NACEPT Vulnerable Populations Recommendations
Cross-Cutting Information Technology Capabilities/Architecture

Karl Benedict, kbene@edac.unm.edu
Director, Earth Data Analysis Center
Research Assistant Professor, Department of Geography and University Libraries
University of New Mexico

Summary for Integration into full letter

In implementing IT capabilities in support of the monitoring, reporting, and mitigation activities defined the the scenarios described in this letter a key set of general requirements have been defined (described in greater detail in Appendix XXX). These requirements relate to the use of open interoperability standards to streamline both collection of measurements being generated by monitoring systems, and dissemination of data products derived from those systems. These standards range from general-purpose web services based upon the REST web service model (which in turn is based upon the HTTP standard protocol), to the suite of more specific open standards from the Open Geospatial Consortium (OGC) relating to data visualization (Web Map Service - WMS), data access (Web Feature and Web Coverage Services - WFS and WCS respectively), and sensor control and communication (Sensor Web Enablement - SWE).

These services are the key components in the development of a services oriented architecture (SOA) that

I. lowers the barriers to data acquisition - decreasing the time required for collected data to be entered into the core management systems.

II. Provides a logical separation between internal data management systems and the clients that consume products based upon the contents of that system.

III. Publication of standards-based services that may be both used by EPA developers to provide specialized data access and visualization tools, but *also* may be used by external developers to provide custom *mashups* in support of specific user communities - particularly vulnerable populations*.*

EPA has initiated a number of programs that are developing these capabilities: EPA's "Apps for the Environment" challenge", "Environmental Dataset Gateway", "Geospatial Data Download Service", "National Geospatial Program" are all examples of programs that are making use of this SOA approach. What is needed within EPA's IT planning is a routine consideration and assessment of where interoperable services may be integrated into the development of new capabilities or updates to existing ones.

Needs Identified in Individual Scenarios

Assessment of the various information technology (IT) capabilities applicable to the outlined scenarios yields IT requirements that primarily fall into four categories: monitoring, reporting/communication, mitigation, and remediation. The recommended capabilities interconnect and have a common set of underlying IT architectural components that may support elements of all four requirement areas. The first section of this document outlines the requirements that have been extracted from the developed scenarios while the second section outlines a recommended IT architectural model that may address the defined requirements.

Monitoring

The identified monitoring needs discussed in the provided scenarios highlight a requirement for continuous monitoring from both fixed and mobile platforms/sensors, with the sensor systems ideally being low in cost (allowing for more broad installation and distribution to community members), and, in the case of community participation monitoring, aligned with the capacity and capabilities of the community members engaged in the monitoring effort.

In the case of community deployed monitoring or data acquisition systems (i.e. mobile applications, sensor packages that may be 'checked-out' by community members for use, in-situ monitors installed in community member defined locations), the systems must be designed with an understanding of the usability of the provided systems for particular users, and the technical capabilities available to those users. For example, while in-situ mold detection systems might be designed with integrated data communication capabilities (through connection with a local wireless data network), those capabilities will be useless in remote areas (such as some tribal lands) where there is no Internet or wireless data connectivity available.

Reporting (Communication)

A common theme in many of the provided scenarios is one of real-time, or near-real-time availability of measurements being collected by monitoring systems. This requirement has significant implications for how the data communications infrastructure (both for ingest of data from sensors, and for delivering data) is developed. If the alerting systems highlighted in a number of the scenarios are to be effective, the data needed to trigger the alerts needs to be available in a timely manner, typically defined in terms of minutes instead of hours or longer. This, in turn, suggests that data ingest, processing, and delivery systems need to be as automated as feasible, eliminating the inevitable time lag involved when human action is required within data processing or delivery workflows. While recognizing that there will be an absolute need for human intervention in some processing and alert activities, minimizing (through the definition of appropriate thresholds, documentation, or instruction for the use of provided products) the time from data ingest to availability is a key requirement for providing useful alerts to vulnerable communities.

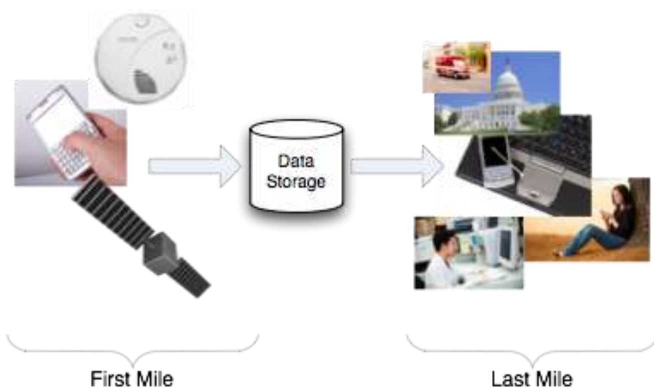
Mitigation

A side benefit of the rapid ingestion and processing of sensor data (described above) is the opportunity to have a more detailed picture of an evolving situation (increasing situational awareness), allowing for a more effective response during an incident, potentially mitigating some potential effects through more rapid decision making. The rapid availability of data, and the ability to inform decisions about the deployment of additional data collection resources (i.e. the development of an enhanced data collection network in a specific location) can improve both during-incident response, and post-incident **remediation** and analysis activities.

All of the above described requirements have implications for the supporting cyberinfrastructure that can enable the needed capabilities.

Translation of Described Requirements into IT Capabilities (Cyberinfrastructure)

Two key challenges that must be met when attempting to address the requirement outlined above, may be described in terms of the *first* and *last mile*, where the *first mile* relates to the capacity to rapidly ingest *data* from a wide variety of sources, and the *last mile* relates to the ability to deliver *data* and *information* to a wide variety of users and platforms.



In both cases, web services represent the most common model for rapid ingest and delivery of data and information. While a variety of general purpose web service standards or models have been developed over the past decade, the REST-based[1, 2] model, based upon the HTTP protocol[3], has become the dominant model for publicly accessible data and general purpose services (e.g. Yahoo's geocoding service[1], Google Maps[2], Amazon Web Services[3]). In the realm of geospatial data and information and sensor interactions, the service standards developed by the Open Geospatial Consortium (OGC[4]) are key standards that continue to increase in their adoption. Of particular interest in the context of the scenarios developed herein are five OGC standards (or standards group in the case of SWE):

1 http://developer.yahoo.com/geo/placefinder/guide/
2 http://code.google.com/apis/maps/documentation/javascript/
3 http://aws.amazon.com/
4 http://www.opengeospatial.org/

- Web Map Service (WMS[4]) - a geospatial data *visualization* standard that supports the delivery of requested map images via HTTP
- Web Feature Service (WFS[5]) - a geospatial data *delivery* standard that is designed for the delivery of *features* (typically points, lines, or polygons) and their associated attributes via HTTP
- Web Coverage Service (WCS[6]) - a geospatial data *delivery* standard that is optimized for the delivery of *gridded (i.e. raster)* data via HTTP
- Catalogue Service (CSW[7]) - a data discovery service that is based upon *documentation* (metadata) associated with specific data products
- Sensor Web Enablement (SWE[8]) - a family of standards that define data access, and interactions with sensor systems.

It is through the development and deployment of services based upon the REST web service model, and the OGC standards that a flexible and scalable services-based architecture (SOA) may be developed that enables the development of new and innovative applications (both within EPA and in the broader community) that are based upon a core set of data and services.

EPA is already moving forward in the development of capabilities in this area, particularly through the following published data resources and related public outreach efforts:

- EPA's "Apps for the Environment" challenge[5] - an open competition for the development of creative and innovative applications that make use of published EPA data and services
- EPA's "Environmental Dataset Gateway"[6] - a metadata catalog and associated web services (including REST-based services) for interacting with EPA's metadata collection
- EPA's "Geospatial Data Download Service"[7] - a resource for downloadable geospatial data products
- EPA's "National Geospatial Program"[8] - an access point for a number of OGC and other web services based upon an older ArcIMS and a new ArcGIS server platform.

These initiatives within EPA indicate the the recommended service models included here are already under development with the Agency, providing an opportunity for the EPA to leverage its growing capability in the area of web services to expand the exposure and access to other data products and services that are not currently available through these interoperable interfaces.

Specific capabilities that may be enabled through these web services include:

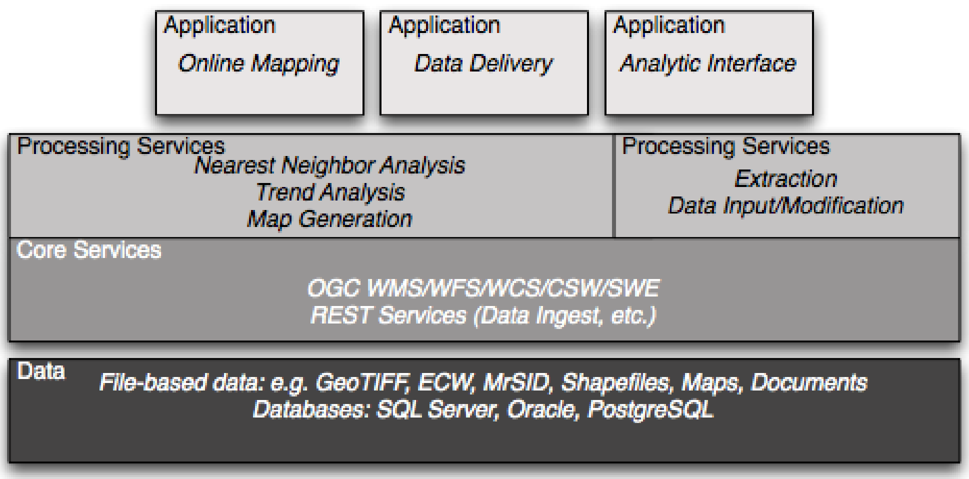
---

[5] http://www.epa.gov/appsfortheenvironment/
[6] https://edg.epa.gov/metadata/catalog/main/home.page
[7] http://www.epa.gov/enviro/geo_data.html
[8] http://www.epa.gov/geospatial/data.html#other

- Rapid visualization of current measurements captured by sensor system via WMS services integrated into platforms such Google Earth, or mobile apps
- Streamlined data ingest from sensor systems via REST-based or SWE service interfaces
- Automated validation of raw data coming from sensor systems to develop higher-level products that have had some defects (such as anomalous readings, missing data, etc.) eliminated prior to publication as publicly available data
- Data access (via WCS, WFS, or REST) by community developed alerting systems that evaluate current measurements from continuous sensor streams that, if a threshold is crossed, automatically generate alerts that community organizations or individuals can receive
- "Crowd-sourced" data collection via web services accessed through web applications or mobile apps, producing rapidly integrated products (an excellent example of such a system is the Ushahidi[9] platform)

Tiered Architectural Model

The above described services are most productively deployed within a tiered SOA in which there is a logical separation between system components, allowing for the differential rates of system evolution that is inherent in the development of integrated systems. In it's simplest form, the recommended SOA consists of three tiers: a data management tier, a services tier, and a client interface (or application) tier.



This model allows for the common situation in which the data management tier tends to change very slowly in terms of the core technologies employed (i.e. relatively stable database-base or file-system based models), while the open standards-based services tier tends to evolve more quickly in response to updates to the standards or demands for new products. Given the rapidly changing environment of web-based applications, the client tier (even when considering desktop applications such as Geographic Information System [GIS] clients) exhibits frequent changes in user expectations and employed technologies. Given these differential rates of change between these tiers,

[9] http://www.ushahidi.com/

developing a system in which the connections *between* the tiers are based upon a limited number of open standards will allow for updates within a tier without necessarily requiring changes to the neighboring tiers.

For example, if a long-term collection of sensor observations is stored in an Oracle database and there was a desire to migrate to PostgreSQL as a database platform, that migration would entail the database migration itself and a requisite update to any services that connect to that database using Oracle-specific drivers. From the perspective of any client applications or data ingest processes that make use of standards-based services the database the changeover would be invisible.

Likewise, Twitter is a current platform that is currently *popular* for broadcasting information to users that are interested in specific information. Six months from now a new social media platform could emerge as a new medium for broadcasting information (such as incident alerts). If a general model for broadcast services was adopted where a data access REST service were developed and broadcast services made use of that data service to emit system specific (e.g. Twitter, Google+, Facebook, SMS) messages, the implementation of support for new services would only require the system specific service development with no required changes to the capabilities of the data or services tiers.

Authentication & Authorization

While recognizing that some data products and information will necessarily be limited to specific users and groups, the general trend has been towards the lowering of barriers for the use of published Application Programming Interfaces (APIs) when the desire is to encourage as broad adoption as possible. In terms of the web service interfaces that are recommended herein, it is suggested that whenever possible the data and data services be made widely available, with access limited through industry standard security models (e.g. HTTPS authentication, public-key encryption, shared-keys, etc.) only when absolutely necessary to ensure data integrity, availability, and appropriate use.

Bibliography

1. Richardson, L. and S. Ruby, *RESTful Web Services*2007, Sebastopol, CA: O'Reilly Media, Inc. 446.
2. Fielding, R., *Architectural Styles and the Design of Network-based Software Architectures*, in *Information and Computer Science*2000, University of California, Irvine: Irvine, CA. p. 162.
3. Fielding, R., et al. *Hypertext Transfer Protocol -- HTTP/1.1 (RFC 2616)*. 1999 5/20/2011]; Available from: http://tools.ietf.org/html/rfc2616.
4. de la Beaujardiere, J., ed. *OpenGIS Web Map Server Implementation Specification, Version 1.3.0*. Vol. OGC® 06-042. 2006, Open Geospatial Consortium. 85.
5. Vretanos, P.A., ed. *Web Feature Service Implementation Specification, Version 1.1.0*. Vol. OGC 04-094. 2005, Open Geospatial Consortium. 117.
6. Whiteside, A. and J.D. Evans, eds. *Web Coverage Service (WCS) Implementation Specification, Version 1.1.0*. Vol. 06-083r8. 2006, Open Geospatial Consortium. 129.
7. Nebert, D., A. Whiteside, and P.P. Vretanos, eds. *OpenGIS® Catalogue Services Specification (OGC 07-006r1)*. Version 2.0.2, Corrigendum 2 Release ed. 2007.
8. Open Geospatial Consortium, *OpenGIS SWE Service Model Implementation Standard*, 2011, Open Geospatial Consoritum. p. 155.