10B.3

Using Open Source Software to Deliver Weather Data and Products to NOAA Users

Micah J Wengren*
Office of Coast Survey, National Ocean Service, National Oceanic and Atmospheric Administration (NOAA)
Matthew L. Austin Office of Systems Development, Satellite and Information Service, NOAA
Chi Y. Kang Office of the Chief Information Officer, NOAA

1. INTRODUCTION

In 2011, NOAA funded a research proposal titled “An Enterprise GIS Web Services Hosting Environment for NOAA” as part of the High Performance Computing and Communications (HPCC) Incubator program. The project's primary goal was to answer the question of whether open source geospatial software could be deployed on centrally-managed, enterprise scale system to support the web service hosting needs of NOAA's diverse collection of geospatial data producers. The resulting prototype Open Geospatial Consortium (OGC)-compliant web services hosting platform has been deployed on NOAA web hosting infrastructure since June 2012. Since that date, it has been made available to NOAA data providers to publish their spatial data, thereby allowing them to publish these types of services without needing to maintain the underlying infrastructure themselves.

Experimental services have been created for several of NOAA's critical weather products, including National Hurricane Center tropical cyclone forecasts, and National Weather Service short-duration warnings, among others. Although the system was designed to accommodate NOAA spatial data regardless of domain, it is particularly useful for publishing weather-related data that is constantly changing and requires responsive and reliable update mechanisms.

This paper provides an overview of the HPCC project: hypotheses, goals, and results and findings. It also discusses the larger question of whether and how open source software can be used to solve Information Technology (IT) problems in the Federal IT landscape, particularly in the area of geospatial IT. Examples of experimental NOAA weather data services that were created on the system are used to demonstrate its capabilities.

2. PROBLEM STATEMENT

NOAA produces spatial data in many different domains, touching every area of the agency's matrix of responsibilities and a Federal science and stewardship organization. In the world of Geographic Information Systems (GIS) and spatial data technology, data consumers increasingly expect to be provided ready access to public data for analysis, visualization, and other types of GIS applications. Much in the way the smartphone has raised expectations of users for ease-of-access to timely and relevant information for a particular need at hand, the adoption of web services for data delivery in the spatial data profession has raised the bar in terms of expectations for how a data provider should deliver data to its customers.

NOAA has traditionally left the group of data-originator organizations under its umbrella the responsibility to determine how they should go about publishing their geospatial data and web services to their users. While this is not necessarily an undesirable situation – often those closest to the data are best able to determine how it should be represented and disseminated to users - it has had the secondhand effect of delegating to data providers the burden of procuring, configuring, and managing the Information Technology (IT) infrastructure necessary to meet these data publishing needs.

This approach is inefficient in many ways. Repeated and duplicative deployments of the same or similar software stacks is not efficient; especially when you consider that not all deployments may be scaled appropriately according to demand, and therefore may be leaving unneeded capacity on the table while costing the full amount in terms of power supply, cooling, and physical space required to run the systems. It also necessitates the same – often high-level – technical expertise to be cultivated within each data-producing organization in order to maintain the software and systems.

Although advantageous in terms of allowing autonomy for data producers to publish geospatial web services as they see fit, it can lead to inefficiencies agency-wide, and justifies the need to at least investigate alternatives that data providers might be able to leverage if they lacked the resources to maintain a system on their own, or otherwise chose not to.

The HPCC project discussed in this paper, 'An Enterprise GIS Web Services Hosting Environment for NOAA', was proposed as an alternative to the status quo geospatial web services deployment model that

* *Corresponding author address: Micah J Wengren, Office of Coast Survey, NOAA, 1315 East-West Hwy N/CS 12, Silver Spring, MD 20910; e-mail: micah.wengren@noaa.gov. Twitter: @micahwengren*

existed in NOAA at the time.

3. PROJECT GOALS

The primary goal of the 2011/2012 HPCC project 'An Enterprise GIS Web Services Hosting Environment for NOAA' was to answer the question:

> *Can freely-available open source software be leveraged to deploy a scalable, centrally-managed hosting system that allows NOAA data providers to publish standards-compliant web services of their geospatial data, without being responsible for maintaining the system themselves?*

Of particular interest to the primary investigators was whether the capabilities that existed in some popular and widely-used open source geospatial software packages would meet the needs of a Federal agency such as NOAA, in an enterprise-scale deployment.

Open source software offers several advantages over more typically feature-rich commercial software that made it a good choice for addressing the problem statement.

First, without the costs and procurement requirements of licensed, commercial software, a system that is entirely open source can be scaled out more easily and without proportional cost increase for software licensing. The only cost to scale is for underlying hardware required.

Second, functionality can be added directly to the code base if needed, since it is available for free download and generally can be modified at will. With sufficiently experienced developers, new features can be implemented or bugs can be patched as needed, and not at the mercy of a patch cycle dictated by the commercial software vendor. This can add a lot in terms of efficiency and responsiveness in system deployment.

Third, established open source software almost always exists alongside and as a result of a robust and active developer community, so technical support for issues small and large can be easily obtained, and at no cost.

Fourth, professional-level support is available for many open source software projects. In some cases professional support will include direct access to lead developers, or core-committers, of a project, so it can be very efficient and effective. In addition, professional support in the form of direct feature enhancements to the core software can often be procured. This allows organizations willing to make a direct investment in the improvement of the software the ability to influence the project roadmap according to their needs. Even better, the added transparency and inclusiveness found in the open source software project steering process can allow organizations will similar needs to pool resources to sponsor functionality enhancements that benefit both themselves as well as the community as a whole. The return on investment therefore can be substantially greater than closed-source software, as features that are paid for once are then available free of charge for use by all.

For these reasons and others, including the effect of decreasing allocations for IT spending across the Federal government, the project team focused on the exclusive use of open source geospatial software in the target system design.

4. PLANNING, DEVELOPMENT, AND PROTOTYPE DEPLOYMENT

At the outset of the HPCC Incubator period in August 2011, the project team identified a target open source software stack they determined to be best-suited to succeed in answering the problem statement. This included the following:

- Operating System: Debian Linux
- Web Server: Apache HTTPD and Apache Tomcat
- Runtime Environment: Java 6
- Geospatial Server: GeoServer
- Spatial Database: PostgreSQL with PostGIS spatial extension

The core of the hosting system is GeoServer, a Java-based web application that provides OGC-compliant web services for spatial data that is uploaded and stored, configured, and styled within its internal data catalog, or inventory. GeoServer requires a Java Servlet container to run, and Apache Tomcat was selected to serve this purpose. For data storage, the open source relational database PostgreSQL was selected with the PostGIS spatial extension.

Based on the above target software stack and knowledge of its existing capabilities, a gap analysis was performed between the current capabilities and what was expected to be needed in order to be deployed on an enterprise, scalable fashion at NOAA. The requirements of the infrastructure providers who would host the prototype, NOAA's Web Operations Center (NOAA WOC), were also factored in the analysis.

A request for proposals and associated procurement paperwork were prepared, and a contract was awarded to OpenGeo (now Boundless) in September 2011 to provide feature enhancement development for GeoServer and consulting in system deployment.

The gap analysis performed on GeoServer's capabilities revealed that, in order to meet NOAA's needs for a centrally-managed hosting system it would require enhancements in the following areas:

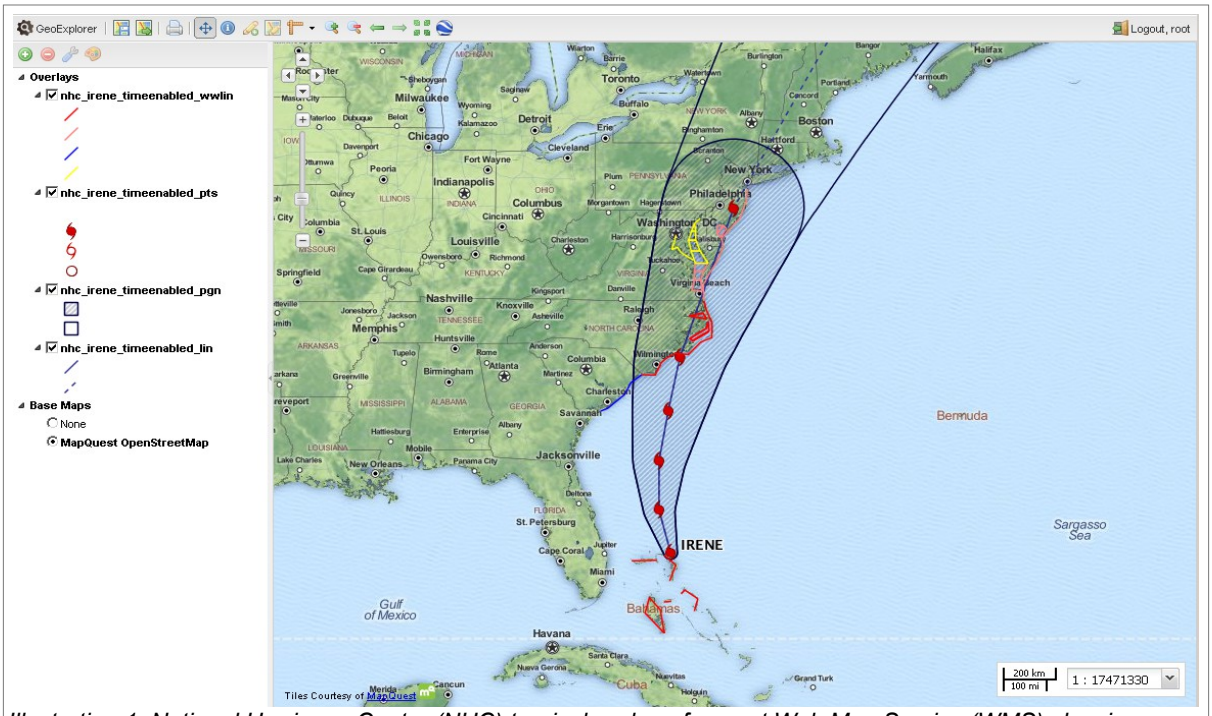

Illustration 1: National Hurricane Center (NHC) tropical cyclone forecast Web Map Service (WMS) showing Hurricane Irene from August 2011. The prototype WMS contains the official forecast track, points, cone of uncertainty and watches/warnings and is updated as new forecasts are released by NHC.

- Multi-tenancy: allowing several distinct organizations to upload data and maintain exclusive access and control to their data. Also, they needed to have the ability to customize OGC service endpoints to reflect their individual organizations (contact information, service descriptions and keywords).
- Authentication: the ability for GeoServer to connect to an external authentication provider: Lightweight Directory Access Protocol (LDAP), in particular.
- Groups: support for the notion of groups of users within GeoServer, with a subset of permissions to access GeoServer configurations menus via the web user interface (Web UI) and the ability for a 'group administrator' to control their group's membership.

These requirements were provided to Boundless at the outset of the performance period of the contract. NOAA's requirements – in particular the authentication enhancements – were in-line with some existing development work that had been done on GeoServer. This expedited the delivery of this new functionality to NOAA, and the contract with Boundless provided a targeted use-case that could be used to improve and finalize prior development work on an enhanced security subsystem for GeoServer that provided LDAP authentication. NOAA's requirements for the other two major enhancements, multi-tenancy and groups with appropriately limited access to Web UI controls, were also more broadly needed in the GeoServer user community, so there was advantageous alignment there as well.

Development work by the contractors was performed for the first six months following the contract award, and custom builds of GeoServer that included the extended functionality were delivered to NOAA for verification on the NOAA WOC infrastructure. An iterative feature testing and feedback process between the NOAA project team and the contractors refined the functionality implemented in the custom builds.

When the full functionality enhancements were accepted by NOAA, a final custom GeoServer build was delivered, and the NOAA WOC staff configured and deployed the full software stack including GeoServer, PostgreSQL/PostGIS, and other software on a redundant, high-availability cluster at two of its data centers in Largo, MD and Boulder, CO. Final deployment of the prototype was completed in June 2012.

Since NOAA's requirements for GeoServer enhancements for this project were closely in-line with needs within the broader GeoServer user community, all of the features included in the custom GeoServer build provided by Boundless to NOAA were subsequently merged back into the core GeoServer code base and included in the official GeoServer 2.2 release

(September 2012)[1]. The process to add functionality to GeoServer is known as the GeoServer Improvement Proposal process (GSIP), and the NOAA HPCC project contributed to GSIP #'s 66,67,71,73, and 74[2].

This was highly beneficial to NOAA, because the initial custom GeoServer build deployed at the NOAA WOC could be swapped out for future releases of the standard GeoServer software without any lost functionality. NOAA, as well as any other Federal agency or organization using GeoServer, will benefit by having these enhancements added to the freely-available GeoServer product. NOAA also does not have the burden of maintaining a custom build in the future for ongoing support of the hosting system, an important factor that could be critical to the success or failure of other custom open source development projects such as this.

5. DATA PROVIDER TESTING

NOAA data publishers were given access to the system following the June 2012 deployment. Publicizing the system was done with the help of the NOAA GIS Committee, a group responsible for coordinating GIS and geospatial efforts across the agency, as well as through other channels including NOAA mailing lists. An agreement with the NOAA WOC provided support for an evaluation phase until September 2013, initially (it has since been extended until September 2014).

There was significant interest by users in testing the system during the prototype evaluation phase of the project. Data providers from the National Ocean Service (NOS), NOAA Fisheries Service (NMFS), NOAA Satellite and Information Service (NESDIS), and National Weather Service (NWS) signed on for trial accounts.

Many experimental OGC-compliant Web Map Services (WMS) and Web Feature Services (WFS) were set up during the prototype evaluation phase. The following weather-specific data services were deployed to the system:

- National Hurricane Center (NHC) tropical cyclone forecasts
- NHC tropical cyclone 'preliminary best track' tropical cyclone tracks
- NWS short duration warnings
- NWS Climate Prediction Center short term temperature and precipitation outlooks
- NWS Storm Prediction Center severe weather reports for tornadoes, hail, and wind
- NWS Advanced Hydrological Prediction Service (AHPS) river flood observations and forecasts

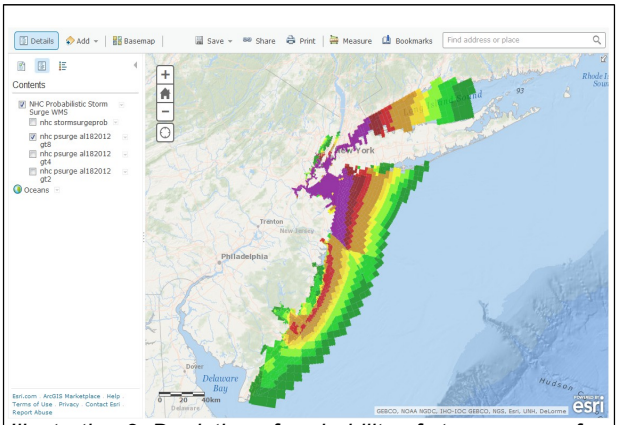

Illustration 2: Depiction of probability of storm surge of eight feet or more during Hurricane Sandy (October 2012). This prototype WMS includes probabilistic storm surge forecasts released by the NHC during land-falling tropical cyclones.

These services were made available publicly, and many were added to the NOAA GeoPlatform site – a cloud-hosted data catalog and mapping platform NOAA contracted with Esri to provide that supports external WMS services. The NOAA GeoPlatform catalog was harvested by the Federal Data Catalog website[3], that now provides public discovery of the services via its search interface. The NHC tropical cyclone service was incorporated into NOS' Environmental Response Management Application (ERMA) Atlantic regional website[4].

Additionally, the hosting platform was used to host web services for a project to demonstrate data interoperability for the NOAA Climate Portal[5] during 2013. The data interoperability pilot project demonstrated the use of OGC-compliant services to allow users to find, display, manipulate and download NOAA and partners' climate data hosted and served from different data centers[6]. The project developed a web-based prototype portal that could be used to interact with such services in a map-based environment[7].

1 http://blog.geoserver.org/2012/09/21/geoserver-2-2-released/

2 http://geoserver.org/display/GEOS/GeoServer+Improvement+Proposals

3 http://catalog.data.gov/

4 http://response.restoration.noaa.gov/maps-and-spatial-data/environmental-response-management-application-erma/atlantic-erma.html

5 http://www.climate.gov/

6 http://www.cpc.ncep.noaa.gov/products/outreach/workshops/CDPW38/4_Thursday/1_Morning/Shresth a.pdf

7 http://wiki.esipfed.org/images/b/b0/AbstractforESIPDataInteroperabilityTalk-May2-2013.pdf

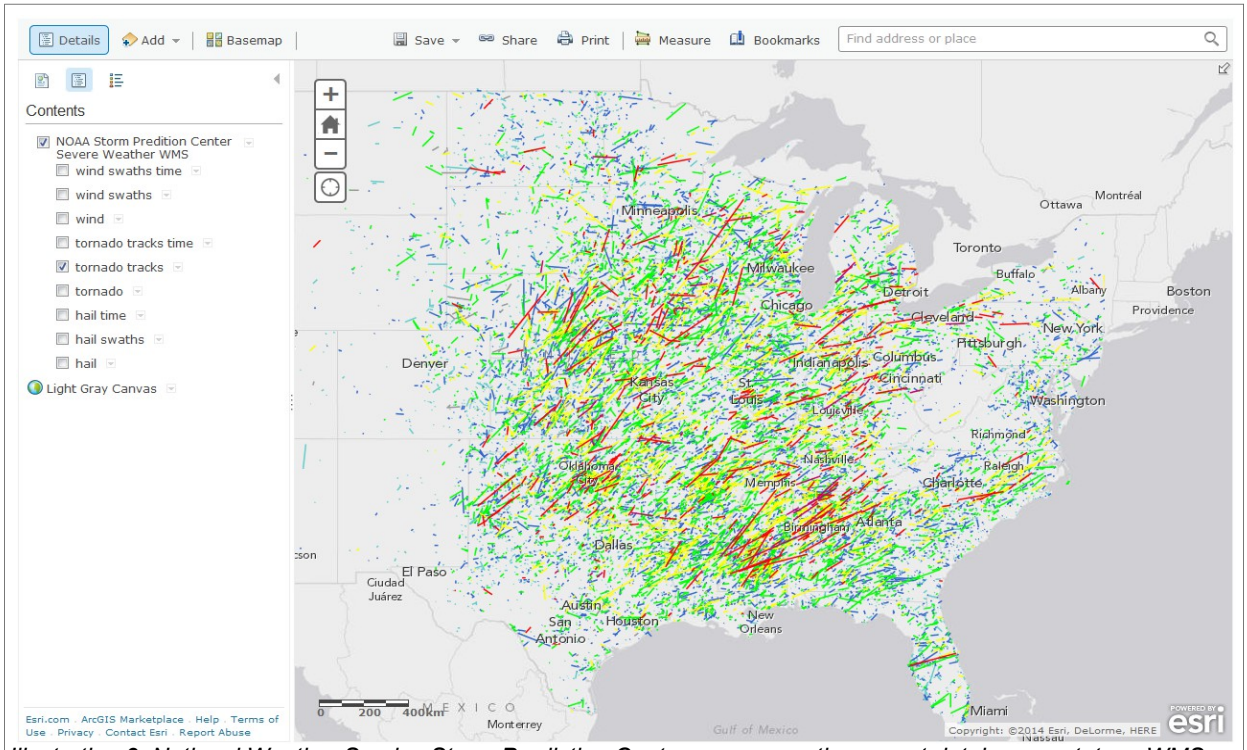

Illustration 3: National Weather Service Storm Prediction Center severe weather report database prototype WMS depicting the full record of reported tornado tracks for 1950 through 2012. The service includes WMS 1.3.0 time dimension support for time-based animation of tornado track features. Map displayed on Esri ArcGIS.com website.

6. CONCLUSIONS AND FURTHER WORK

As mentioned previously, the HPCC prototype hosting system trial is ongoing, and is expected to conclude in September 2014. Before that time, feedback will be requested from data providers and a summary of findings created. In general, participants will be asked questions along the lines of "did the hosting system meet your expectations for a shared NOAA OGC services hosting platform," and "what aspects of the system do you feel worked well, and what are the areas with room for improvement?". Usability of the system will be a particular focus, as that was viewed by the project team as a potential barrier to adoption by data providers.

In general, the HPCC project investigator team believes the effort was a success for several reasons. Regardless of usage by NOAA data providers, the initial design objectives were achieved and functionality delivered was consistent with the original proposal. The investment made by NOAA's HPCC committee resulted in several improvements to the core software component of the system, GeoServer, as highlighted earlier. Additionally, because these were implemented in the core GeoServer, NOAA can utilize them for future GeoServer projects without maintaining a custom fork of the software. Lastly, several prototype services of critical NOAA geospatial data products were created and made available for public evaluation and use, and also were incorporated into NOAA web-based mapping applications and the Federal Geospatial Data Catalog.

There were some notable challenges encountered in the testing period. Despite not yet having gathered feedback from test users, a few general issues can be identified. These were mostly related the high learning curve required for users who were unfamiliar with using GeoServer.

The main shortcoming of the overall system had to do with the difficulty presented when loading geospatial data into it. Without a desktop GIS companion to the GeoServer system, as well as a fairly limited web-based data upload capability, the process of loading data requires development of scripts or programs that can interact with GeoServer's Representative State Transfer (REST) Application Programming Interface (API). This extra step alone is enough to turn some users away. When compared with more established and tightly coupled commercial GIS software options, where a user can manipulate a desktop Graphical User Interface (GUI) to configure their data layers, associated styling and other rules, and then seamlessly publish the entire package to a remote server system for web service deployment, making the switch to a GeoServer-based system that is much less polished is not always appealing. This shortcoming turned some test users during the prototype evaluation period away from using it very heavily.

It should be noted that in 2013 a plugin was developed for the open source Quantum GIS (QGIS) desktop GIS system that provides some of these desktop GUI-based capabilities for GeoServer, including the ability to create Styled Layer Descriptor (SLD) styles for layers within a GUI environment. If this plugin had been available early in our testing period we might have had much greater adoption by users and the impact described above may have been greatly reduced.

A second notable limitation was that the HPCC system did not allow raster geospatial data formats. This was due to the versions of PostGIS and PostgreSQL available in the Linux distribution the system was deployed to. At the time it was deployed in June 2012, PostGIS 2.0, which supports internal database storage of rasters, had not yet been released.

The third limitation worthy of mention was the necessary restrictions enforced on ways to upload data into the system. Separate from the issue of requiring data publishing scripts, this functionality limitation was more a result of the enterprise deployment pattern used. Because the system was deployed in NOAA's WOC web hosting data center, users could not be allowed to have any system-level access, including direct transactional write access to the PostgreSQL database via firewall hole, proxy, VPN or some other means. Only the GeoServer software itself was allowed write access to the database, and the only exposed TCP ports were those used by GeoServer, Hypertext Transfer Protocol (HTTP) on port 80, and HTTPS on port 443. This is a necessary limitation for any secure enterprise-scale service, regardless or software used. Despite the GeoServer REST API's highly capable handling of automated data publishing workflows over HTTPS, the lack of direct database write access still resulted in limitations on how frequently data could be pushed into the system. This meant that real-time data writes to the hosting system weren't possible, for example if a client required a 3rd party application that would have permissions to directly modify an underlying database table of a service. An example use case for this would be to connect NWS' AWIPS software directly to the hosting system so forecasters could commit changes that directly updated a live service.

Notwithstanding the general issues discussed above, as well as and additional specific feedback from system users at the end of the test period in September 2014, the HPCC OGC web services hosting platform has been successfully configured, deployed on a centralized web hosting platform at the NOAA WOC for nearly two years of testing, and has demonstrated the ability to host live prototype web services of many of NOAA's critical weather data products. NOAA's investment of HPCC resources towards the project have been successfully invested in both supporting a prototype shared-service model for geospatial web services, and contributed to the direct enhancement of the GeoServer software for the benefit of all users who download and deploy it for free. The project team believes NOAA should investigate investment in open source geospatial software for future projects of a similar nature.