

Intelligent Methods for Managing Geospatial Data at the US Federal Government

Feras A. Batarseh
College of Science
George Mason University
Fairfax, Virginia, USA
fbatarse@gmu.edu

Ruixin Yang
College of Science
George Mason University
Fairfax, Virginia, USA
ryang@gmu.edu

Abstract— in many operations, the US government depends on geospatial data to obtain insights and drive policy making. Federal data are usually pulled from multiple sources, aggregated differently, and formatted in an assortment of standards. Analysts have been overwhelmed by the ample amounts of inconsistent and unstructured data available to them; the main quandary is that they are unable to draw clear conclusions from that data without proper analytics. Moreover, when dealing with geospatial data, the challenge of data management can exacerbate. Geospatial data (including longitudes, latitudes, country codes, addresses and many other forms) are applied at many governmental use-cases such as: census, military operations, natural disasters, economical research, agricultural commodities' production, US trade, and many others. To address such dire needs, this paper introduces a novel method (called G-Fed) for geospatial data management at the federal government. G-Fed provides tools for streaming, validating, aggregating, and presenting data. Furthermore, G-Fed utilizes the power of big data analytics to drive its processes. Software engineering best practices are also applied within G-Fed; such as multithreading, security roles, on-demand data tools, standardization, and accessibility. The United States Government Accountability Office (GAO) published a report to point to the matter of federal data management. GAO's report identified fourteen issues that governmental agencies need to solve to overcome their data challenges. Efficiency pressures and federal budget constraints are pushing data implementations towards more automation, extra validation, and higher security. Additionally, the Open Data and the Big Data Initiatives (signed in 2012) have put more pressure on the backs of agencies to make their data (including geospatial) open and available to the public. However, without having state-of-the-art data repositories, it is not possible to address such demands. This paper addresses these issues by presenting a step-by-step set of guidelines. G-Fed aims to help agencies organize geospatial data in a manner that makes it possible to publish, and be provided online to the keen American public. A data repository of public government data is used for experimentation of G-Fed. Experimental results are recorded and presented.

Keywords— *Federal Geospatial Applications, US Geological Survey, Data Management, Visualizations, Government, Geo-Mapping, Geo-Data*

I. INTRODUCTION

A. A Historical Background

As early as 1890, the US government has been collecting data to map the states of the union in the most accurate ways possible [1]. During the revolutionary war, a good map could have influenced a win or a defeat. Maps were very essential in planning war tactics, drawing lines between states, acknowledging water bodies, rivers, and geographical landscapes. In their book 'Revolution: Mapping the Road to American Independence' [2], Richard Brown and Paul Cohen illustrated how the British monarchy was trying to keep the maps of the US away from the hands of the American revolutionaries. They understood that once the landscape is shared by both sides of the battle, war will be much more difficult to win. Sir Thomas Hyde Page was badly wounded when he was drawing a 'live' map during the war (Refer to Figure 1 below). His map, included even the flight of bullets during the war to capture the hottest areas of conflict (think of heat maps with recent mapping technologies) [1].



Fig. 1. A Handmade map from the revolutionary war [1].

After independence, detailed maps were put in place for all the states. The maps were needed to outline areas for potential agricultural projects, arid lands, rivers, lakes, and much more. The government published maps of different states in consecutive years, as shown in Figure 2.

The importance of geographical information and mapping continued throughout the years. In 1906, President Roosevelt, created the United States Geographic Board [4], to drive and highlight geo-mapping activities at the federal government. Executive order is shown in Figure 3 below.

The Department also publishes separate maps of the several States and Territories in which public lands of the United States are located, issued as follows:

| | | | |
|---------------------------------|------|--------------------------|------|
| Alabama | 1895 | Minnesota | 1894 |
| Alaska (double sheet) | 1898 | Mississippi | 1890 |
| Arizona * | | Missouri | 1891 |
| Arkansas | 1901 | Montana | 1897 |
| California (2 sheets) | 1900 | Nebraska | 1890 |
| Colorado | 1897 | Nevada † | 1903 |
| Florida (2 sheets) | 1893 | New Mexico * | |
| Hawaii | 1901 | North Dakota † | 1903 |
| Idaho | 1899 | Ohio | 1895 |
| Illinois | 1885 | Oklahoma | 1898 |
| Indiana | 1886 | Oregon | 1897 |
| Indian Territory | 1899 | South Dakota | 1901 |
| Iowa | 1885 | Utah | 1902 |
| Kansas | 1898 | Washington * | |
| Louisiana | 1896 | Wisconsin | 1896 |
| Michigan * | | Wyoming | 1900 |

* At present out of print.

† Will soon be published.

Fig. 2. State-level mapping as early as 1890 [3].

EXECUTIVE ORDER

The official title of the United States Board on Geographic Names is changed to UNITED STATES GEOGRAPHIC BOARD.

In addition to its present duties, advisory powers are hereby granted to this board concerning the preparation of maps compiled, or to be compiled, in the various bureaus and offices of the government, with a special view to the avoidance of unnecessary duplications of work; and for the unification and improvement of the scales of maps, of the symbols and conventions used upon them and of the methods representing relief. Hereafter, all such projects as are of importance shall be submitted to this board for advice before being undertaken.

THEODORE ROOSEVELT

THE WHITE HOUSE,
August 10, 1906

Fig. 3. Executive order for the creation of US Geographic Board [4].

Another important federal agency (that deals with mapping) is the United States Geological Survey (USGS) [5]. The dataset used in this paper is taken from its public data repository. Not only in the US, but all around the world, Maps are critical to governments, and they serve political, and economical interests in so many forms. The next sub-section introduces the motivation of this paper. Section 2 provides a review of existing methods (and federal geo-variables), Section 3 presents the contribution of this paper G-Fed, and lastly, Section 4 presents experiments, results and conclusions.

B. Motivation

Adopting new technologies is always central to the improvement of any organization (including federal agencies). Nonetheless, the challenges of change tend to vary among

organizations. Although it usually takes extended amounts of time to migrate an entire agency towards a new technology, some decisions that are taken at resource-constraint divisions could prompt a great deal of positive change. What is presented in this paper is an example of that: a free mapping platform applied at an agency to improve its geospatial data management processes – that is done without the need for legislative decisions, expensive contracts or extensive bureaucratic cycles [6 and 7]. Visualizing data with geospatial properties is becoming increasingly more prevalent due to the recent wide spread dissemination of devices, sensors, databases, and services that reference the physical world (which highlights the need of a well-organized meta-data). In most cases, animated maps are used to communicate facts and stories related to geospatial information in various application domains. However, while interactive geo-visualizations are useful in a variety of domains, the tools that can be employed to generate them are either very expensive (such as: Maptitude, ESRI’s Arc GIS, or Microsoft’s Map Point), requiring extensive training, cumbersome to use, lacking functionalities that are specific to the needs of an agency, or not flexible enough. Many amateurs and non-GIS researchers are now creating geospatial data representations more than ever before. To that point, the main goal of the method presented in this paper is to provide a framework that helps implement such visualizations; that is done by using open-source APIs for interactive maps (*Unfolding Maps* libraries in Java [8 and 9]). The dataset used in the experimental work of this paper has been taken from the USGS. The presented solution injects UI design principles in the development process. To encourage exploration of data, the presented example visualization provides additional understanding of seismic data through a set of responsive interactions. It showcases spots of seismic activities on the world map (to be discussed in Section 3) [6 and 7]. However, it is important to note that the proposed system could be used to map any type of geographical data – as long as it has the right variables (columns and rows). These variables are discussed in the following section.

II. STATE OF THE ART OF FEDERAL MAPS

Similar to most data science processes, before geo-mapping, it is important to clean the dataset, standardize it, aggregate it to the right level, and pivot it. Aggregation is especially important in federal geo-data. The government –for policy and decision making– looks at data on all geo-levels: continents, countries, regions, states, provinces, zones, zip codes, cities, and many more [7]. More importantly, maps are developed to **highlight** certain areas of interest, for example, conflicts in the world (such as Eastern Europe), certain areas of agriculture (such as the Corn-belt), watershed regions, and geo-locations with high hurricane activities (such as the Gulf of

Mexico). To represent this variety of geo-mapping requirements, it is important to include a comprehensive list of *geo data types* that enable such representations. These variables (meta-data) include: Longitudes, Latitudes, Geo-Type, Country Codes, and many others presented in section 3 of this paper (as part of presenting G-Fed). Not only in the US, countries such as Germany, China, Canada, and others have developed new policies to facilitate geo-mapping. For example, Heinze et al. [10] presented a method through geo-mapping that measures the levels of communication between citizens in certain areas of the country with their local governments. The goal was to highlight and encourage such communications. In 2007, Leung [11] published a doctorate dissertation showing how mapping helped with the urban development in Shanghai, China. In his dissertation, he showed that most ESRI Arc GIS maps in china are relevant to government use (refer to Figure 4), and that the Chinese government is investing in mapping more than ever before [11].



Fig. 4. ESRI Maps webpage in China (mostly government applications) [11].

Similarly, in the US, Seattle was one of the leading cities in integrating geo-mapping within its city functions. The city created maps for the following [12]:

- Public Land Survey Systems
- Surveys Control
- Street Rights-of-Way, Plat Boundaries and Lot Lines
- Geographic Features of the City
- Municipal and Park Boundaries

Moreover, many federal agencies opened up their geo-data to the public for mapping through open-source tools. NASA, for example, provided their tool, *World Wind Viewer* for open source libraries and general consumption [13]. In 2014, the USAID did the same thing [14], opening up their maps for open-source tools. Many other studies presented benefits of having a robust, open-source, and dependable geo-mapping system. G-Fed, the system presented in this paper is an example of that, and is presented in the next section.

III. FEDERAL GEO-MAPPING

Before drawing a map using any tool, certain data points need to be available. In G-Fed, a list of geographical data variable was collected after discussing geo-requirements with multiple federal analysts. The SQL database table that includes these meta-data variables and their data types is shown in Figure 5.

| | Column Name | Data Type |
|---|---------------------------------------|--------------|
| 📍 | ERSGeography_ID | int |
| | ERSGeography_Desc | varchar(100) |
| | ERSGeography_LongDesc | varchar(500) |
| | ERSGeography_PSD_Code | varchar(10) |
| | ERSGeography_UST_Code | varchar(10) |
| | ERSGeography_State_ANSI_Code | varchar(10) |
| | ERSGeography_State_FIPS_Code | varchar(10) |
| | ERSGeography_County_Code | varchar(10) |
| | ERSGeography_Zip_Code | varchar(5) |
| | ERSGeography_ZipPlus4_Code | varchar(9) |
| | ERSGeography_ERSGeographyDimension_ID | int |
| | ERSGeography_LatitudeDD | float |
| | ERSGeography_LongitudeDD | float |
| | ERSGeography_LatitudeDMS | float |
| | ERSGeography_LongitudeDMS | float |
| | ERSGeography_Country_Code | varchar(10) |
| | ERSGeography_County_ANSI_Code | varchar(10) |

Fig. 5. G-Fed list of geo-data variables (meta-data).

The variables shown in Figure 5 are saved in a database, all the country codes, geographical dimensions of datasets are presented in what we call a *geo-enabled database*. This database has the meta-data that represents any type of data point. For example, if a federal analyst is measuring the production of corn in a number of states in the US, that data includes multiple <values>, and is presented like this:

Statistic Type <Production>, of Commodity <Corn>, in Unit <Tons>, in Country <US>, States <California, Oregon and Washington>, Latitude <120.554° W>, Longitude <43.8041° N>, in Year <2017>, Month <May>.

It is important to note that the example above is only one data point; every data point has a similar structure though. Any value in the <bracket> represents actual data and resides in the Hadoop cluster, everything else is considered meta-data. The cluster is connected to the *Unfolding Maps* [9] Java geo library that federal analysts use to view, share, build and modify their maps. All the other meta-data components and structures (such as Unit, States, and Latitudes), are saved in the geo-enabled database. The main advantages of unfolding maps include: it is open source, automated through the mapping engine, data is organized through the meta-data, it is free, rich with libraries, data-type agnostic, and highly interactive. An example map built (earthquakes activity) using the system is shown in Figure 6. Besides being a rich map, the visual in Figure 6 creates intensity indicators, detailing geospatial information (such as location descriptors, areas, and magnitudes) when hovered over

or clicked on by the user. One key to enhancing the effectiveness of response and recovery activities during natural disasters for example - is through promoting the adoption of new technologies. That would further empower federal agencies involved in disaster management and mitigation activities. The proposed *example* is beneficial to multiple federal agencies (such as FEMA, USGS, NASA, and USAID) that monitor, get effected, manage or directly handle natural disasters. The presented *example* helps emergency management teams with allocating more resources and personnel in hazard-prone areas, which might reduce the financial burden caused by natural hazards and eventually save lives. The overall structure of the G-Fed system is shown in Figure 7.



Fig. 6. World map and earthquake activity around the world through G-Fed.

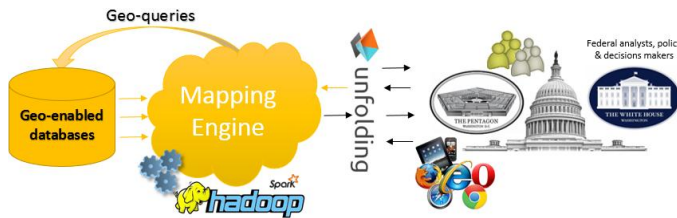


Fig. 7. The G-Fed System Overview.

Unfolding Maps, is a wide array of features that the federal analyst can benefit from. Libraries for maps include: Choropleth maps, Map overlays, Image markers, Circular snapshots, Colored transit lines, and Detailed maps with view ports. The main libraries and packages included in G-Fed are:

- java.util.HashMap
- java.util.List
- processing.core.PApplet
- de.fhpotsdam.unfolding.UnfoldingMap
- de.fhpotsdam.unfolding.data.Feature
- de.fhpotsdam.unfolding.data.GeoJSONReader
- de.fhpotsdam.unfolding.marker.Marker
- de.fhpotsdam.unfolding.utils.MapUtils
- processing.core.PConstants
- processing.core.PGraphics
- processing.core.PImage
- de.fhpotsdam.unfolding.geo.Location
- de.fhpotsdam.unfolding.marker.AbstractMarker
- processing.core.PApplet
- de.fhpotsdam.unfolding.UnfoldingMap
- de.fhpotsdam.unfolding.geo.Location

- de.fhpotsdam.unfolding.providers.Microsoft
- de.fhpotsdam.unfolding.utils.MapUtils
- And many more [8 and 9].

For more information on Unfolding Maps, please refer to their online repository [9]. As Figure 7 shows, after a request is initiated by a federal agency or analyst, the first step is to check whether the data are available in the Hadoop cluster, the **Mapping Engine** searches for the *'unstructured'* data based on the inputs provided by the user. After that, the data points are mapped through the metadata database. The data are then sent back to the engine to be opened within Unfolding Maps. The analyst then can use the *'structured'* data and draw a map to address the needs of the federal agency – the mapping is performed by the Mapping Engine, by filling the row values for the pre-defined metadata columns (shown in Figure 5). A wide array of features are possible through the Java geo-map libraries. The analyst is then empowered with the required tools to geo-map in a flexible manner. Commercial software vendors and 'expensive' government contractors tend to 'lock' the federal agency down to one mapping technology (for profit reasons). By using open source geo tools however, the agency has more power in terms of choice of tools, visualizations and would also save ample amounts of tax-payers' money. A survey was conducted to measure federal analyst user satisfaction, the questions, results, and other conclusions are presented in the next section.

IV. EXPERIMENT AND CONCLUSIONS

The technical sanity of the presented system was tested through being used at a federal agency. However, the most important challenge to its success is user satisfaction – i.e. the satisfaction level of the federal analyst. Due to issues with federal data privacy and confidentiality, data examples are not publishable. However, a detailed experimental survey was conducted to evaluate G-Fed. Although a small number, 32 analysts, filled out the survey; the questions introduced in the survey were comprehensive. The survey included:

- 1- Do you use maps in your daily tasks?
- 2- How many maps did you build using G-Fed?
- 3- How long (on average) did it take you to build one map?
- 4- What is your overall satisfaction level with unfolding maps?
- 5- What is your overall satisfaction level regarding G-Fed?

Answers for question 1 were: 70% yes, and 30% no. The average answer for question 2 was 11; while it was 5 hours for question 3. Unfolding maps had an overall satisfaction of 62%. The results from question 5 of the survey are presented in Figure 8 below.

At government, culture and bureaucracies are major challenges when it comes to adopting new technologies. Most agencies have contracts with software vendors that keeps them hostage to certain tools. The method presented in this paper is 'the first step' towards using free, open-source, and vendor-agnostic mapping at the government. The method presented requires much more evaluation and experimentation. However, initial results shown user satisfaction – which is very important in a federal setting.

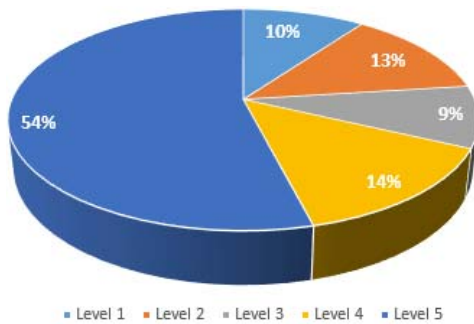


Fig. 8. Question 5 survey results.

Future work for G-Fed include: building more maps for different agencies, surveying more analysts, providing a compelling user interface, automating the process of building maps, adding more structures to the metadata, applying data mining algorithms to the geo-data through the Hadoop Spark library, expanding the metadata, and using different types of data sets within G-Fed.

ACKNOWLEDGMENT

Many thanks to Stuti Deshpande and Amogh Gaikwad for their coding efforts using unfolding Maps, and for providing some of the maps and material used in this paper. Additionally, much appreciation to the federal analysts that were involved in evaluating the new data system, and its tools.

REFERENCES

- [1] S. Warrell, "Making Maps under Fire during the Revolutionary War", An article published by National Geographic, 2016. Weblink: <http://news.nationalgeographic.com/2016/01/160117-american-history-revolutionary-war-map/>
- [2] R. Brown and P. Cohen, "Revolution: Mapping the Road to American Independence, 1755-1783", Published by W.W. Norton & Company, ISBN: 0393060322, 2015
- [3] E. Dawson, "Government Maps: Friends' Intelligencer 1853-1910", Published at the American Periodicals, pg. 192, 1904
- [4] W. Bowie "Board of Surveys and Maps of the Federal Government", Published by American Association for the Advancement of Science, Vol. 51, No. 1314, pp. 233-235, 1920
- [5] United States Geological Survey, documents and reports available at: <https://www.usgs.gov/>
- [6] C. Thomas, "Mapping the nation: government and technology making a difference", Published by Esri press, Redlands California, ISBN: 9781589483156, 1963
- [7] F. Batarseh and R. Yang, (Editors), "Federal Data Science: Transforming Government and Agricultural Policy using Artificial Intelligence", A Book Published by Elsevier's Academic Press, 2017
- [8] T. Nagel, J. Klerkx, A. Vande Moere and E. Duval, "Unfolding - A library for interactive maps", Published at the Interaction Design Lab at KU Leuven, 2016
- [9] Unfolding Maps Libraries, available at: <http://unfoldingmaps.org/>
- [10] J. Heinze, H. Schneider, and F. Feric, "Mapping the consumption of government communication: a qualitative study in Germany", Published at the Journal of Public Affairs, 2013
- [11] K. C. Leung, "Geographic Information and Urban Development in Shanghai", A Dissertation Published at The University of Minnesota, UMI number: 3264606, 2007
- [12] Nation's Cities Weekly Report, "Seattle integrates geomapping into array of city functions", Vol. 17, No. 3, ISSN: 0164-5935, 1994
- [13] J. Jackson, "NASA opens World Wind viewer as a mapping tool", A report published by GCN Magazine, 2007
- [14] USAID report, "USAID welcomes the crowd to use geo-mapping tools for open source development", available at: www.usaid.gov