data.gov 2.0 and Geospatial Platform Integration

Doug Nebert

Senior Advisor for Geospatial Technology
USGS–CSS, FGDC Secretariat
The Geospatial Platform and data.gov environments were redeployed in October 2011 but maintain separate catalogs and UIs. The objective to maintain one catalog for all government data assets and to develop Platform presence to provide extended features for geospatial resources. Re-engineering project nearly complete to deploy new single catalog and create new Platform website and Community capabilities.
Components

- CKAN will be the catalog with UI, CSW and native interfaces
- Geospatial Platform will be a primary UI to link to search, access, communities, visualization, market place
- Data.gov will continue to provide primary UI for all government data
- Communities module from data.gov has been updated to Drupal 7 and to be used in Platform
- ArcGIS Online will be used as backend for mashups, web app creation, geospatial viewer
Data.gov Catalog Redesign

- GSA selected the open source CKAN metadata catalog system to replace the geospatial catalog initially and the data.gov DMS and index by June 2013.
- geo.data.gov and the Geospatial Platform will both use the new data.gov catalog to harvest and store all data descriptions.
- GSA is developing a specification for representation of ‘raw’ (non-geo) data descriptions to support the Open Government Initiative and will provide tools to help agencies create and edit those records.
Redesign, continued

- There will eventually be one registration process for ‘raw data’ and geospatial data, services (APIs), and applications.
- Primary method will be harvest from registered agency metadata WAF, CSW, or portal interfaces.
- Secondary option is form-based for Core ISO.
- Search facets are created to help users browse all resources quickly by topic, agency, geography, resource type combinations.
- Planned links to visualization applications and APIs for data graphing, charting, and mapping.
Timelines

- Tasking to the Open Knowledge Foundation is in place for CKAN customization, which will be folded into open source and OGPL codebase and integrated with existing Drupal framework.
- Several development spirals have been completed to provide initial functionality by March 29.
- Current spiral to add geospatial search and map interface.
- Public deployment on GSA Cloud infrastructure will occur in May 2013 at catalog.data.gov.
CKAN prototype with records from geo.data.gov

The UI is flexible. AGOL “Groups” are being implemented as GSA Communities hosted either at data.gov or under Platform.
System will include a single registration and approval workflow for ‘raw’ data and geospatial, for catalogs, APIs, services, and apps from federal and non-federal publishers.
We will go beyond datasets to include data, services, applications, and metadata collections, and visualizations.
Search facets provide instantaneous combination search filtering on multiple filters as selected by users on the left. This example is of USGS source in Alaska. This is a sort of browse that facilitates data exploration. We plan to populate these search facets based on controlled vocabularies on the processed metadata, such as ISO Topic Category, A–16 theme, agency, resource type, and geographic extent.
When the resource type is determined to be viewable in a map viewer, a viewer will be launched. It should be possible to have multiple access methods described and exploited from the metadata.
OpenLayers viewer will be used to display KML, KMZ, GeoRSS, and WMS sources. AGOL–hosted javascript viewer will be used to display ArcGIS REST and webmap resources.

WMS service

From the dataset abstract

The National Land Cover Database (NLCD) 2001 was produced through a cooperative project conducted by the Multi-Resolution Land Characteristics (MRLC) Consortium. The MRLC Consortium is a...

Source: The National Map - National Land Cover Database (NLCD) WMS
Q&A

- FGDC Metadata are collected and saved for display, but managed as ISO 19115-2. ISO can also be harvested.
- Validation of metadata occurs before ingest, errors flagged for publishers; validation against one of six schemas.
- Developing a Best Practices or recommendations doc for metadata to optimize CKAN functionality.
- Current registration info will be carried over to new shared IDP/IDM system based on SAML 2.0.
- CKAN content will be based on fresh harvest and review.
- Support management of large homogeneous collections to enable two-step query.
- Are deploying “Communities” module from data.gov into Platform with links into CKAN records flagged for each community.
- Marketplace is being revived to support “Planned” data.
- Tutorials will be built for roll-out soon.