



Meeting • • • • Many Needs

Geoportal aids users and contributors

By Jessica L. Zichichi, Vice President, Geospatial Solutions Group, Innovate!, Inc., and Michelle A. Torreano, Agency Geospatial Metadata Coordinator, US EPA

A comprehensive geospatial metadata sharing and management framework implemented at the US Environmental Protection Agency (EPA) is promoting transparency, participation, and collaboration in government while meeting user, agency, and federal requirements.

A Gateway for Contributors and Consumers

The GDG is the central gateway to EPA's geospatial resources. It enables data consumers to discover, view, and access geospatial resources (e.g., data, services, or applications). Likewise, it enables EPA data contributors to publish and manage their geospatial information for others to use.

GDG maintains a repository of metadata about the geospatial resources users are seeking to discover or make available. Metadata records in the repository describe the nature and scope of registered data items or services and provide information that enables users to locate and view them. Metadata descriptions also provide information necessary to access items and services for use in other systems or mapping applications.

The search widget lets users search GDG and obtain formatted results within their own applications.

Hundreds of agency geospatial resources are available to the public from the GDG. These resources are added and managed in GDG by EPA regions, programs, and labs from across the agency—organizationally and geographically—using a defined governance structure.

A Comprehensive Framework

GDG was implemented as a comprehensive geospatial metadata sharing and management framework to meet multiple user, agency, and federal requirements. Since geospatial data has a prominent role in government decision making, EPA thought this data should be maintained as a shared enterprise resource that would provide end users with multiple benefits. For example, properly documenting and managing agency geospatial resources would ensure that federal metadata maintenance and sharing requirements were met.

Creating GDG also allowed the EPA to have a single access point for these resources and provided users with a comprehensive view of the agency's geospatial assets. This reduced the time users spent searching across multiple locations and data redundancy by connecting data producers with consumers, thus preventing duplication of effort.

It also has allowed contributors to reuse information to support their mission goals by embedding content directly into their applications. This not only saves time and money but also provides better consistency across the agency. This catalog also streamlines EPA's contributions to external metadata sharing portals by being the single source for contributions to Geospatial One-Stop (GOS) and Data.gov.

Serving Users

EPA generally groups GDG functions into two categories: functions that support GDG consumers and functions that support GDG contributors.

Consumers can access GDG resources in a number of ways. One of the most common methods is a simple keyword search using the tool located on the GDG home page for accessing GDG resources. This tool returns an interactive list of matching metadata records from the GDG catalog.

Users can hover over list items to see more information about a resource and view its footprint in the GDG map viewer. Data can then be previewed or accessed, other details on the resource viewed, or a search further refined.

With the advanced search features in the same interface, users can limit results by specific content type, data category, or time period. Advanced users may also perform more targeted searches in this same interface employing Lucene search syntax that allows the specification of additional operators or metadata fields. The combination of these search methods lets users search for exactly what they want.

Consumers can easily embed results into web applications or pages using the GDG REST interface. Links to REST outputs are displayed for users each time they perform a search so REST URLs can be reused.

Consumers can access GeoRSS, HTML, or KML outputs. REST URLs can be configured using additional parameters to narrow or expand the search results to create outputs in a wide variety of metadata views and access points to suit a variety of needs.

The search widget, another useful feature, lets users search GDG and obtain formatted results from within their own applications. The search widget can be embedded simply by copying two lines of code into a web page.

Serving Providers

The GDG metadata management suite gives contributors tools to manage content. Contributors can control how and how frequently (where applicable) records are contributed to the central catalog, control how they are furnished to different types of users, and control which records belong to which user groups. GDG contributors can contribute metadata using either automated (harvesting) or manual (upload) methods. Once metadata has been contributed, users can log in and manage individual records. A key function of metadata management in GDG is the ability of users to identify which records should be made available to the public (unrestricted) and which should be labeled as internal only (restricted). Each contributor can apply this setting to individual records. Records identified as unrestricted are made available to all general public users through GDG and are automatically contributed to GOS and Data.gov. Restricted records are only available to GDG consumers that have logins.

Meeting Specific Agency Needs

Although EPA was able to utilize many of the basic features provided by Esri's Geoportal extension for ArcGIS Server software, the agency also worked to customize the application to best meet its needs. One key customization was to structure GDG access and metadata management using the agency's central Identity and Access Management (IAM) system.

Rather than setting up separate logins and passwords for GDG, EPA used the agency's enterprise IAM system to control user logins, group membership, metadata ownership, and access. Using this system automates group membership for basic metadata access for EPA employees and contractors and reduces tasks associated with ID and password management for GDG administrators as well as reducing the number of logins and passwords that EPA staff must memorize.

EPA structured groups at the IAM system to control how metadata is managed. Each GDG contributor is assigned to a particular regional or program office group to oversee metadata management for that group (e.g., EPA's Office of Water [OW]). Individuals are also assigned to functional groups (e.g., GDG reader, GDG steward, or GDG administrator) that control functional rights. When users log in at the GDG, they are provided with the appropriate metadata management functions so they can modify records based on their group membership. GDG contributors can belong to as many groups as necessary.

The intranet-based REST URL is another customization that automates authentication for EPA intranet users. GDG's default REST interface honors metadata access levels. If a user is not logged into GDG, they can view only unrestricted records when using the basic REST URL. The intranet-based REST URL allows users with access to the EPA intranet to view the full set of GDG metadata records (restricted and unrestricted) without logging in separately at GDG. This lets EPA intranet users have access to the full set of their user groups' records at the GDG using the REST interface from within other intranet web applications and web pages. Because the intranet URL is only available to those who can access the EPA intranet, this provides EPA with a secure way to provide access to all GDG metadata records using the power of the REST interface while maintaining the security of the original Internet-based REST URL, which requires authentication.

Finally, a harvesting extension tool called WAFer provides EPA with a mechanism to pull records directly from contributors' ArcSDE databases. Many EPA personnel wanted a direct connection so metadata →

The screenshot displays the EPA Search interface. At the top, there is a search bar with the text 'toxic release' and buttons for 'Search' and 'Clear'. Below this, the search criteria are set to 'This Site (the GDG)' for 'Search In:', 'Any' for 'Content Type:', and a list of categories for 'Data Category:'. The 'Metadata Update Date' is set to '(yyyy-mm-dd)' with 'Start' and 'End' date pickers. The 'Sort Results By' is set to 'Relevance'. Below the search filters, there are radio buttons for 'Anywhere', 'Intersecting', and 'Fully within', and a 'Where:' field with a 'GO' button. A map of the United States is shown with a red bounding box around the eastern half of the country. On the right side, the search results are displayed, showing 'Results 1-15 of 15 record(s)'. The first result is 'EPA Region 1 Toxics Release Inventory (TRI) Facilities', followed by 'U.S. EPA Image Service for Toxics Release Inventory (TRI) Sites (Public Version)', 'U.S. EPA Web Feature Service for Toxics Release Inventory (TRI) Sites (Public Version)', 'EPA Region 2 Regulated Facility Data', 'EPA Region 1 Regulated Facilities: RCRA Sites', 'U.S. EPA's Geospatial Data Access Project', 'EPA Region 1 Regulated Facilities: RCRA Generators', and 'EPA Region 1 Regulated Facilities: RCRA Transporters'.

↑ Simple search interface

records stored in databases could be viewed rather than having a separate location for these records. In addition, the rgw standard harvester in the Geoportal extension for ArcGIS Server was configured such that it required that EPA connect from a publicly available server to a number of intranet-only

gov and using the search and discovery resources there, and embedding GDG components in web applications. User input and recommendations will be key to the future success of the application. Provide feedback at geogateway.epa.gov/geoportal/catalog/identity/feedback.page.

For More Information

Jessica L. Zichichi

Vice President, Geospatial Solutions Group
Innovate!, Inc.

774-487-0445

jzichichi@innovateteam.com

Michelle A. Torreano

Geospatial Metadata Coordinator
US Environmental Protection Agency

202-566-2141

torreano.michelle@epa.gov

About the Authors

Jessica Zichichi has been working in the field of GIS for more than 13 years. She holds a master's degree in computer science and bachelor's degrees in geography and environmental studies. Her GIS experience ranges from basic mapping and cartographic production to desktop and web-based geospatial application development and programming. Zichichi's recent GIS efforts have been focused on enterprise geospatial solutions, geospatial metadata, and policy and planning.

Michelle Torreano is the geospatial metadata coordinator for the US EPA. She leads the implementation of the agency's geospatial metadata management framework, as well as coordinates with inter-agency data sharing groups. With the US EPA since 2001, she has worked on a variety of projects to advance the agency's National Geospatial Program, including system management, geospatial data acquisitions, and policy development. Torreano graduated from Purdue University with a degree in natural resources and environmental science.

GDG Resources

GDG REST URL https://geogateway.epa.gov/geoportal/webhelp/en/geoportal_931/GDG_REST_URL_Options.pdf

GDG search widget <https://geogateway.epa.gov/geoportal/catalog/components/components.page>

The screenshot displays a GeoRSS feed interface. At the top, a yellow box contains the text: "You are viewing a feed that contains frequently updated content. When you subscribe to a feed, it is added to the Common Feed List. Updated information from the feed is automatically downloaded to your computer and can be viewed in Internet Explorer and other programs. Learn more about feeds." Below this is a "Subscribe to this feed" button. The main content area shows two feed items. The first item is titled "EPA Region 1 Toxics Release Inventory (TRI) Facilities" and includes a timestamp "Wednesday, August 18, 2010, 3:28:56 PM" and a description of the TRI database. The second item is titled "EPA Region 1 Regulated Facilities: RCRA Treatment, Storage, and Disposal (TSD) Sites" and includes a timestamp "Wednesday, August 18, 2010, 3:28:30 PM" and a description of the RCRA sites. On the right side of the feed, there is a sidebar with a "Displaying" section showing "15 / 15" items, a "Sort by:" section with options for "Date" and "Title", and a "Filter by category:" section with options for "downloadableData" (10), "geographicActivities" (1), and "liveData" (4).

↑ GeoRSS output

servers, which would not have been acceptable for EPA security. The agency developed the WAFer tool to serve as a single access point from the public GDG server and relay all harvesting to internal EPA servers.

WAFer produces a web-accessible folder (WAF) for each back-end endpoint. These WAFs are then made accessible to the public GDG server using a single protocol and base URL. This allows EPA to connect to a single internal computer and extends the back-end connection types that can be harvested, such as ArcSDE databases and FTP sites. The agency could maintain harvesting from a public GDG server while also extending harvesting in a way that was critical for GDG contributors.

Soliciting Feedback

GDG is a public resource. EPA encourages public involvement with the GDG. Individuals can use GDG resources by visiting the GDG website at geogateway.epa.gov.

What's Next?

Federal agencies are being asked to take specific actions to implement the principles of transparency, participation, and collaboration enunciated in President Barack Obama's *Memorandum on Transparency and Open Government*. As part of this effort, EPA has focused on documenting nongeospatial data for sharing. Because of its success, GDG will be extended to become the central location for exchanging and managing metadata for all agency datasets. An expanded GDG will allow contributors to easily manage and publish metadata for various types of data at a single location and will serve as a central access point for internal users, external users, and interagency data sharing applications. An expanded GDG will advance EPA's ability to meet its mission goals by providing an improved framework for data sharing that increases reuse of resources, promotes transparency, and facilitates data exchange across the agency and with external parties.