IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-6032-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/IGARSS53475.2024.10641425

THE VISUALIZATION, EXPLORATION, AND DATA ANALYSIS (VEDA) PLATFORM: A MODULAR, OPEN PLATFORM LOWERING THE BARRIER TO ENTRY TO CLOUD COMPUTING

Brian M. Freitag[1], Manil Maskey[1], Aimee Barciauskas[2], Jonas Solvsteen[2], James Colliander[3], James Munroe[3]

[1] NASA Marshall Space Flight Center, Huntsville, AL, USA
[2] Development Seed, Washington DC, USA
[3] 2i2c / Code for Science and Society, Portland, OR, USA

ABSTRACT

Increasing data volumes and migration of data to the cloud introduces challenges related to scalability and complexity to users of National Aeronautics and Space Administration (NASA) Earth science data. The Visualization, Exploration, and Data Analysis (VEDA) platform is an open-source modular cyberinfrastructure that leverages community standards and builds upon existing open-source capabilities for data services and geospatial data visualization. Four major components make up the VEDA platform – a cloud-optimized data store, backend data services, a web-based visualization and data-driven storytelling dashboard, and an analysis hub. The modular design of the VEDA platform enables the tailored use of the platform components or the platform in whole as has been demonstrated by other applications leveraging VEDA within NASA and the U.S. government.

Index Terms— Data visualization, Data Analysis, Data Science, Open Science, OGC, STAC

1. INTRODUCTION

The recently released open-source science initiative by NASA focuses on transparency, inclusiveness, accessibility, and reproducibility of technology and data produced using NASA funding [1]. With growing data volumes, data migration to the cloud, and cloud-computing resources innovative solutions are required to further support these open science principles. Leveraging existing community-supported software, the VEDA platform is designed to provide an infrastructure to lower the barrier to entry to cloud-hosted data and cloud compute resources for NASA. With an interactive web mapping interface, supporting documentation, and a cloud-based analysis hub VEDA users can explore cloud-hosted data within a web browser, directly access cloud-hosted data, and leverage cloud computing for scalability by querying a single application programming interface (API) URL. Built to community standards such as those managed by the Open Geospatial Consortium (OGC) [2] and the Spatiotemporal Asset Catalog (STAC) specification [3], the VEDA platform also optimizes data access and visualization of cloud-hosted data products through publicly accessible data service APIs. By aligning with community standards, the VEDA platform enables interoperability of data services with other platforms built to the same community supported standards. This provides a turnkey, redeployable infrastructure for geospatial science data applications. The rapid deployment of VEDA components for different applications is made possible by the modular design of the platform which is described in more detail throughout the remainder of the paper.

2. VEDA PLATFORM

The target audience for the VEDA platform is broad ranging from science-curious public to research teams. In general, users across this spectrum can be binned into two primary user groups, contributors and consumers. Researchers leveraging the analysis hub and the data services components of VEDA can contribute data, algorithms. Additionally, through data-driven story telling capabilities of the dashboard, researchers are also able to communicate the value of their science contributions and how it can be leveraged by the community and public to better understand the Earth system. This workflow encompasses the VEDA content lifecycle for contributors to the VEDA platform (Figure 1). Consumers of the VEDA platform are broader and extend beyond the research community to the public and students interested in Earth science and learning more about the Earth system. Serving as the bridge between the research community and the science-curious public, VEDA enables open science through transparency and accessibility of otherwise complex scientific concepts. There are four major components of the VEDA platform that enable these connections between researchers and the science-curious public.

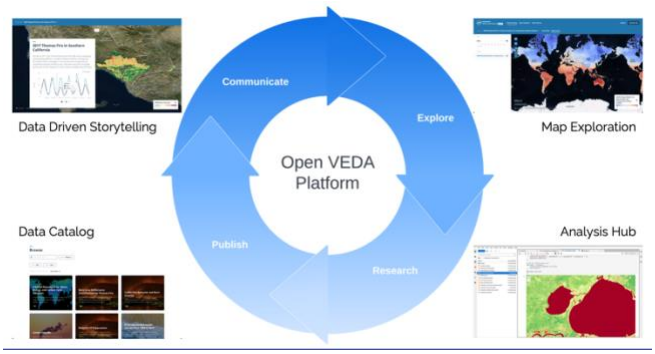

Figure 1. VEDA content lifecycle.

2.1. Data Catalog

The data catalog within VEDA contains cloud-optimized formats of user contributed and high-value datasets for existing platform users in an S3 bucket. Each dataset in the catalog contains collection level and asset level metadata that adhere to STAC specifications. By doing so, the VEDA data catalog is interoperable with other platforms supporting STAC and supports easy integration with GIS applications.

2.2. Data Service APIs

Data services within the VEDA platform include a suite of APIs for ingesting, searching, and visualizing both raster and vector datasets. The open-source suite of APIs, named eoAPI, leverages STAC for data ingestion, hosting, and querying. Services included within eoAPI are fully scalable, interoperable, and open source aligning with NASA's open science principles, but also supporting rapid adoption and adaptation of these services. The VEDA data services APIs enable dynamic web-based raster and vector visualizations for cloud-hosted data.

2.3. Web Dashboard

The VEDA dashboard is the primary point of entry into the VEDA platform. It serves as the primary mechanism for data exploration, analysis and communication. Visualization of data layers hosted in VEDA leverage the data services APIs to enable intuitive exploration of data in a web mapping client (Figure 2). The VEDA dashboard includes a data exploration and analysis environment that allows for more detailed interaction with the various datasets including data layer toggling and date comparison sliders. Within that same page users are also able to compute basic time-series area-weighted spatial statistics for a user-defined area of interest. A key feature of the dashboard is the data-driven storytelling component where content contributors can relay complex scientific information in an interactive, intuitive way with capabilities like map comparison sliders and highlighting locations with a geospatial overlaid with an information card for communicating key points. This is key for bridging the gap between the science curious public and science research teams for understanding the Earth system.

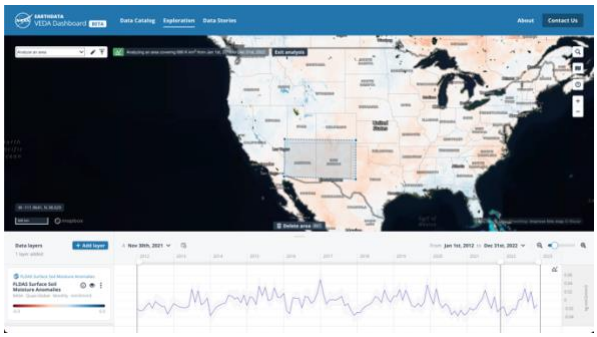

Figure 2. VEDA dashboard map exploration page

2.4. Analytics Platform

For scientific users interested in more detailed scientific analysis, the VEDA platform includes an analysis hub where users can seamlessly launch into a cloud compute environment (Figure 3). The analysis hub includes capabilities to launch into Python, R, and QGIS environments as well as the option for users to bring their own image for data processing and analysis. Seamless integration with the dashboard and the analysis platform abstracts the complexity of cloud-computing and facilitates a migration from legacy to cloud-based workflows within the science community. Moreover, the analysis hub is collocated with NASA's official data archives in the cloud for optimized S3 access to all of NASA Earth science data. This enables science teams to integrate and compare their data analysis or user-generated data products with official NASA products.

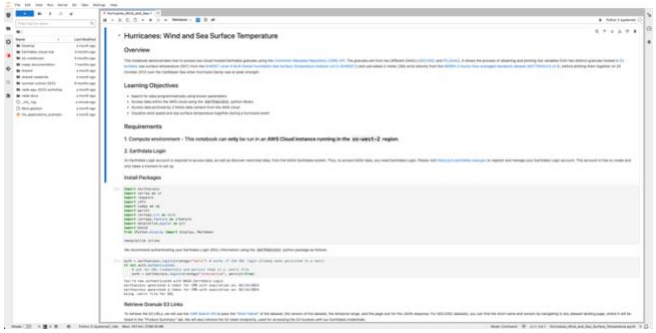

Figure 3. VEDA analysis hub with science example notebook

2.5. User Engagement

VEDA is designed to support a broad user community from the public to scientific research teams to decision makers. User interaction with the VEDA platform varies for each of these users based on their specific needs (Figure 4). Additionally, the modular design of the platform enables program managers to select the components of the platform they want to address the needs of their stakeholders.

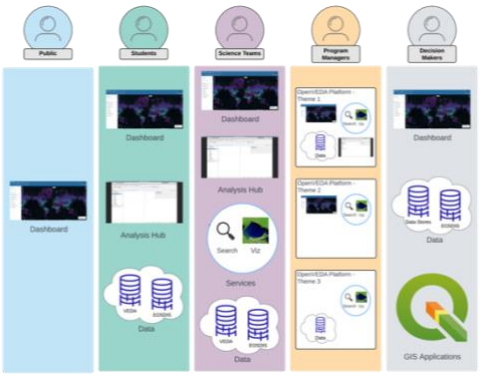

Figure 4. User engagement with the VEDA platform

3. VEDA DEPLOYMENTS

Capabilities of the VEDA platform expands upon an existing legacy of open-source geospatial data software developed within NASA. Dynamic visualization capabilities from the Multi-mission Algorithm and Analysis Platform (MAAP) [4] are the core of web-based dynamic data visualization within VEDA. Front-end capabilities of VEDA are adapted from the Earth observation dashboard for exploring and communicating the impacts of the COVID-19 pandemic [5]. Incorporation of open standards and incorporating existing open-source solutions for data visualization and exploration capabilities has allowed for rapid deployment of VEDA for other NASA-led initiatives. Further, the modular design of the platforms components has enabled technical teams within these different initiatives to quickly deploy VEDA in part or in whole.

3.1. Earth Information Center

The multi-agency Earth.gov portal [6] launched in December 2023 leverages the VEDA dashboard for data exploration and storytelling of Earth science information (Figure 5). The standards-based data services within the VEDA platform allow for data from disparate catalogs to be incorporated into the Earth.gov data analysis environment. In other words, Earth.gov does not have its own data catalog but rather visualizes data from existing catalogs at NASA within the exploration environment.

3.2. U.S. Greenhouse Gas Center

The multi-agency U.S. Greenhouse Gas Center [7] launched in December 2023 contains a data portal that is a full deployment of all the VEDA components. The modular design of the VEDA platform enabled deployment of the U.S. Greenhouse Gas Center data portal in under four months (Figure 6). Upon initial release the platform included 13 datasets and 4 data insights. Added capabilities identified by U.S. Greenhouse Gas Center stakeholders were rolled back into the core VEDA architecture demonstrating the ease of incorporating new technologies into the various services.

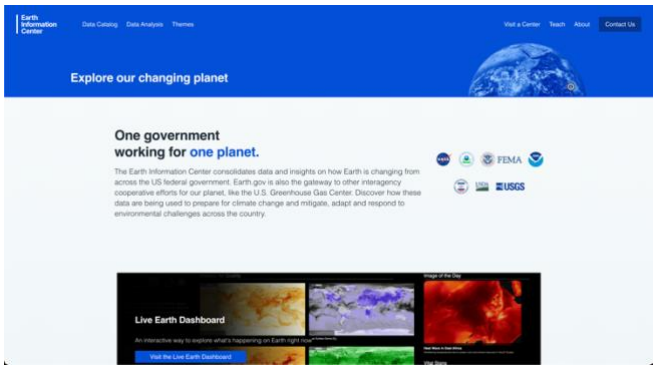

Figure 5. Screenshot of the Earth Information Center homepage

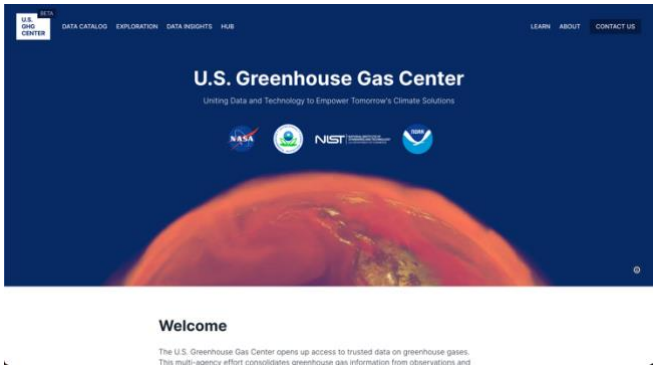

Figure 6. Screenshot of the U.S. Greenhouse Gas Center homepage

3.3. Fire Information for Resource Management System (FIRMS)

The FIRMS web portal [8] integrates many different data sources supporting the monitoring of wildfires, prescribed burns, and fire recovery. The dynamic visualization tiling service in VEDA serves shortwave infrared (SWIR) false color composite imagery of NASA's Harmonized Landsat Sentinel-2 (HLS) dataset (Figure 7). Source HLS data within the FIRMS application is served directly from the official NASA archive managed by the Land Processes (LP) Distributed Active Archive Center (DAAC). Integration of the VEDA dynamic tiling service for HLS data visualization within FIRMS demonstrates the power of VEDA leveraging cloud-optimized data formats and aligning with community standards and specifications. Additionally, FIRMS end users are better able to track fire line progression and burned area using the HLS SWIR false color composite imagery.

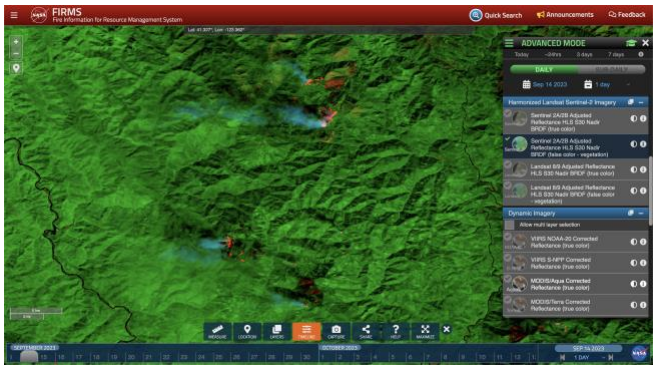

Figure 7. Screenshot of Fire Information Resource Management System home page

4. CONCLUSION

The VEDA platform is a modular cloud cyberinfrastructure that reduces complexities associated with migrating usrs to an ecosystem of cloud-hosted data and cloud computing. Built to community standards and specifications, backend data services within VEDA are interoperable with similar geospatial data catalogs built to similar standards as is demonstrated with the Earth.gov deployment. The VEDA dashboard provides a bridge from the science curious public to research teams through interactive web mapping with intuitive data exploration and analysis and engaging data driven storytelling capabilities. The component-based design of the VEDA platform allows for the rapid deployment of the whole platform or platform components as has been demonstrated within several NASA and U.S. Federal government initiatives.

5. REFERENCES

[1] National Aeronautics and Space Administration, "Science Mission Directorate Policy Document SPD-41a," 26 September 2022. [Online]. Available: https://smd-cms.nasa.gov/wp-content/uploads/2023/08/smd-information-policy-spd-41a.pdf.

[2] Open GeoSpatial Consortium, "OGC Standarsds," [Online]. Available: https://www.ogc.org/standards/.

[3] STAC, "SpatioTemporal Asset Catalogs," [Online]. Available: https://stacspec.org/en.

[4] C. A. Albinet, A. S. Whitehurst, L. A. Jewell, K. Bugbee, H. Laur, K. J. Murphy, Frommknecht, B. Frommknecht, K. Scipal, G. Costa, B. Jai, R. Ramachandran, M. Lavalle and L. Duncanson, "A Joint ESA-NASA Multi-mission Algorithm and Analysis Platform (MAAP) for Biomass, NISAR, and GEDI," *Surveys in Geophysics,* vol. 40, pp. 1017-1027, 2019.

[5] M. Maskey, M. Falkowski, K. Murphy, O. Veerman, R. Mestre, I. Gurung, M. Ramasubramanian, L. Thomas, Z.-F. Yi, D. Bollinger, A. Seadler and Y. Ivey, "Visualizing, Exploring, and Communicating Environmental Effects of COVID-19 Using Earth Observation Dashboard," *2021 IEEE International Geoscience and Remote Sensing Symposium,* pp. 1370-1373, 2021.

[6] National Aeronautics and Space Administration, "Earth Information Center," [Online]. Available: https://earth.gov/.

[7] National Aeronautics and Space Administration, "U.S. GHG Center," [Online]. Available: https://earth.gov/ghgcenter.

[8] National Aeronautics and Space Administration, "FIRMS: Fire Information Resource Management System," [Online]. Available: https://firms.modaps.eosdis.nasa.gov/map/#m:advanced;d:2023-09-13..2023-09-14;l:country-outline,aws-hlss30-false-color,earth;@-123.87,41.47,10.53z.