The 2015 CDI DataBlast was held at the Denver Federal Center on Tuesday, May 12 during the 2015 Community for Data Integration Workshop. Below are the abstracts and PDFs for the posters that were presented at the DataBlast.

- A Controlled Vocabulary for Marine Planning Data
- A Culture of Data Management within the USGS Chesapeake Bay Studies
- A Distributed, Standards-Based Framework and Open-Source Software Stack for Searching, Accessing, Analyzing and Visualizing Met-Ocean Data: Application to Hurricane Sandy
- A Rubric to Evaluate Core Data Management Plan Components
- Alaska Science Center Data Distribution Workflow
- Coastal and Marine Geology Program Video and Photograph Portal
- Compass: A new direction in data management
- Data Mine: Mobile legacy data management tool
- Data Publication and Access via Approved Database
- Expression of Controlled Vocabularies from Web Services Using Terminological Grids
- Government Open Data
- Have FAQs About New USGS Data Policies? So Do We!
- Integration of Land Cover Trends Field Photography with an Online Map Service
- Making Unmanned Aircraft System (UAS) Data Available to USGS Scientists and the Public
- National Dam Removal Database: A living database for information on dying dams
- National Water Census Data Portal and Archives
- Project Lifecycle Tracking
- ScienceCache: A framework for engaging citizen scientists in data collection through geocaching
- State of Colorado Open Data Initiatives
- Survey Manual Changes Affecting Everyone
- The Water Quality Portal: A single point of access for water quality data
- Web-enabled Visualization and Access of Value-added Disaster Products
- Woods Hole Coastal and Marine Science Center is En Vogue with a New FADD
Providing a data foundation for the National Ocean Policy process of regional marine planning is a tremendous data integration challenge. A network of independent people, organizations, and machines will need to coordinate their efforts to produce a transparent framework of high-quality data covering a broad range of topics, for the whole area of the U.S. coast, Exclusive Economic Zone, and Great Lakes. An initial list of necessary data was produced in 2011 by an interagency working group. In the last two years, scientists from USGS, EPA, and NOAA have worked together to convert the list to a set of controlled vocabulary terms with definitions that form a logical structure for organizing the data in a catalog. Currently, we are working on putting the vocabulary online in a vocabulary service and creating a taxonomy-based interface to the Data.gov collection. If the independent data providers use the controlled vocabulary in metadata records that are submitted to Data.gov, the marine planning data foundation can be automatically identified and organized to be easily found by regional planning groups.
A Culture of Data Management within the USGS Chesapeake Bay Studies

Cassandra Ladino (ccladino@usgs.gov) - USGS Eastern Geographic Science Center

The USGS Chesapeake Bay Studies program faces many of the same challenges that other resource constrained groups deal with in trying to develop a systematic implementation of Data Management principles into project workflows. Over the last three years much effort has been spent understanding project level interactions and the needs of our lead scientists. This year is especially pivotal for three reasons: 1) new Data Management fundamental science practices adopted by the Survey require adherence to essential Data Management principles; 2) the financial resources to staff a permanent Data Manager to assist in the creation of proper documentation and reinforce use of data management applications like USGS’s ScienceBase; and 3) the requirement to develop a database that allows integration of data from the multiple science themes (Land and Climate Change, Water Quality, and Fish, Wildlife, and Habitats) of highest importance to the USGS Chesapeake Bay studies to help answer complex ecosystem questions. This poster will describe the people, discoveries, and challenges of work in FY15 to provide a sense of a Data Management culture being built at the project level within the USGS Chesapeake Bay studies.
A Distributed, Standards-Based Framework and Open-Source Software Stack for Searching, Accessing, Analyzing and Visualizing Met-Ocean Data: Application to Hurricane Sandy

Rich Signell (rsignell@usgs.gov) - USGS Coastal and Marine Geology Program
Andrew Yan (ayan@usgs.gov) - USGS CIDA
Filipe Fernandes (ocefpaf@gmail.com) - SECOORA
Kyle Wilcox (kyle@axiomdatascience.com) - Axiom Data Science

SSF Category: Computational Tools & Services

There have been significant recent advances in common data models, web services, and python-based tools for search, access, analysis and visualization. It is now possible to supply an individual or organization with a complete, free, open-source framework that enables: (1) providers to easily serve their data in standardized form without impacting their existing workflows (using Unidata THREDDS Data Server, sci-wms and pycsw); (2) users to perform standardized search for data (using pyoos and OWSLib); (3) users to analyze and visualize data (using Iris, Cartopy, Pandas, Folium, mpileaflet, Jupyter notebooks and Wakari Enterprise). This framework relies on the CF-Standard for describing data in grids, profiles, time series, on the UGRID standard for describing data on unstructured (e.g. triangular) grids, and on the newly developed SGRID standard for describing data on staggered grids (commonly used in atmospheric and oceanographic models). These standards enable unified access via WMS image service endpoints and OPeNDAP/SOS data service endpoints. The entire infrastructure will be demonstrated using Hurricane Sandy datasets from multiple institutions.
A Rubric to Evaluate Core Data Management Plan Components

Brian Westra (bwestra@uoregon.edu) - University of Oregon

SSF Category: Management, Policy, & Standards

Research funding agencies are increasingly requiring that Data Management Plans (DMPs) accompany funding proposals. Our poster will describe the rubric we’ve created for analyzing DMPs and how this analysis may yield insight into the kinds of data researchers are generating, and how they intend to manage and share those data. This project has focused on NSF DMPs, but the framework could be applied to other agencies.

Awareness of researcher practices and intentions is fundamental to providing RDM services that are tailored to the needs of researchers. The information gathered through the rubric could also support improvements in guidance and resources provided by research institutions and agencies.

The rubric might also be employed by researchers and support staff to critique a DMP before it is submitted alongside a grant application, thus avoiding submitting a plan with missing or limited content.

Although tools - such as the DMPTool or DMP Online - help with the creation of a DMP, there is no standardized tool to aid with the evaluation of the quality of a DMP. Further, nothing has been developed to enable large-scale evaluation of DMPs for research purposes. We expect our rubric to fill this need.
Revisions to the Fundamental Science Practices (FSP) in response to recent directives issued by the Office of Science and Technology Policy (OSTP) and Office of Management and Budget (OMB) require Principal Investigators (PI) to make the data resulting from their research publicly available. Absence of an established and proven workflow for assessing data quality, preparation of metadata, and selection of approved repositories has posed a challenge to PIs and USGS Science Centers alike.

The Alaska Science Center (ASC) has prototyped a data publication workflow which has now been applied successfully to five separate data publication projects with more currently in the pipeline. Being a new workflow, each project has informed the overall process by encompassing new requirements and identifying opportunities for increased efficiency. The workflow is a ‘process flow diagram’ documented in ‘swim-lane’ format for easy comprehension of the responsibilities belonging to each of twelve identified roles. We have found this diagram an effective means of communicating the activities of data release to all parties and have plans to automate tracking of data products in the workflow in the near future.
Coastal and Marine Geology Program Video and Photograph Portal

Nadine Golden (ngolden@usgs.gov) – USGS Pacific Science Center
Seth Ackerman (sackerman@usgs.gov) – USGS Woods Hole Coastal and Marine Science Center

SSF Category: Data & Information Assets

This portal provides access to Coastal and Marine Geology Program video and photography of the seafloor off of coastal California and Massachusetts, and aerial imagery of the coastline along segments of the Gulf of Mexico and mid-Atlantic coasts. These data were collected as part of several USGS Coastal and Marine Geology Program Seafloor Mapping projects and Hurricane and Extreme Storm research.

Most CMGP video and photographic data has not been available to the public before. Though version one includes only a few of the largest CMGP video/photo data sets, iterations of the portal will create one online location for all CMGP video and still photographs.

California GIS base layers offshore of Santa Barbara Channel are included. We will continue to add GIS spatial data layers in order to display seafloor morphology and character, identify potential marine benthic habitats, and illustrate surficial seafloor geology and shallow subsurface geology.

Significant impacts:

1. provides information for ocean planning decision-support tools;
2. establishes baselines for monitoring long-term change;
3. geologic framework provides basis for local earthquake and tsunami hazard assessment;
4. assessment of sediment distribution/thickness provides important input to sediment management and sediment transport modeling; and
5. documenting habitat provides basis for ecosystem management.
Supporting our scientists by avoiding duplication of their effort is vital for successful implementation of the USGS requirement of Data Management Plans (DMP). The USGS Coastal and Marine Geology Program (CMGP), plans to expand the functionality of an internally developed system called Compass. Compass was recently created to ensure that all federally funded data collections be managed and preserved for the CMGP offices in Woods Hole, MA, St. Petersburg, FL, and Santa Cruz, CA. In addition to maintaining metadata about field activities and linking data collected by projects, we are working to increase Compass’ functionality to include creating, managing, and updating DMP’s, since much of the DMP information can be/are currently collected for active projects in Compass.

**Compass - A New Direction in Data Management**

Carol Reiss (creiss@usgs.gov) - USGS Pacific Coastal and Marine Science Center
Fran Lightsom (lightsom@usgs.gov) - USGS Coastal and Marine Geology Program

SSF Category: Data & Information Assets

A team of CMGP data managers have been meeting bi-weekly since February 2015 to develop this recommendation for implementing the new USGS requirement that all projects have Data Management Plans (DMPs). In general terms our recommendation is this:

- Use existing Compass resources.
- Establish required DMP elements.
- Implement, with guidance.
- Refine and improve.

**Implementing Data Management Plans in CMGP**

<table>
<thead>
<tr>
<th>Compass resources</th>
<th>DMP elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Name</td>
<td>Project Start Date (month/yr)</td>
</tr>
<tr>
<td>Principal Investigator</td>
<td>Project End Date (month/yr)</td>
</tr>
<tr>
<td>Institutional Sponsor</td>
<td>Name of project</td>
</tr>
<tr>
<td>Collaborating agencies</td>
<td>Notes</td>
</tr>
</tbody>
</table>

- Will field activities be collected (yes/no)?
- Will data be stored in a central database?
- How will data be stored from external sources?
- Will data be shared through other methods?
- Type of archiving笈笈
- Data preservation
- Data retention
- Data storage
- What will data be stored?
- Where will data be stored?
- When will data be stored?
- Will data be shared?
- What will backups of data be stored?
- Where will backups of data be stored?
- Who is responsible for data archiving?
- Who is responsible for sharing data?
- Who is responsible for data preservation?
- Who is responsible for backups of data?
- Will data be shared?
- How long do you expect the data to be available?

US Department of the Interior
US Geological Survey
In 2013 the USGS Center for Data Integration (CDI) funded the “USGS Data Mine” project to evaluate the CDI data lifecycle framework by applying it to selected legacy datasets. The primary objective was to gain a realistic understanding of the data release process and develop a simple tool to conduct legacy data inventories and estimate data release resource needs. In 2014 we conducted a legacy data inventory of the USGS Fort Collins Science Center and selected 6 legacy datasets to process, document and release for public distribution. Based on that experience we are developing a simple, truly mobile-first application that enables USGS data managers to:

- create a categorized inventory of their program’s legacy data;
- describe each dataset in their inventory in its current state and the work required to convert it to a distributable open-data format;
- document each dataset’s progress through the USGS review and approval;
- archive their approved legacy dataset and its metadata in a selected Sciencebase community; plan and report on legacy data recovery projects.

**Data Mine v.1**

*Legacy Data Inventory Tool*

- Mobile design for mobile users
- Designed by/for data stewards
- Built on USGS frameworks

---

**SSF Category: Data & Information Assets**

Lance Everette (everettel@usgs.gov) - USGS Fort Collins Science Center
A USGS approved database provides a mechanism for quality-controlled conforming datasets data to be published quickly online without separate publication and additional bureau-level oversight. Establishment of such a database requires the development and documentation of the data acquisition, review, and approval process. An example of an approved database is the USGS Oceanographic Time-Series Database, published on the Woods Hole Coastal and Marine Science Center’s “stellwagen” server (http://stellwagen.er.usgs.gov/). The rigorous review-and-approval process for these data is described in USGS Open File Report 2007-1194 (http://woodshole.er.usgs.gov/pubs/of2007-1194/).

The data on stellwagen were collected during scientific research projects carried out from 1975 to the present and are organized by region. Periods of data collection were typically one month to several years. The experiments commonly focused on observations near the seafloor, and most also obtained some current velocity or hydrographic data in the water column. “Stellwagen” provides on-line access to measurements via file download and THREDDS OPeNDAP services. Data discovery and access methods will be discussed. The metadata and storage strategy allows these data to be accessed by a variety of clients and services. For example, a portal is under development that will allow users to browse and view this data set, along with model data and other kinds of observations.
Expression of Controlled Vocabularies from Web Services Using Terminological Grids

Peter N. Schweitzer (pschweitzer@usgs.gov) - USGS
Alan O. Allwardt (aallwardt@usgs.gov) - USGS
David G. Govoni (dgovoni@usgs.gov) - USGS
Frances L. Lightsom (flightsom@usgs.gov) – USGS

SSF Category: Communities of Practice

Along with web services that provide controlled vocabularies, developers need to consider carefully how best to present terms to users. Alone, a user with a lot of time may be shown all of the terms. But most of the practical situations in which people need controlled terms are not like that. Instead the user needs a small number of terms at a time, with some way of seeing relationships among those terms, and scope notes explaining them. Ignoring these distinctions may cause the user to be overwhelmed with information or frustrated by not having enough.

An unusual use of controlled vocabulary is to arrange terms in a small grid such as 5x5. This offers enough terms to help the user understand the terms and to locate them easily. The situation in which this arrangement is helpful is the keynote or other overarching presentation given to a large group, in which the listener’s mind may wander. By focusing on the chance hearing of terms from the grid, the listeners inadvertently pay more attention than they otherwise might. Anecdotal evidence from a prior experiment of this type was encouraging; participants of this session will be offered the opportunity to test this idea.

Government Open Data

Mike Frame (mike_frame@usgs.gov) - USGS Core Science Analytics and Synthesis Program
Lisa Zolly (lisa_zolly@usgs.gov) - USGS Core Science Analytics and Synthesis Program
Viv Hutchison (vhutchison@usgs.gov) - USGS Core Science Analytics and Synthesis Program

SSF Category: Management, Policy, & Standards

The U.S. Geological Survey’s Core Science, Analytics, Synthesis, and Libraries (CSAS&L) program has been a leader in the agency’s response to the Open Data Initiative launched by the Obama Administration in 2013. USGS has generated new applications, policies and ongoing initiatives for education and outreach, resulting in improved access and usability of the bureau’s data. The DOE Oak Ridge National Laboratory (ORNL) has been a strong collaborator with CSAS&L in meeting these initiatives through the development of various tools, technologies, and best practices.

This poster will reference the resulting projects connected to USGS open data efforts including the Science Data Catalog, a searchable public data listing of officially released science data; the Science Data Lifecycle Model, a framework on which to convey best practices for the flow of scientific data in use by multiple organizations; the Digital Object Identifier creation tool, used to assign persistent identifiers to items; the Online Metadata Editor, a form-based tool that asks straightforward questions about a research dataset and generates a metadata record compliant with federal standards; as well as continued education and outreach methods. This poster addresses the challenges faced when managing, analyzing and sharing large quantities of earth science data; USGS CSAS&L goals relate to strengthening and progressing applied research computing, data best practices, and preparing the next-generation of data scientists for the global workforce.
Have FAQs About New USGS Data Policies? So Do We!

Heather Henkel (hhenkell@usgs.gov) - Southeast Science Center

SSF Category: Management, Policy, & Standards

The Data Management Policy Team and CDI, working in conjunction with the outreach efforts of the Fundamental Science Practices Advisory Committee (FSPAC), are developing FAQs to answer common questions related to the new data release policies. These FAQs are intended to complement and cross-reference the USGS Data Management web site and represent a Bureau-wide consensus to ensure interpretations of Fundamental Science Practices (FSP) policies are uniformly applied throughout USGS.

Feedback on these documents is needed to determine if they are adequate to answer frequently asked questions, to decide if the answers are clear and useful, and to find what other questions should be added. This poster will provide copies of the draft FAQ documents, along with the “Data Release Via Web Site or Web Service” process diagram, and allow people to read and comment on them.
Integration of Land Cover Trends Field Photography with an Online Map Service

Christopher Soulard (csoulard@usgs.gov) - USGS
Jason Sherba (jsherba@usgs.gov) - USGS
Ryan Longhenry (rlonghenry@usgs.gov) - USGS

SSF Category: Data & Information Assets

The Land Cover Trends field photography collection is a national-scale, ground-reference dataset which initially served as a research tool to aid in Landsat-derived land-use/land-cover change analyses and assessments. Between 1999 and 2009, Land Cover Trends scientists collected over 33,000 geographically referenced field photos with associated keywords describing the underlying LULC and change process taking place. This controlled, reliable field photography collection represents the most comprehensive national database of geo-referenced photography in the United States. CDI funding will support the effort to add geotags and keywords to digital copies of each photo, ingest, manage, and host all tagged photos in Earth Explorer with help from collaborators at the USGS Earth Resources Observation and Science (EROS) Center, and complete ongoing web portal environment to serve digital photography alongside other geospatial data. Within 5 months, researchers, land managers, and citizens will be able to efficiently search and download over 20,000 Land Cover Trend field photos within the USGS Earth Explorer web-based user interface.
Making Unmanned Aircraft System (UAS) Data Available to USGS Scientists and the Public

Raad Saleh (rsaleh@usgs.gov) - USGS EROS Data Center
Jennifer Lacey (jlacey@usgs.gov) - USGS EROS Data Center
Jeff Sloan (jsloan@usgs.gov) - USGS UAS
Bruce Quirk (quirk@usgs.gov) - USGS UAS
Josip Adams (jadams@usgs.gov) – USGS Rocky Mountain Geographic Science Center

SSF Category: Data & Information Assets

As Unmanned Aircraft System (UAS) technologies advance, so are the sensors they carry. This results in hundreds or even thousands of gigabytes of data in a single flight. Where do these data go, who can access them, and does it adhere to standards? These and similar questions are ones that need to be addressed with the increasing use UAS technologies. Currently, there are no mechanisms or protocols to make such data available to the public and government agencies on a streamlined, structured, and managed way. As the USGS, and other government agencies, conduct more UAS missions, we need to address how we conduct data management activities associated with data acquired by UAS. We are proposing to develop a plan that would leverage on existing data management infrastructure to perform a pilot study on how UAS data sets can be properly archived, adhered to standards, and then disseminated to the public and the larger science community via the internet.

USGS Unmanned Aircraft Systems Data Management
National Dam Removal Database: A living database for information on dying dams

Jeff Duda (jduda@usgs.gov) – USGS Western Fisheries Research Center
J. Ryan Bellmore (jbellmore@usgs.gov) – USGS Forest and Rangeland Ecosystem Science Center
Jon Warrick (jwarrick@usgs.gov) USGS Pacific Coastal and Marine Science Center
Sky Bristol (sbristol@usgs.gov) - USGS Core Science Systems
Vivian Hutchison (vhutchison@usgs.gov) - USGS Core Science Systems
Daniel Wieferich (dwieferich@usgs.gov) - USGS Core Science Systems
Katherine Vittum (kvittum@usgs.gov) - USGS Western Fisheries Research Center
Laura Craig (lcraig@americanrivers.org) - American Rivers
Erin McCombs (emccombs@americanrivers.org) - American Rivers

SSF Category: Data & Information Assets

The last 20 years has seen an exponential increase in the number of dam removals as the Nation is faced with aging infrastructure and increased river restoration goals. The goal of this project is to create a dynamic, National Dam Removal Database (NDRD) that will make significant contributions to forecasting ecological and geomorphic responses to dam removal and guiding dam removal efforts through an ability to leverage metaknowledge from the entire body of dam removal science. This project will use USGS ScienceBasea dynamic data and information management system to manage and display interactive information about the geography, demographics, and science of dam removal to be used by researchers studying dam removal, natural resource and restoration practitioners, decision makers, and the public.

Figure 1: USGS dam removal science database: a living database for information on dying dams

**Scope:** The removal of dams has increased in recent years and this trend is expected to continue (Fig. 1). A synthesis of scientific information on dam removal studies is needed to inform decision makers, practitioners, and the public about ecosystem outcomes and the potential for river restoration.

**Objectives:** Create a dynamic, National Dam Removal Database (NDRD) that will make significant contributions to forecasting ecological and geomorphic responses to dam removal and guiding dam removal efforts through an ability to leverage metaknowledge from the entire body of dam removal science.

**Benefits:** A living repository of structured information about dam removal projects, outcomes, and scientific studies will provide managers, decision makers and scientists with a dynamic tool for exploring questions about dam removal.

**Project Architecture:** The USGS dam removal science database will be hosted within ScienceBase and leveraged with existing relevant information sources.

- The database is formed from data extracted from published scientific literatures and other contributed studies. ScienceBase vocabulary provides functionality for controlling input.
- Users can upload dam removal information through a spreadsheet upload and/or data entry form.
- Other existing geopolitical databases will be dynamically integrated into the NDRD, providing additional context relative to dam removal. Existing sources include USGS national data sets like HHM, NHD, as well as other relevant databases (e.g., NRD, American Rivers).
- A user interface serves as a portal to the dam removal science database, where project can be explored, data can be visualized, and spatial and temporal trends examined. ScienceBase API will drive interface functionality.

**Key Features:**
- Interactive, updated information on US dam removals.
- Up-to-date clearinghouse for dam removal project information and associated studies.
- User interface allowing new projects and studies to be added to database.

**Missing Features? Other Ideas? Let us know!** Please post your ideas for additional features or information assets that you think could help us.
The National Water Census Data Portal is a new web outlet for information from the USGS water use and availability program. It contains national estimates of water budget components for local watersheds, water use data for counties, tools to calculate statistics of daily streamflow records, modeled daily streamflow at ungaged stations, and access to records of aquatic biology observations. The portal is provides a subset of the research and information being conducted by the National Water Census. A new community in the ScienceBase system has been put in place to document the full scope of research conducted. This poster and accompanying interactive demo will summarize the National Water Census data archiving and dissemination efforts as well as the data management life cycle being implemented with researchers.
Project Lifecycle Tracking

Gail Montgomery (montgomeryg@usgs.gov) - USGS Fort Collins Science Center
Tim Kern (kernt@usgs.gov) - USGS Fort Collins Science Center
Lei Ann Wilson (wilsonl@usgs.gov) - CSG Contractor to USGS Fort Collins Science Center
Megan Eberhardt Frank (eberhardt-frankm@usgs.gov) - CSG Contractor to USGS Fort Collins Science Center
Haylee Schweizer (schweizerh@usgs.gov) - CSG Contractor to USGS Fort Collins Science Center
Dell Long (jlong@usgs.gov) - USGS Fort Collins Science Center
Emily Fort (efort@usgs.gov) - USGS National Climate Change Wildlife Science Center
Jacob Juszak (jjuszak@usgs.gov) - CSG Contractor to Fort Collins Science Center
Sebastien Nicoud (snicoud@usgs.gov) - CSG Contractor to USGS Fort Collins Science Center

SSF Category: Management, Policy, & Standards

NCCWSC, CSAS&L and FORT have collaborated to build a suite of tools designed to help managers and PIs work with project records, metadata and products stored in ScienceBase from initial record creation to display on a public website. These tools include RFPManager, DEPTH, PDash, and DMPEditor. They provide agencies value by facilitating the effort required to document projects and capture data and information related to those projects. They also help ensure records are accurate and accessible to others and support efforts such as the USGS Data Lifecycle, the LCC IDMN Catalog, and the application of data management standards from cradle to grave.
ScienceCache: A framework for engaging citizen scientists in data collection through geocaching

Tabitha Graves (tgraves@usgs.gov) - USGS Northern Rocky Mountain Science Center
Dell Long (jllong@usgs.gov) - USGS Fort Collins Science Center
Tim Kern (kernt@usgs.gov) – USGS Fort Collins Science Center
Jake Weltzin (jweltzin@usgs.gov) - National Phenology Network Program

SSF Category: Communities of Practice

We propose to develop a scientific geocaching mobile application framework that will target 2 user groups for citizen science data collection: youth and geo-cachers. By melding training and games into the hunt for place-based data collection sites, and incorporating photo uploads as data or for authentication, new volunteers can collaborate in robust data collection. Scientists can build a project on a website, specifying locations or goals for new data collection sites, clues for established sites, and questions to answer, measurements, or other activities for the site based on their individual data needs. We will develop the framework by building on the success of the USA National Phenology Network (NPN) and Science Base, using a case study assessing phenology of bear foods in Glacier National Park and applying those lessons to a second project evaluating tree invasion into alpine meadows using repeat photography. Data will flow to ScienceBase and NPN’s Nature’s Notebook databases when appropriate. In addition to the mobile app and website, this will contribute to the USGS knowledge base in interactive mobile application design patterns. This framework will enable increased spatial and temporal data collection and increase engagement of the public in science across divisions.
State of Colorado Open Data Initiatives

Andrew Cole (andrew.cole@sos.state.co.us) - Colorado Secretary of State
Matt Tricomi (mtricomi@usgs.gov) – USGS Core Science Systems

SSF Category: Data & Information Assets

The Colorado Secretary of State is working with the Governor's Office of IT (CIO group) and all Colorado Agencies via the GoCode Colorado initiative to help increase use by citizens, industry, and government and increase quality data published by agencies.

The first and only statewide effort of its kind, gocode.colorado.gov brings together entrepreneurs, business partners, and developers to make use of public data through a series of events. Last year at least three businesses were created around apps that use this data. Awards have flipped traditional government procurement on its head, creating value to Colorado businesses and society. Near 30 businesses including Companies like Google, Esri, SendGrid, Rally Software and Gnip (now Twitter) volunteered their time and event donations near $200,000 the first year because they desire more access to government data—they know how important this is to economic development. GoCode is the most concerted effort in the state to increase public data in data.colorado.gov adding over 150 data sets. This has increased agency participation, and the GoCode team supported the “data wrangling” and ETL to help the “have not” agencies get their data up and used.

Survey Manual Changes Affecting Everyone

John Faundeen (faundeen@usgs.gov) – USGS EROS

SSF Category: Management, Policy, & Standards

The release of the Instructional Memorandums was an outgrowth of work begun in 2011 to develop a data lifecycle model representing how USGS science does or should flow from creation to publishing. The work from the lifecycle model was directly usable in response to the 2013 OSTP and OMB directives related to increasing access to federally funded science research and managing information as an asset.

This poster graphically relates the lifecycle model, the 2013 directives, and the Instructional Memorandums. USGS staff will be available to discuss the elements and drivers as well as how the bureau will need to evolve its publishing and management of our rich science heritage.
The Water Quality Portal: A single point of access for water quality data

James Kreft (jkreft@usgs.gov) - USGS Center for Integrated Data Analytics

SSF Category: Data & Information Assets

The Water Quality Portal (WQP) is a cooperative project between the U.S. Geological Survey (USGS) (USGS) and the U.S. Environmental Protection Agency (EPA) overseen by the National Water Quality Monitoring Council (NWQMC). It was launched in April of 2012 as a single point of access for discrete water quality samples stored in the USGS NWIS and EPA STORET systems. Since launch thousands of users have visited the Water Quality Portal to download billions of results that are pertinent to their interests. Numerous tools have also been developed that use WQP web services as a source of data for further analysis. Since the launch of the Portal, the WQP development team at the USGS Center for Integrated Data Analytics has worked with USGS and EPA stakeholders as well as the wider user community to add significant new features to the WQP. WQP users can now directly plot sites of interest on a web map based on any of the 14 WQP query parameters, and then download data of interest directly from that map. In addition, the WQP has expanded beyond just serving out USGS and EPA data and is now providing data from the US Department of Agriculture’s Agricultural Research Service STEWARDS system, and is working with others to bring in additional data. Finally, the WQP is now linked to another NWQMC-supported project, the National Environmental Methods Index (NEMI), so WQP users can easily find the method behind the data that they are using. Future work is focused on incorporating additional biological data from the USGS BioData system and from STORET, as well as adding biologically relevant query parameters to broaden the scope of discrete water quality sample types that the WQP can provide. The WQP team is also exploring ways to further integrate with other systems, such as those operated the U.S. Department of Agriculture Forest Service, to facilitate the overarching goal of improving access to water quality data for all users.
Web-enabled Visualization and Access of Value-added Disaster Products

This project will provide enhanced capability for search, access, and visualization of the disaster-related products and images that are developed and contributed by USGS and other collaborators during the course of an emergency response event. These products are already hosted and delivered via the existing USGS Hazards Data Distribution System (HDDS), but they are not easily accessed. The project will support expanded ingest capabilities for HDDS, in order to allow disaster products to be more easily shared, discovered, visualized, and accessed by the end user community.

The HDDS is an interactive map-based web portal that provides a consolidated point-of-entry and distribution system for remotely sensed imagery and other geospatial datasets related to emergency response. When disasters occur, the system provides a critical source of satellite and aerial imagery for the emergency response community and many other end users. The HDDS allows rapid selection, preview, and download of relevant pre- and post-event imagery, and has supported several hundred emergency events since its inception in 2010. The hosted imagery is routinely accessed by end users from all levels of government (Federal, State, local, tribal, and international) along with many other organizations and communities who may be engaged in an emergency event support.
Traditionally, the Woods Hole Coastal and Marine Science Center (WHCMSC) has released data through U.S. Geological Survey (USGS) publication series such as an Open-File Reports or Data Series. In order to publish non-interpretive digital data more quickly via a web interface, WHCMSC has developed a Field Activity Data Display, or FADD; effectively a data catalog page linked to a WHCMSC Field Activity. The publication requirements for a FADD are similar to a USGS series publication: data must be accompanied by complete metadata (CSDGM FGDC), catalogued in IPDS, internally reviewed by at least two people, and receive a DOI number. All the data associated with the field activity are processed and released with the exception of some raw data that are locally archived and available upon request. The data requirements of the FADD ensure that ancillary data such as navigation and sound velocity profiles will be released with metadata. Because the FADD contains non-interpretive data, the review process only requires Center approval. The FADD is citable and fulfills the need for faster turn-around times from data collection to data release. Since only the data and metadata are released, this has the potential for a more thorough review of these elements.