02895flgacount	icpsr:icpsr02895flgacount	postgis	✓	EPSG:4326
	icpsr02895fstates	icpsr:icpsr02895fstates	postgis	✓	EPSG:4326
	icpsr02895nmdvcount	icpsr:icpsr02895nmdvcount	postgis	✓	EPSG:4326
	icpsr02895ohcounty	icpsr:icpsr02895ohcounty	postgis	✓	EPSG:4326
	icpsr02895ohio	icpsr:icpsr02895ohio	postgis	✓	EPSG:4326
	icpsr02895pacounty	icpsr:icpsr02895pacounty	postgis	✓	EPSG:4326
	icpsr02895pazip	icpsr:icpsr02895pazip	postgis	✓	EPSG:4326
	icpsr02895penn	icpsr:icpsr02895penn	postgis	✓	EPSG:4326
	icpsr02895powstates	icpsr:icpsr02895powstates	postgis	✓	EPSG:4326
	icpsr02895scnccount	icpsr:icpsr02895scnccount	postgis	✓	EPSG:4326
	icpsr02895stateregs	icpsr:icpsr02895stateregs	postgis	✓	EPSG:4326
	icpsr02895us_states	icpsr:icpsr02895us_states	postgis	✓	EPSG:4326
	icpsr02895uspgh95	icpsr:icpsr02895uspgh95	postgis	✓	EPSG:4326
	icpsr02895westv	icpsr:icpsr02895westv	postgis	✓	EPSG:4326
	icpsr02895wvcounty	icpsr:icpsr02895wvcounty	postgis	✓	EPSG:4326
	icpsr04546ucnij	icpsr:icpsr04546ucnij	postgis	✓	EPSG:4326

Spatial metadata edit



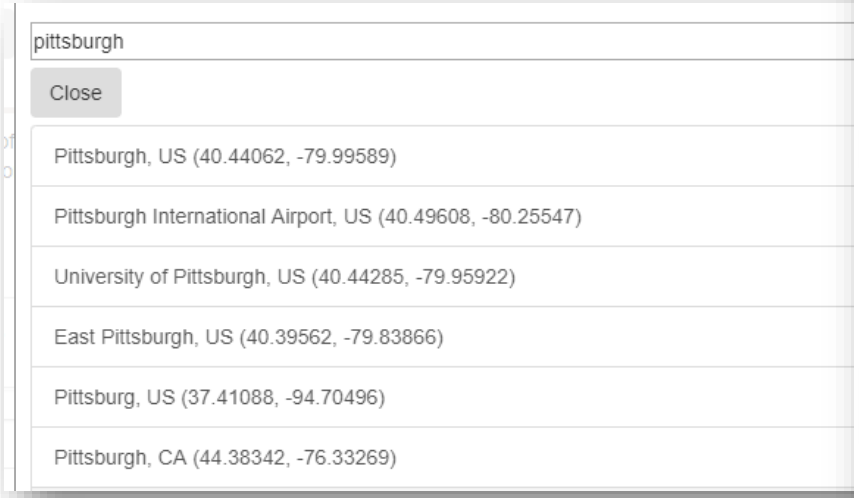
Spatial metadata edit

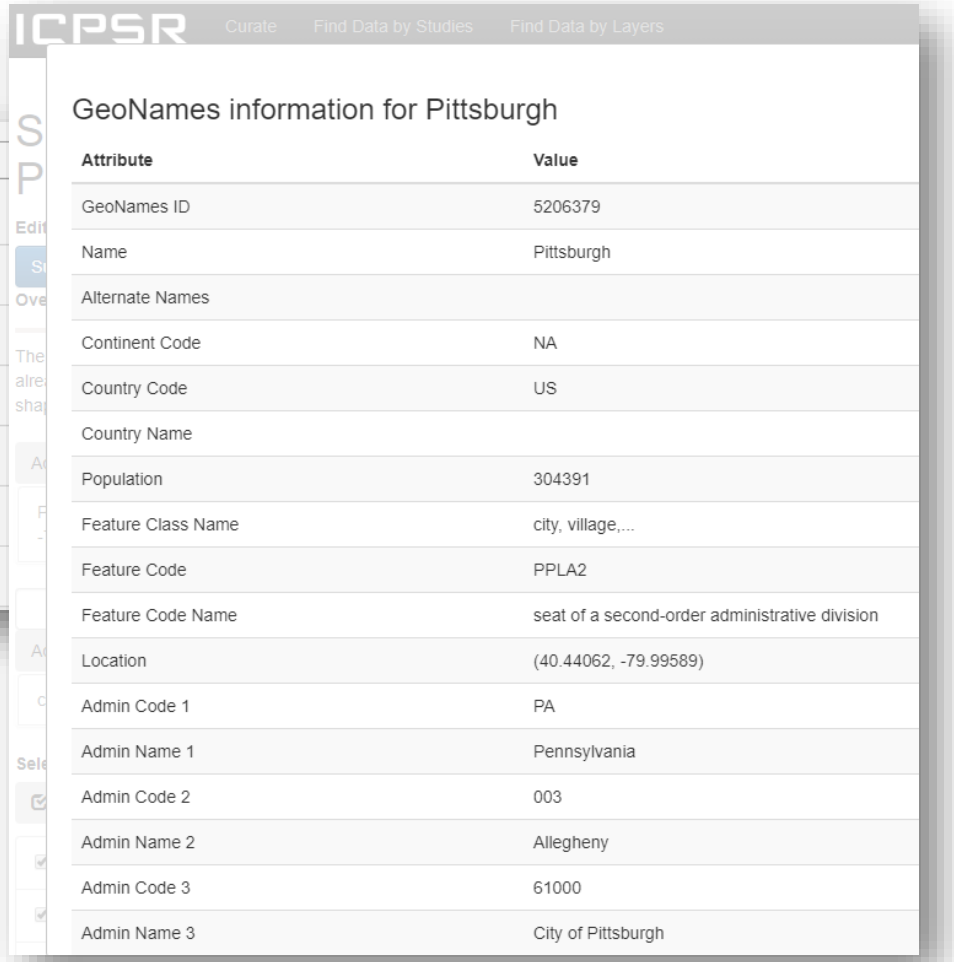
- Edit geographical coverage



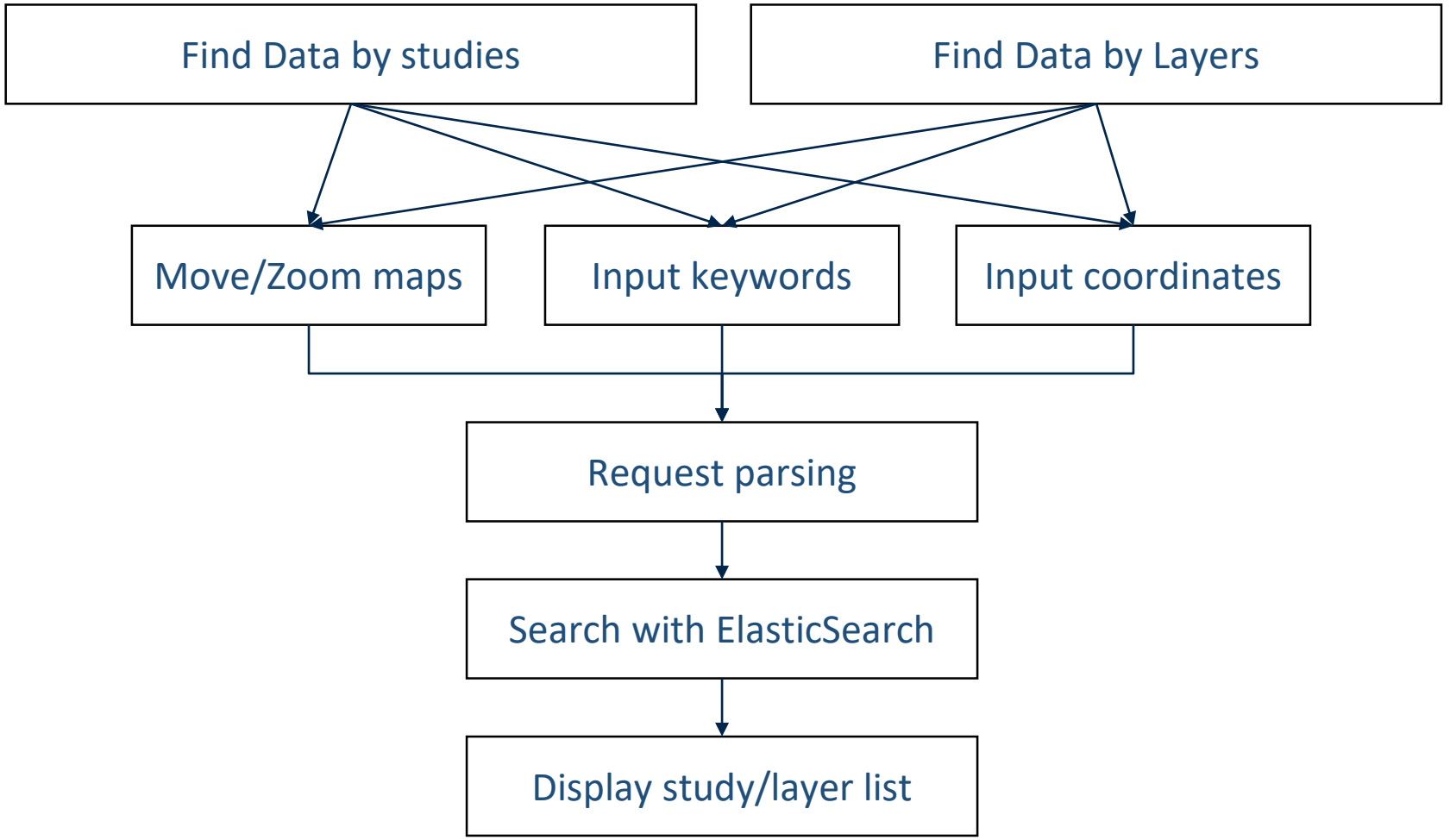
Spatial metadata edit

- Geotagging
- Keyword tagging



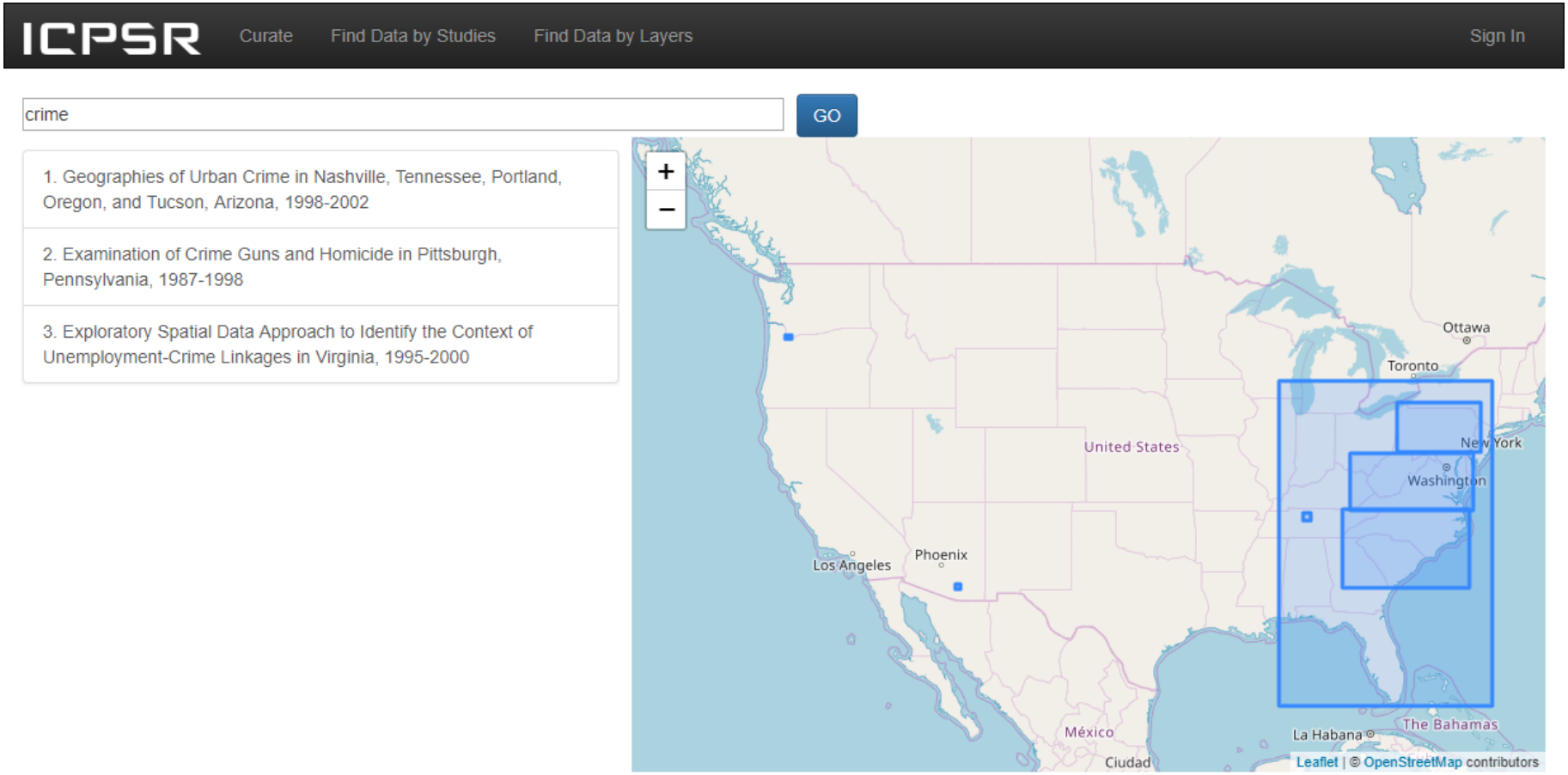


Spatial search



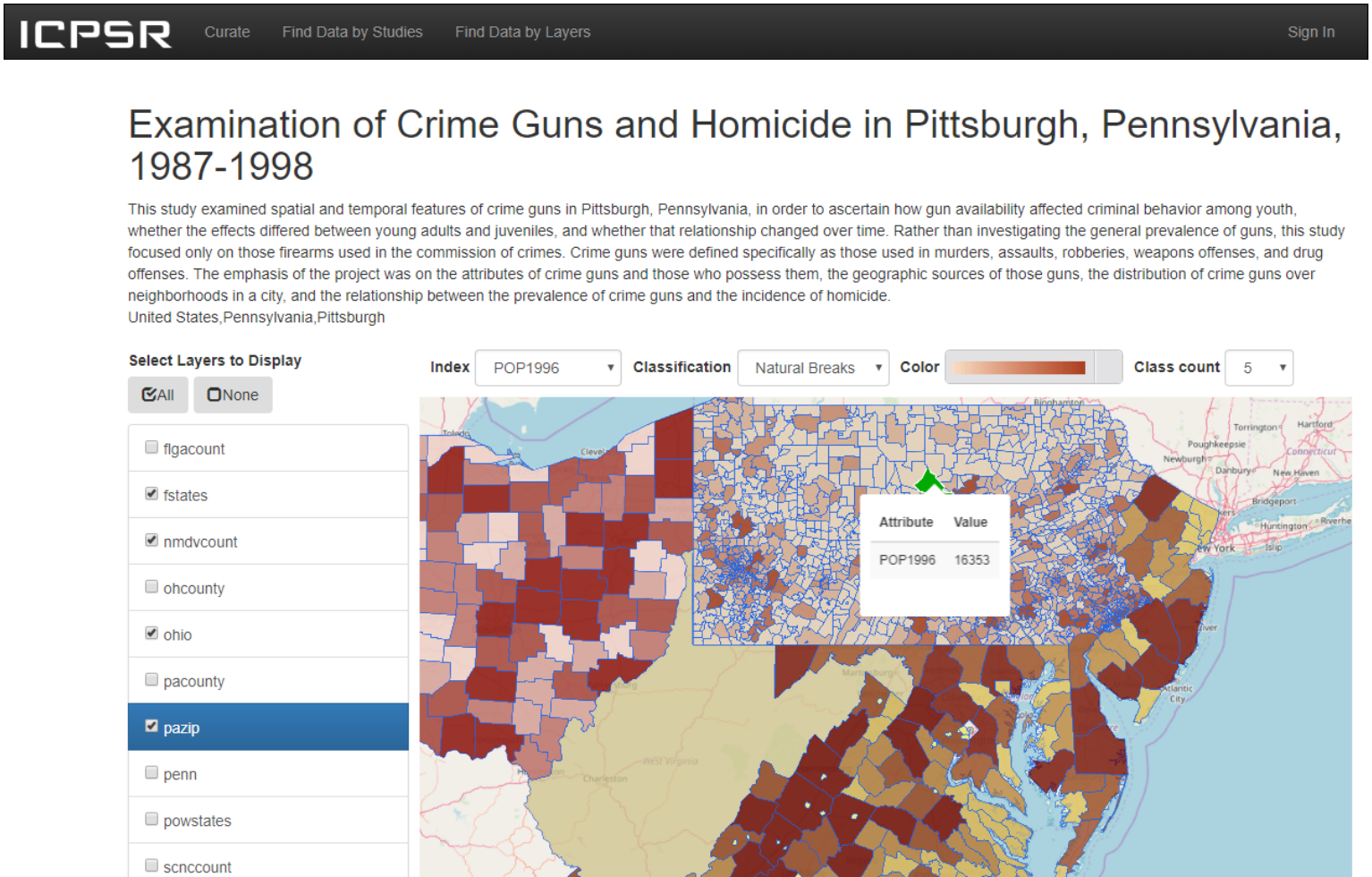
Spatial search

- Text + Map + Coordinate

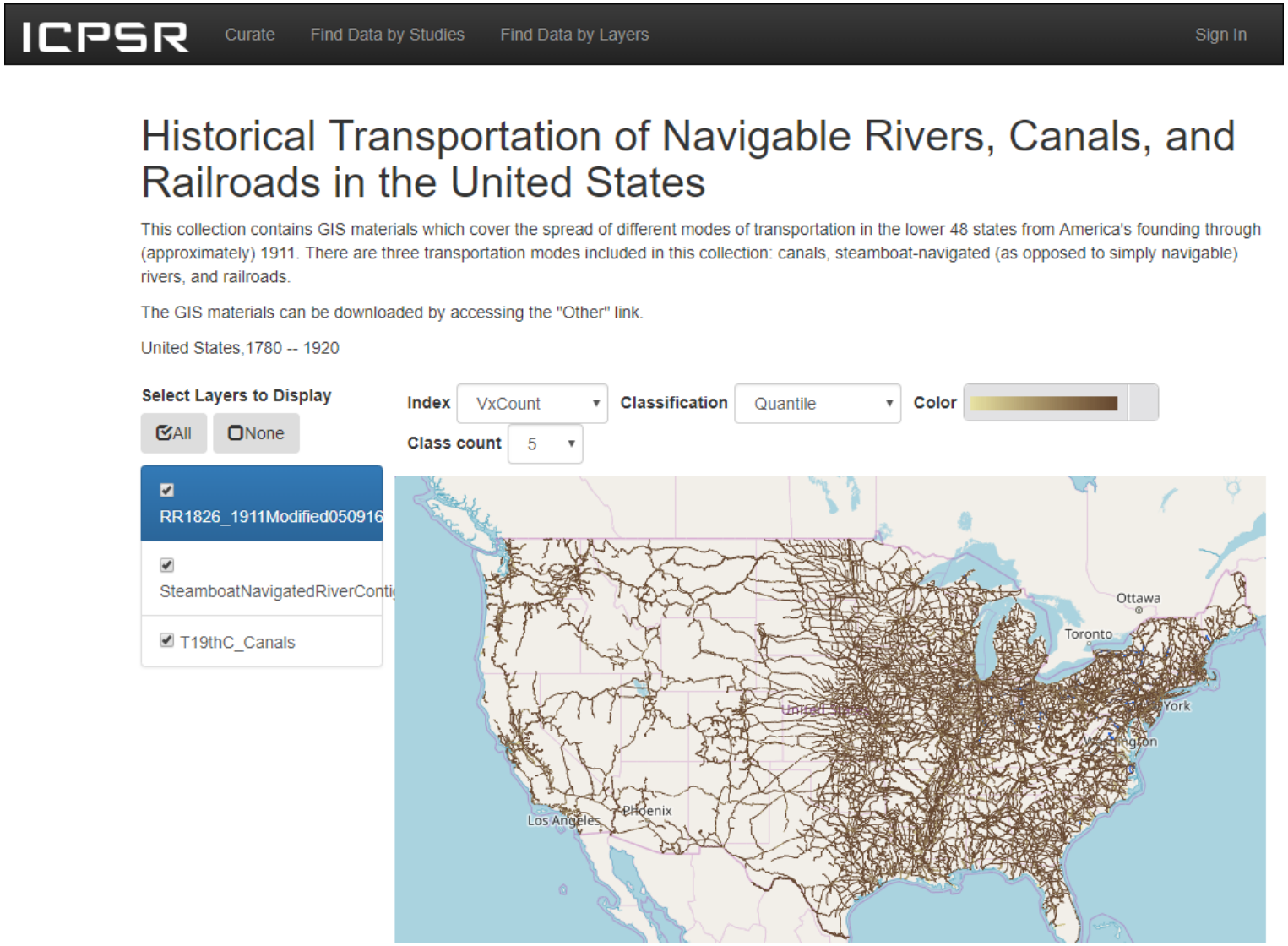


Study view

- Vector tiles + Thematic mapping



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
Ottawa
Toronto
York
Phoenix
Los Angeles

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