NASA WRANGLER

Automated Cloud-Based Data Assembly in the RECOVER Wildfire Decision Support System

John Schnase, Mark Carroll, Roger Gill, and Margaret Wooten
NASA Goddard Space Flight Center, Greenbelt, MD 20771

Keith Weber, Kindra Blair, Jeffrey May, and William Toombs
Idaho State University GIS Training and Research Center, Pocatello, ID 83209

ABSTRACT

NASA Wrangler is a loosely-coupled, event driven, highly parallel data aggregation service designed to take advantage of the elastic resource capabilities of cloud computing. Wrangler automatically collects Earth observational data, climate model outputs, derived remote sensing data products, and historic biophysical data for pre-, active-, and post-wildfire decision making. It is a core service of the RECOVER decision support system, which is providing rapid-response GIS analytic capabilities to state and local government agencies. Wrangler reduces to minutes the time needed to assemble and deliver crucial wildfire-related data.

Index Terms— cloud computing, decision support

1. INTRODUCTION

Each year wildfires consume an average of 4.2 million acres of land in the United States, according to the National Interagency Fire Center. The long-term recent decade average is even higher [1]. Fire suppression activities have been employed since the early 1900s to preserve land and protect people and infrastructure. National coordination of fire suppression activities between federal agencies is performed by the National Interagency Fire Center in Boise, Idaho. Fire management begins when a fire is named and an incident command team is assigned; it progresses through the stages of fire suppression including initial and extended attack, followed by containment, control, and extinguishment. If necessary, when the fire is contained, burned area emergency response (BAER) teams may be assigned. These teams have seven days to make an assessment of post-fire conditions and develop a preliminary stabilization and rehabilitation; they have 21 days to submit a final plan once the fire is controlled [1].

Over the past two decades, major advancements have occurred in remote sensing technologies and geographic information systems (GIS). These Earth observational data and software have been employed by fire managers and those who support them to map and characterize fire locations and their extent. These maps can be combined with other geospatial data depicting resources, infrastructure and population centers to identify areas of importance.

The current process of preparing a fire rehabilitation plan typically begins with the BAER team leader requesting a fire severity product. A difference normalized burn ratio (dNBR) layer derived from Landsat imagery is generally the product that is delivered. This product may be integrated with other data, such as topographic information, soil properties, land use, presence of threatened or endangered species, threats to life and property, historic and recent conditions of soil moisture, to create the knowledge base upon which a remediation plan is crafted.

1.1. Challenge Being Addressed

Data assembly, analysis, and decision-making processes must happen quickly in order to meet the statutory requirement of producing a preliminary BAER plan within seven days. However, right now, this process involves substantial human intervention, and the information gathered depends on the availability of staff, time, and data for a particular region. Even though there is a wide array of information services available to the wildfire community, these services tend to focus on research coordination, information sharing, fire risk assessment, active fire management, and fires on forested lands. None of the existing services address the specific needs of post-fire stabilization and restoration planning and monitoring vegetation recovery for semiarid lands.

To assist the effort to manage savanna fires, we are developing an automated decision support system (DSS) called the Rehabilitation Capability Convergence for Ecosystem Recovery (RECOVER) [2]. This system compiles all the necessary datasets for the BAER teams rehabilitation planning and provides them in an easy to use web map interface. In this paper, we describe the RECOVER system and the RECOVER project, focusing particular attention on an automatic, cloud-based data assembly component of the RECOVER system we refer to as NASA Wrangler.

1.2. The RECOVER Project

The RECOVER project is focusing on the restoration of fire-impacted ecosystems as well as post-fire management and rehabilitation. The work is being funded by the NASA Applied Sciences Program. Idaho State University's (ISU) GIS Training and Research Center (GIS TReC) and NASA Goddard Space Flight Center's Office of Computational and Information Sciences and Technology and the Biospheric Sciences Laboratory are the lead organizations.

The primary objective of the RECOVER project has been to build a DSS for BAER teams that automatically brings together in a simple, easy-to-use GIS environment the key data and derived products required to identify priority areas for reseeding and monitor ecosystem recovery in savanna ecosystems following a wildfire.[1] While our specific focus has been on the semiarid regions of the Western US, with RECOVER framed around the problems and challenges faced by the BAER program and the special requirements of post-wildfire decision-making with regard to reseeding in savanna ecosystems, RECOVER technologies are broadly applicable to pre- and active- fire applications and other natural disasters.

1.3. Partner Organizations

Our primary partner organizations to date include the US Department of Interior (DOI) Bureau of Land Management (BLM) and the Idaho Department of Lands (IDL). BLM is the second largest agency in the eight-member National Interagency Fire Center the nation's wildfire coordinating center located in Boise, ID. BLM has operational responsibility for wildland fire on approximately 250 million acres of public land in the US, including approximately 12 million acres, or 22%, of the land base in Idaho. Idaho Department of Lands is the primary state-level agency responsible for dealing with wildfire in Idaho.

The initial stages of the RECOVER project have focused on developing and evaluating capabilities in Idaho. ISU's GIS TReC has over many years developed a close working relationship with BLM and IDL in Idaho, which has created a congenial and highly productive environment for this work. There has been significant involvement of key BLM and IDL collaborators at all stages of the project, and we are working directly with, BLM's National Program Leader for Post Fire Recovery, as well as regional and state coordinators, field office personnel, and incident team leaders. Activities are now expanding to include collaboration with the US Forest Service.

[1] Burned Area Emergency Response (BAER) is the name of the US Forest Service program; the corresponding program within the Bureau of Land Management is named Emergency Stabilization and Rehabilitation (ESR). For simplicity, we use BAER throughout.

2. NASA WRANGLER

The RECOVER DSS comprises servers and clients. Its servers perform two primary functions: they automatically gather site-specific data, and they serve those data to various clients. Currently, RECOVER's most widely-used client is a full-featured Adobe Flex Web Map GIS analysis environment that users can access on a variety of platforms.

Since rapid response is crucial to the RECOVER process, performance of the server's automatic data assembly function has been of particular importance. ISU has implemented these capabilities at GIS TReC using state-of-the-art ESRI technologies, which are particularly adept at serving data to RECOVER's primary client, the Adobe Flex Web Map. In addition, GIS TReC provides the trained technical staff that has been essential to the successful use of the RECOVER DSS over the past two fire seasons.

Complimenting this work, the NASA Goddard team, which has expertise in high-performance computing, has developed an innovative and cost-effective cloud-based data aggregator that further automates the labor-intensive task of "wrangling" data: NASA Wrangler (Fig. 1).

In a typical scenario-of-use, using as input only a site name and coordinates specifying an area of interest, Wrangler uses the rapid resource allocation capabilities of cloud computing to automatically gather various Earth observational and ancillary data products from Internet-accessible data services. Additional data can be added after the initial request if needed, and the entire data collection is refreshed throughout a burn so that when a fire is contained, BAER teams have at hand a complete and ready-to-use RECOVER dataset that is customized for the target wildfire. Wrangler continues to gather data after the fire to support long-term monitoring of ecosystem recovery.

2.1. System Architecture and Implementation

NASA Wrangler has been built as a classic model-view-controller (MVC) application, with some additions and modification that enable system scalability. The MVC design pattern divides an application's objects into three interconnected components. The *model* is the main component. It directly manages the application's data and logic, and it does so independent of the user interface. The *view* can be any output representation of the system's information. Importantly, multiple views of the information are possible. The *controller* accepts input and converts it into commands for model or view objects, and it controls how the application's behaviors are presented to the user.

Wrangler's model component manages *sites*, which are defined by a name and the upper-left and lower-right coordinates of the sites' geographic extent. *Endpoints* represent external, Internet-accessible data sources from which Wrangler can retrieve data. These endpoints are associated with *protocols* that define the communication procedures used by the endpoints' corresponding service. Wrangler currently

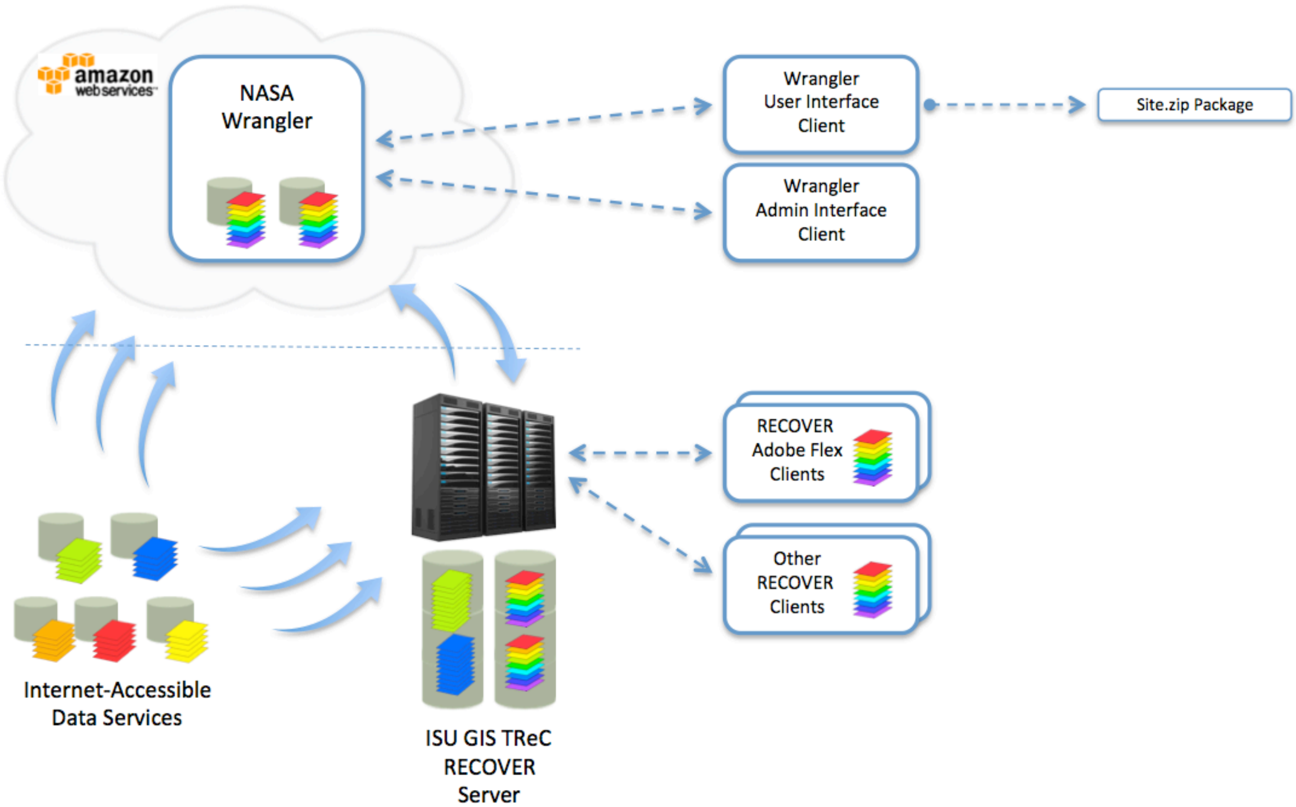

Figure 1. Architecture of the RECOVER decision support system showing component interactions. NASA Wrangler can deliver rapidly-assembled data to other RECOVER servers or allow users to download site-specific data packages for use in other GIS environments.0

supports FTP, WCS, WFS, WMS, and idiosyncratic web services, such as those for Landsat, MODIS, and MERRA, which we describe in greater detail below. Finally, Wrangler's model manages *layers*, which are the data collected from each endpoint for a specific site.

Wrangler's view component currently supports three interfaces: a command-line interface (CLI) that is used primarily for development, a web service application programming interface (API) that enables machine-to-machine communication, and a web interface, which is the primary user interface to the system.

Wrangler's controller communicates with model and view to implement the overall behavior of the system. The controller accepts input from view objects (the interfaces), sends commands to the model to update system state as directed by the system's logic, and sends commands to the view to change the interfaces representation of model state. In practical terms, Wrangler's behavior is very simple: when a new site is saved in Wrangler's database, layers are automatically created for every enabled endpoint. When all the layers for a site have been collected, they can be download from the system or transferred to a calling application. Wrangler was built using Django, an open-source web framework for Python particularly well suited for data-driven applications. User interface components were built using Bootstrap and Leaflet.

Wrangler gathers data in parallel and decouples this process of parallel data aggregation from its interfaces. When a site is created — i.e. when a new site name and spatial extent are added to Wrangler's database — a job daemon launches the various endpoint calls and monitors the arrival of requested data. All the processing is performed by the database subsystem and focuses entirely on responding to data requests, not site management. This decoupling allows sites to be created by various interfaces simultaneously, speeds the gathering of data from multiple sources, and makes it easy to update, re-run, or add layers.

NASA Wrangler has been built entirely in the Amazon Web Services (AWS) cloud. Cloud computing provides agility and cost-savings, because our Amazon cloud servers are an "elastic" resource that can be dynamically created and removed as needed. Another benefit to cloud computing is that Wrangler's compute and storage resources are acquired as an operational cost to the project, rather than through a time-consuming and potentially complex IT procurement.

2.2. Wrangler Data Sources

Earth observations and ancillary data play a crucial role in BAER decision processes. Wrangle's key observational inputs included Landsat 8, MODIS, and AMSR-E. These data are used to develop fire intensity, fire severity, NDVI, fPAR, ET, and many other products of specific use to the wildfire community [3, 4]. In addition to observational data, Wrangler automatically gathers several dozen other data products, including information on the fire site's vegetation cover and type, agroclimatic zone, environmental site potential, fire regime condition class, geology, hydrology, soils, historic fires, topography, and evapotranspiration.

Wrangler automatically assembles parameters that can be important in understanding pre-existing conditions. These include previous years' monthly averages for soil moisture, NDVI, temperature, precipitation, relative humidity, and surface winds. To support long-term monitoring, Wrangler can automatically updates NDVI and fPAR data on a monthly basis post-fire.

The RECOVER project is breaking new ground by introducing reanalysis data into wildfire decision processes. Wrangler is acquiring its historic climatology data from the Modern Era Retrospective-Analysis for Research and Applications (MERRA) collection. MERRA uses NASA's GEOS assimilation system to produce a broad range of climate variables spanning the entire satellite era, 1979 to the present, essentially integrating *the entire NASA EOS suite of observations* into key RECOVER variables [5].

3. DISCUSSION

Data assembly is the most significant bottleneck in wildfire-related decision making. The RECOVER decision support system is transforming this information-intensive process by reducing from days to a matter of minutes the time required to assemble and deliver mission-critical data [2]. NASA Wrangler is a cloud-based implementation of RECOVER's core rapid data assembly capability that offers advantages to RECOVER and other potential applications as well.

Wrangler acquires all its data from original sources and works with most web services as well as ftp protocols. It subsets and transforms data to a common projection and extent on-the-fly to create a coherent stack of GeoTIFF data layers immediately consumable by most GIS analysis environments. It is fully automated: human intervention is not required after the system receives an initial request. It uses only open source software, which means it can be shared with others without complex licensing arrangements.

3.1. Cloud Computing Advantages

Wrangler is a loosely-coupled, event driven, highly parallel data aggregation service designed to take advantage of the elastic resource capabilities of cloud computing. As a result, Wrangler is an efficient, scalable, and cost-effective capability that is particularly well-suited for local deployment in high-performance compute-storage environments or can be provided to others a Software-as-a-Service (SaaS) offering, as we do in the RECOVER project. It also can be conveyed as an intact cloud image. This later option has important implications for technology transfer and funding.

A word of caution: this example represents a system configuration, our experiences, and an Amazon cost structure at a particular time and serves only to make a general point about one of the most important advantages of cloud computing. A Wrangler AWS base server with 100 GB of storage, supporting seven wildfire sites having an average size of 12,000 square KM, each gathering data from 15 endpoints three times a day and downloading several 140 MB data packages per day, has been costing about $375 per month over the past two years. The economic appeal of cloud computing is often described as converting capital expenses (CapEx) to operating expenses (OpEx). This pay-as-you-go stance has greatly benefited our project and enabled an agility otherwise impossible. We believe this holds great promise for cash-strapped management agencies.

3.2. Next Steps

In the coming months, we will complete work on a fully automated means of ordering, downloading, projecting, and subsetting Landsat and MODIS time series data in a single step. This will represent a significant advance in current capabilities. We also hope to expand our partnership to involve other agencies and collaborators.

4. CONCLUSIONS

The RECOVER project provides automated data assembly and web map decision support that enable BAER teams to expedite the process of creating post-fire rehabilitation plans. NASA Wrangler is a cloud-based implementation of RECOVER's core data aggregation capability. Wrangler has become an valuable off-shoot of the RECOVER project and is proving to be a low-cost, highly effective technology for rapid data assembly that is broadly applicable to other data-intensive GIS and remote sensing applications.

5. REFERENCES

[1] National Interagency Fire Center (NIFC). 2017. *www.nifc.gov.*

[2] Schnase, J.L., M.L. Carroll, K.T. Weber, et al. 2014. RECOVER: An Automated Cloud-Based Decision Support System for Post-Fire Rehabilitation Planning. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XL–1, pp. 363–370.

[3] Weber, K.T., S. Seefeldt, C. Moffet, et al. 2008. Comparing fire severity models from post-fire and pre/post-fire differenced imagery, *GIS Science and Remote Sensing* 45(4): 392-405.

[4] Weber, K.T., S. S. Seefeldt, J. Norton, and C. Finley. 2008. Fire severity modeling of sagebrush-steppe rangelands in southeastern Idaho. *GIScience and Remote Sensing* 45: 68-82.

[5] Rienecker, M.M., & Coauthors. 2011. MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, Vol. 24, No. 14, pp. 3624–3648. Available online at *http://dx.doi.org/10.1175/JCLI-D-11-00015.*