

Automating Tabular Data in a Geographic Context While Utilizing the GeoCommons Platform

FGDC NSDI CAP Grant

Award Number G09AC00107, Category 4: Enabling Use of Government Tabular Data in a Geographic Context

Final Report – September 2010

Executive Summary

FortiusOne has completed their FGDC CAP grant by extending the GeoCommons platform, enabling it to dynamically georeference tabular data with both point locations and place boundaries (polygons). This includes geocoding of street-level address data as well as referencing with datasets by a common foreign key. In the first phase, FortiusOne developed an open-source geocoding engine that is currently in use on GeoCommons public website. In the second phase FortiusOne developed an additional spatial join capability across arbitrary data attributes in addition to providing an industry standard service interface to data and maps – also publicly available on GeoCommons. In order to validate and demonstrate the georeferencing capabilities, FortiusOne georeferenced datasets from Grants.gov, USASpending, and FCC. The culmination of these efforts was an adoption of the technology by FCC to deliver IssueMap – a Web based application that turns pasted or uploaded tabular data into thematic maps that can be shared across social media service or embedded on websites.

Key Outcomes

1. Developed a GeoCoder and GeoJoin capabilities
 - a. US TIGER Data
 - b. OpenStreetMap
 - c. Arbitrary Boundaries (FIPS, ISO, et al)
2. Added OGC Standard Services: Catalog Services for Web (CSW), Web Map Service (WMS)
3. Deployed to GeoCommons and GeoIQ Federal, Intel, NGO, and Enterprise Customers
 - a. Deployed to Afghanistan and Haiti
4. Open-Sourced, adopted and growing by several geospatial companies and communities
5. Integrated into the FCC IssueMap Web application

Project Narrative

There exist large amounts of data in tabular formats that contain a geographic reference, but do not include geospatial coordinates. The goal of this project was to build and provide an automated and simple interface for users of the GeoCommons platform to georeference and share geographic data. The project was separated into two primary phases, geocoding address data, and geojoining against datasets based on a foreign key.

In the initial phase of the CAP grant, FortiusOne developed a modular geocoding engine that can accept either structured queries or free string entries and provide a list of street level, or less accurate, geographic location matches and estimates of quality. This engine was open-sourced with the purpose to more broadly promote the ability of tools to provide georeferencing.

The code for this project can be found at <http://github.com/geocommons/geocoder>.

The first version of the geocoder only supports US street address parsing and uses the US TIGER/Line road data. It currently only supports typical geographic entities such as countries, cities, address ranges. In developing the geocoder, there are many special cases in addressing schemes for US data. In addition, there is not a publicly available, global address database to use for world wide geocoding. One potential solution is the growing OpenStreetMap project that is actively adding addressing in many countries.

The FortiusOne geocoder was integrated into the free, public GeoCommons website so that anyone can upload and georeference tabular data. As part of this work, interaction design and usability testing developed a step-by-step walkthrough of uploading data from spreadsheets, syndicated feeds, or other formats, specifying geographic identifier attributes, and reviewing of the georeferenced data. All uploaded data is made available through GeoCommons and can be visualized and analyzed with other datasets using the application as well.

A GeoCommons portable server was also deployed to Jalalabad, Afghanistan to support the data sharing and collaboration between multiple NGO's, US Agencies, and other organizations. The FortiusOne Geocoder was loaded with 30,000 Afghanistan village names and used to geocode tribal information, citizen information, and situation reports leading up to and during the elections in August 2009.

USAID's Global Development Commons, JICA, UNHCR, UNDP, and Ushahidi also collaborated to share information. All data that was shareable was dynamically federated daily from the Jalabad server to the public GeoCommons data portal and multiple visual analyses were shared at <http://news.geocommons.com/afghanistanelection09> and featured by BBC, New York Times, Washington Post and other international media outlets.

There is a growing community of developers that are beginning to build parsers for other addressing schemes and languages for the geocoder, such as the GeoNames database as well as the OpenStreetMap project.

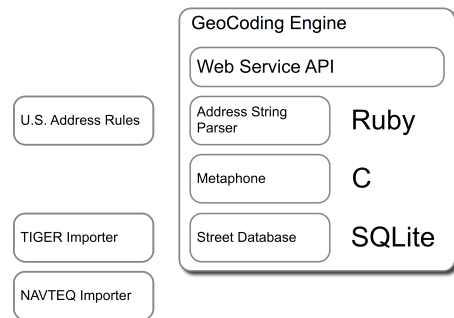


Figure 1: FortiusOne GeoCoder Modular Architecture

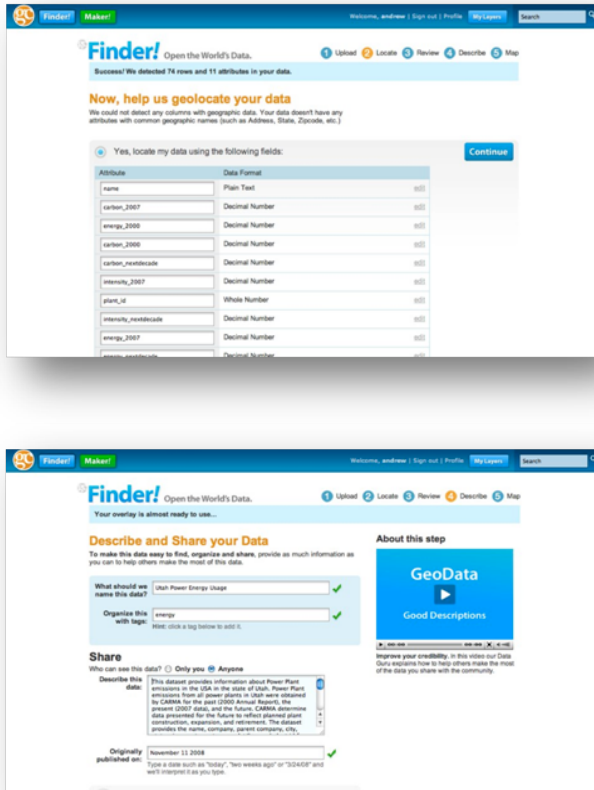


Figure 2: Geocoding Upload Steps

Phase Two

Phase two consisted of adding the spatial join functionality for public GeoCommons and building it into the FortiusOne platform. Any data set loaded into GeoCommons has the option to join their data set to a boundary based on a unique identifier in the tabular data. The workflow joining data to a boundary is highlighted in the diagram below.

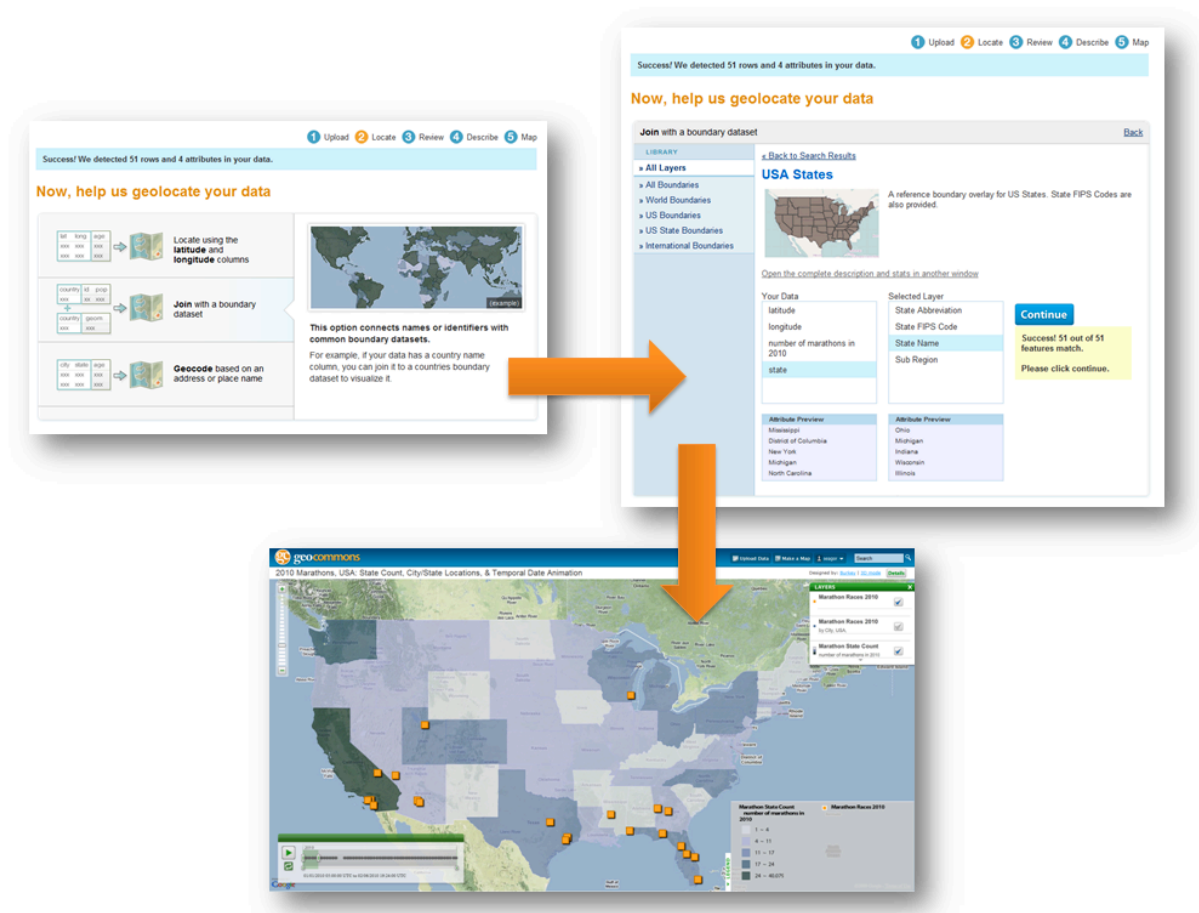


Figure 3: Joining Tabular Data to Boundaries for Thematic Mapping

Along with the spatial join capabilities built into the platform a comprehensive set of boundaries were provided through quick reference tools for facilitating rapid georeferencing. The platform also provides the ability for users to upload custom boundaries and use those for spatial joins as well. This provides a blend of ease of use and flexibility to deal with a wide variety of georeferencing scenarios. In addition to the user interface made available through the GeoCommons website an API has also been made available to provide programmatic access to developers

The deployment of the capability has seen widespread adoption both on GeoCommons as well as government agencies leveraging the API to create new Web applications with the service. A great example of this is the FCC’s IssueMap Web application. IssueMap allows users to cut and paste tabular data or upload it to the site and get a thematic map that is programmatically generated. Users can then share across social media services or embed on their blog or website. The diagram below illustrates the IssueMap work flow – proceeding from a copied table from a website to a thematic map in a few clicks:

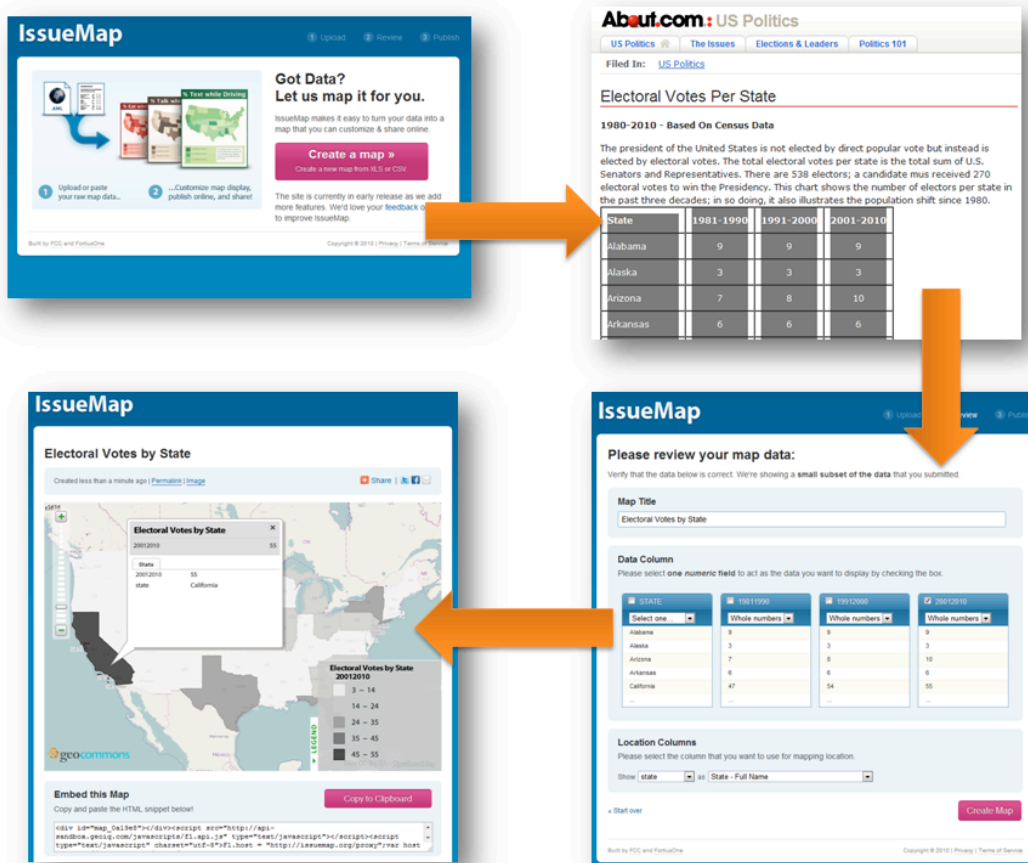


Figure 4: From Table to Thematic Map with IssueMap.org

Phase 2 also included building Open Geospatial Consortium (OGC) specification interfaces for search and federation. This included Catalog Services for the Web (CSW) and Web Feature Services (WFS). Combined with the existing OpenSearch-Geo and Atom syndication, these services will allow for easy federation between multiple GeoCommons servers and other data repositories such as Data.gov and Geospatial One Stop (GOS).

Next Steps

FortiusOne is currently seeking assistance in obtaining datasets and feeds to further test and demonstrate georeferencing data. These example datasets will illustrate the potential applications and benefits of geospatial government information. In addition, FortiusOne is discussing how the technology built in this CAP grant can support providing data to Data.gov and Geospatial One Stop in easy to use, open data formats. Also on the roadmap for continuing the project is the expansion of the service for international geocoding, adding fuzzy string matching for better place identification, and further speed improvement for the service. Lastly FortiusOne is actively promoting the use of the API's developed to enable more innovative Web applications like IssueMap to be developed by the community.

Feedback on Cooperative Agreements Program

The CAP grant program created a great opportunity to build an innovative service for the community without cumbersome restrictions about sharing the technology and making it publicly available. There was a latitude of freedom to create iterative innovations and expand the project as we developed the concepts. Unfortunately there was limited in-situ coordination with the FGDC and intermittent meetings to share progress with limited feedback. In addition, there was no integration into FGDC process to provide a vehicle to integrate project and resulting products beyond the proof of concept. It would have been ideal to have proof of concept metrics or integration opportunities along the way and at the end of the project.

.