Requirements Specification Document

GeoKings - GeoSTAC

Sponsors - USGS Amy Stamile, USGS Trent Hare, USGS Jason Laura

Version 2.0

Andrew Usvat, Zack Bryant, Alexander Poole,
Jackson Brittain, John Cardeccia

CS 476

Dr. Leverington

4/23/2023

Northern Arizona University

For the client: ______________________ ________
Signature *Date*

For the team: ______________________ ________
Signature *Date*

Table of Contents

Introduction

In today's era, there have been huge technological advancements that have opened the door to space exploration and research. This exploration has mainly been conducted by having satellites orbit moons, asteroids, and planets. Having these satellites in orbit allows us to gather tons of data including images. This information is then transmitted back to Earth and observed by scientific communities that use this type of information to understand as well as research the solar system [1].

Given that space exploration is growing rapidly, federal agencies and private citizens see the need to supply the planetary community with an easy way to access maps of planets and data for these planets. All planets in our solar system have the potential to be further explored. That being said, this makes it a necessity that researchers can use images and data from our satellite to attempt further exploration.

There are several resources that allow scientists and researchers to perform an analysis of the data and images they receive from the satellite, however, these resources are underdeveloped. The analysis of these resources by the community is done through sophisticated software which requires extensive knowledge of understanding. In addition to this, the community must store large amounts of data on their personal devices in order to analyze the data and review these images.

The Astrogeology group of the USGS (United States Geological Survey) stands at the forefront of planetary data preservation, analysis-ready data delivery, and planetary data access. Their main goal is to support community data exploration and availability. As part of this plan, the USGS has launched a project called "GeoSTAC", a web-based application created and updated by preceding NAU capstone groups. GeoSTAC allows the serving of planetary data using community-developed standards to access Analysis Ready Data (ARD). ARD refers to any type of data that has been processed through highly sophisticated imaging software, which significantly reduces the extent of data processing and obviates the necessity for the community to download large volumes of data for analysis and research purposes. The primary objective and use of the GeoSTAC application are to implement their focus areas outlined above by providing ARD to the global planetary science community, simplifying their research, and rendering it more efficiently.

Problem Statement

GeoSTAC has demonstrated good delivery of ARD through their STAC API, which exclusively handles raster data. This API fetches digital footprints and attaches them to an existing raster map. While the STAC API is helpful in transmitting information to the user, the digital footprints reveal only a small portion of the existing information that is available to users. In order to enhance the capabilities of the current application, our team aims to introduce the following functionality to GeoSTAC:

- Integration of geological data

- Customizable topography styling
- Implementation of a comprehensive search engine

By accomplishing these functions, our team will be able to transform GeoSTAC into a more useful and advanced research and geospatial data tool.

Solution Vision

We are proposing to integrate the new Vector API with the GeoSTAC application in a way that compliments the application without having to rebuild the entire existing project. We see this implementation as an update to the overall system since we are modifying the application to work with the new Vector API. The solutions to solve the previously mentioned problems from the problem statement are as follows.

- Fetching and rendering the data via a fetch call and rendering all data points with javascript in Leaflet.
- Using a package called Leaflet.SLD in order to stylize the vector lines with a provided SLD file.
- Integrating Elasticsearch to be able to search and query through the Vector API to return a specific type of data.

We are taking all of the information given by the Vector API and incorporating the data in the GeoSTAC application in a similar manner that the currently used STAC API is. The STAC API consists of data that allows us to visualize the maps seen on the

GeoSTAC application. Similarly to how the STAC API works, we are passing the data from the Vector API to a variety of different tools in order to achieve a desired outcome. The only notable difference is we'll be modifying a program in order to display the data from the Vector API in a way that allows us to use symbols when rendering vectors. We currently plan on modifying an existing program called Leaflet.SLD which will allow us to display vectors using symbols.

Project Requirements

I. Functional Requirements

This section will discuss the requirements carefully outlined by our sponsors. Each section is a working part and will contain the specifications required to reach the ultimate goal set by USGS. The GeoKings aim to accomplish rendering vectors, stylizing vectors, selecting specific information using Elasticsearch, and code documentation.

1. **Ability to Render Vector data in the form of GeoJSON received from the Pygeo API**

Vector data will be delivered to the application using an API given to us by the Sponsors. With this API, we will be able to receive GeoJSON data that can then be displayed as polygons, lines, points, etc. These shapes will be structured in a way that makes sense scientifically and will be interactive, e.g. clicking on the shape to access its feature properties and information. This will either be done using vector tiling or base leaflet, depending on the functionality observed while stylizing.

2. **Ability to stylize the vector features using Styled Layer Descriptor (SLD) files**

Vectors that correspond to a certain geologic feature type will be displayed on a leaflet map with respect to a provided SLD file. This SLD file will contain information on how certain geologic features will be styled and additional functions will be provided in order to display more complex designs of these vectors using SVG files. These SVG files will contain images related to standard symbols used in topographic maps in order for users to have familiarity with the symbols [2].

3. **Use the current Elasticsearch backend to query the vector API**

For this newly added vector data to be useful, users of the GeoSTAC application will need to be able to filter through the available data, narrowing what vector data will be displayed down to the data set they are interested in. The current implementation of Elasticsearch allows users to search by keyword for the desired raster data, with properly formatted search criteria being passed to the STAC API.

Our team will update the Elasticsearch back-end so that in addition to the STAC API, it can pass queries to the new Vector API. These queries will need to be formatted according to the Open Geospatial Consortium's (OGC) Common Query Language (CQL) specification.

User interface (UI) enhancements will need to be made to provide users with the ability to find and filter for the correct vector data. We anticipate using checkboxes to select the specific geologic features being sought after, though we may evaluate other options such as the text search currently employed for raster data.

II. Performance Requirements

This performance section ensures that our project is accurate and functions appropriately. The first section we review is data accuracy which is important because we need to ensure that we have the correct data when releasing it to the public. The second section reviews user testing which is crucial for the purpose of ensuring that users can properly use and interact with the GeoSTAC interface.

1. Data Accuracy

This project is a data discovery tool, and because of this, we need to test if the data we are getting from the API is correct. That being said, there are two types of checks that will be conducted: (1) making sure that the API is rendering correct information to its corresponding planet and (2) checking that metadata given to SLD files from the API is accurate. Both of these tests will ensure that we are receiving the correct data all around the platform.

2. User Testing

This project is specifically used for the purpose of making an interface that can easily be used by the community. That being said, it is important to ensure that the various functions used in the tool are easily understandable and can be familiar to regular users in a short period of time. Users should be able to learn these different functionalities from the interface and from any documentation.

Being able to search and select are the features that will be more widely used. Because of this users should be able to understand exactly how to execute both of these functionalities. Both searching and selecting are considered to be more non-intuitive functions on the GeoSTAC application meaning that the user will likely have to view some sort of documentation in order for them to accomplish some task (selecting an area or downloading an image). Searching is not intuitive because it will require the user to perform advanced search operations using the Features API CQL specification. Selecting is not very intuitive because the user is likely not to be familiar with Leaflet icons on the map. That being said, there will have to be some sort of documentation in order to explain both of these functionalities on the application.

To test these features, there will be some sort of user testing done within the GeoKings implementation timeline. This is made possible due to the fact that the application is publicly available to anyone. It is our goal that when using both the Area selection and the Elasticsearch capability it should take no more than thirty seconds to get results, we aim to accomplish this through user testing. The team will provide objectives to users for specific functions (Area selection and Elasticsearch capability). If the time limit of thirty seconds is not met we ask that the user additionally mention how long it took them to complete the task. Once testing is complete, the user response will be recorded as well as any additional comments or suggestions. These types of user tests are essential in order to assure the application with its new additions is working properly and quickly.

3. Code Usage and Documentation

Documentation of the new functionality to the GeoSTAC codebase and how to work with newly interjected tools is required. This ensures that if the project is handed off to other teams they will be able to understand the project and its newly implemented functionality as well as continue the project in a timely manner.

Functions will be documented in the codebase with a comment header, explaining the purpose of the function, the parameters it takes, and what it returns if anything. A user guide will be created as well to explain how certain tools of the project work and why these tools were included in the project. The user guide will also explain the data flow, from fetching the vector data to displaying the vectors with symbols.

III. Environmental Requirements

Given the nature of this project, where the foundation of the project is already in place, there are many environmental requirements that must be met. As stated previously, the GeoSTAC website is already live, lacking the capability of rendering vector data on the integrated map, only being able to display raster data. This leads to the constraints of using the technologies that have already been implemented that have brought the web application to the stage of development that it's in today.

The web app is primarily built using the programming language JavaScript, with elements of CSS and React involved to support the meat of the program. Within the scope of JavaScript, elements of the new functionality are required to use specific libraries. As an example, the usage of the JavaScript library Leaflet for rendering the maps has been specified by the clients. The system will be constructed using the JavaScript library to ensure seamless integration with the technologies already in place.

In terms of the search functionality, using a provided API to access the Elasticsearch backend has been requested by the clients. The system will handle requests to the provided API in order to receive the necessary vector data.

To stylize the vector data, Geologic Symbols will be used, which is a SLD file that stores geologic symbols for mapping applications such as Leaflet, which is a requirement from the clients. The system will allow users to differentiate between the varying stylizations on the map, which represent identifiable features, as well as interact with the vector data being presented, resulting in greater overall comprehension.

All of these constraints fall under a larger scope of constraints, which is outlined by the U.S. Web Design System (USWDS), a set of guidelines expected of all U.S. Government affiliated websites. These guidelines apply to GeoSTAC as it is a product of the United States Geological Survey (USGS). As shown on the USWDS website, “USWDS provides principles, guidance, and code to help you design and build accessible, mobile-friendly government websites and digital services.” [3]. The system will comply with the standards set in place by USWDS.

Potential Risks

With any project, there will inherently be risks that may have consequences of varying degrees following the release of a product. Given the environment in which GeoSTAC exists, there do not seem to be risks that would fall in the severe category, such as causing death. With that being said, there are some risks that could potentially have lasting effects if they happened to occur.

A potential risk is that the system would display inaccurate data to the user of the web application. With the primary demographic being astrological researchers, not just for leisure, providing inaccurate data could prove to be detrimental to the researchers’ work. In this scenario, the likely outcome would be that the researcher is inconvenienced, with the fix being fact-checking the data against another source. In a more extreme example, the issue would cause rippling effects across an academic sector. This could result in the necessity to fix not only the data but also the mindset of the individuals provided with the wrong information and all those that their data reached, which would be much more difficult to remedy.

Another potential risk that the system may have is that it is not compliant with government regulations. With this web application being a product of the United States Geological Survey (USGS), it must follow certain guidelines. One important guideline is the previously discussed U.S. Web Design System (USWDS) guidelines, which dictate how government-affiliated websites should be structured. There was no information about the consequences of not following these guidelines, but one could infer multiple negative outcomes. Some negative outcomes could include cuts in funding, temporary website deactivation, and other demerits until a solution can be put in place.

There will always be risks involved with the development of a new project. By acknowledging the obvious risks early, making a plan to mitigate them, and continuing to be vigilant and aware of new potential risks, the severe outcomes can be nearly negated.

Project Plan

As the foundation for this project becomes cemented, it is important to look forward and make a plan for how development is going to continue. This has been accomplished by setting milestones that will guide and inform about the progress being made. The milestones have been laid out to provide the most seamless workflow transition possible. Although these milestones have been established, over the course of development, more milestones may be added as needed, or the timeline for a given milestone may have to be adjusted to accommodate new circumstances. As of now, the currently established milestones include: rendering vectors with Vector API to leaflet maps, stylizing vectors with symbols, being able to query through vector API and output to the GeoSTAC website with a proper presentation using React, and implementing Elasticsearch. As of now, these milestones encompass the requirements and are broken down into attainable goals to facilitate development.

The development process has been broken down into two time periods: the time remaining in this current semester, and the time given for next semester. Leaving out the time in the summer allows for a realistic assessment of time in case any work cannot be done over that time period.

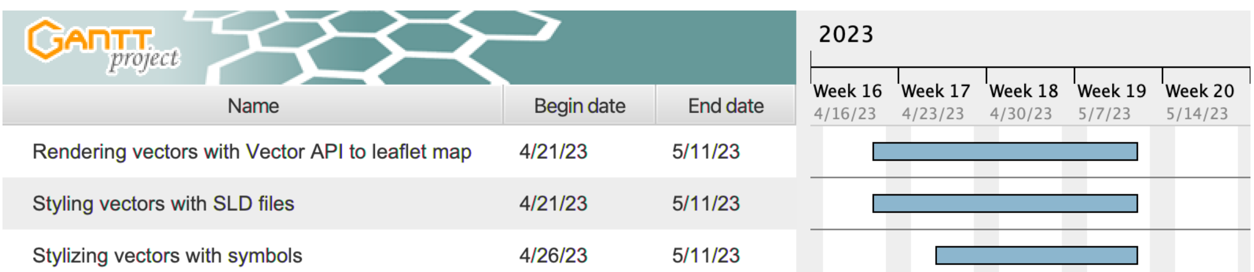

Figure 6.1

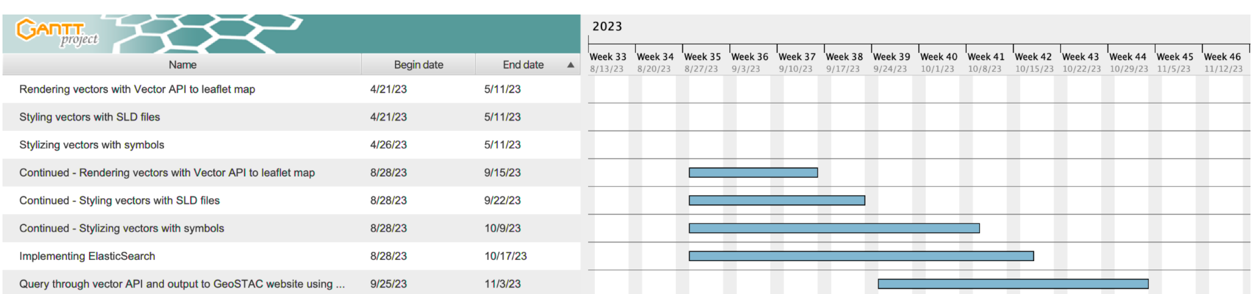

Figure 6.2

Conclusion

The collection of satellite imaging on other planets in our solar system over the last several decades has resulted in a data set capable of generating detailed scientific maps. The USGS has worked to make this data accessible to the scientific community around the world, however, this has brought to light a problem – there are no compelling software solutions available to provide easy and useful access to this map data.

The software that exists capable of rendering planetary map data is often cumbersome to use and is proprietary - meaning that it can be cost-prohibitive to use. The USGS has solved this by tasking two prior NAU capstone teams with developing free, open-source software to fill this technological gap – the result is the GeoSTAC web application, our team is now the third tasked with advancing this software.

A current limitation of GeoSTAC is its inability to render and stylize vector data; the software currently handles only raster data. Our team is evaluating multiple pathways to incorporate the ability to both render vector data and apply stylization to this data using SLD and SVG files. We will also introduce UI enhancements, and update the existing Elasticsearch backbend to allow scientists to search the map for the specific vector data they are interested in.

Incorporating this new feature set into the existing GeoSTAC application will increase the amount of data that can be provided to scientists, as well as improve its accessibility and usefulness of it. In turn, this will aid scientists in their research and help to contribute to our understanding of our planetary neighbors.

References

1. Stamile, A., Cain, J., Kaufman, Z., Nelson, G. (2021, December 4). Documents and presentations. GeoSTAC - Documents. Retrieved April 17, 2023, from https://ceias.nau.edu/capstone/projects/CS/2022/GeoSTAC/webPages/document.html.
2. “Topographic map symbols - USGS,” Topographic Map Symbols. [Online]. Available: https://pubs.usgs.gov/gip/TopographicMapSymbols/topomapsymbols.pdf. [Accessed: 17-Apr-2023].
3. “How to use USWDS,” *U.S. Web Design System (USWDS)*. [Online]. Available: https://designsystem.digital.gov/how-to-use-uswds/. [Accessed: 17-Apr-2023].