

# **Community for Data Integration Fiscal Year 2017 Funded Project Report**

Open-File Report 2018–1154



# **Community for Data Integration Fiscal Year 2017 Funded Project Report**

By Leslie Hsu, Kate E. Allstadt, Tara M. Bell, Erin E. Boydston, Richard A. Erickson, A. Lance Everette, Erika Lentz, Jeff Peters, Brian E. Reichert, Sarah Nagorsen, Jason T. Sherba, Richard P. Signell, Mark T. Wiltermuth, and John A. Young

Open-File Report 2018–1154

**U.S. Department of the Interior  
U.S. Geological Survey**

## **U.S. Department of the Interior**

RYAN K. ZINKE, Secretary

## **U.S. Geological Survey**

James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2018

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <https://www.usgs.gov> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <https://store.usgs.gov>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

### Suggested citation:

Hsu, L., Allstadt, K.E., Bell, T.M., Boydston, E.E., Erickson, R.A., Everette, A.L., Lentz, E., Peters, J., Reichert, B.E., Nagorsen, S., Sherba, J.T., Signell, R.P., Wiltermuth, M.T., and Young, J.A., 2018, Community for Data Integration fiscal year 2017 funded project report: U.S. Geological Survey Open-File Report 2018–1154, 15 p., <https://doi.org/10.3133/ofr20181154>.

ISSN 2331-1258 (online)

## Contents

Abstract.....	1
Introduction.....	1
Community for Data Integration Funded Projects, Fiscal Year 2017.....	1
An Interactive Web-Based Application for Earthquake-Triggered Ground Failure Inventories.....	2
Automating the Use of Citizen Scientists' Biodiversity Surveys in iNaturalist to Facilitate Early Detection of Species Responses to Climate Change .....	3
Flocks of a Feather Dock Together—Using Docker and HTCondor to Link High-Throughput Computing Across the U.S. Geological Survey.....	4
U.S. Geological Survey Data at Risk—Expanding Legacy Data Inventory and Preservation Strategies .....	6
Empowering Decision Makers—A Dynamic Web Interface for Running Bayesian Networks.....	7
Web Mapping Application for a Historical Geologic Field Photo Collection .....	8
Visualizing Community Exposure and Evacuation Potential to Tsunami Hazards Using an Interactive Tableau Dashboard .....	9
Exploring the U.S. Geological Survey Science Data Life Cycle in the Cloud .....	10
Developing Application Programming Interfaces to Support Enterprise Level Monitoring Using Existing Tools.....	11
Extending ScienceCache Mobile Application for Data Collection to Accommodate Broader Use Within the U.S. Geological Survey.....	13
Evaluation and Testing of Standardized Forest Vegetation Metrics Derived from Light Detection and Ranging Data .....	14
Summary.....	15
References Cited.....	15

## Figures

1. Photograph showing shallow landslides (the unvegetated areas) triggered by the 2016 Kaikoura Earthquake in New Zealand.....	2
2. Diagram showing how TaxaTaxi compares online biodiversity databases to find species in common and unique to each database .....	3
3. Diagram showing the schematic of flocking with HTCondor between U.S. Geological Survey (USGS) centers.....	5
4. Image showing a bathythermogram, one of the types of legacy data inventoried and rescued in the Legacy Data Inventory Reporting System.....	6
5. Diagram showing the Bayesian network to predict coastal change as currently configured in proprietary software.....	7
6. Screen capture showing the user interface for the Grand Canyon photo collection web map application.....	8
7. Screen capture showing the Oahu Tsunami Evacuation dashboard.....	9
8. Image showing snapshots from a Jupyter notebook using a scalable computer cluster on the U.S. Geological Survey Yeti High Performance Computing facility to compute maximum wave height on an 80 gigabyte dataset .....	10
9. Diagram showing that Application Programming Interfaces (APIs) allow the connection of information documented in <a href="https://monitoringresources.org">MonitoringResources.org</a> to multiple applications .....	11
10. Screen capture showing how a grid cell selection tool web application for the North American Bat Monitoring Program (NABat) uses newly developed application programming interfaces (APIs) to share documentation and design of data collection efforts helping organize, curate, and facilitate access to NABat data .....	12
11. Screen capture showing how the improved ScienceCache Editor allows U.S. Geological Survey scientists to easily create data collection forms for internal studies in a private mode with user management capabilities.....	13
12. Image showing the data layers involved in extracting height above ground vegetation data from 3DEP light detection and ranging data .....	14

# Community for Data Integration Fiscal Year 2017 Funded Project Report

By Leslie Hsu, Kate E. Allstadt, Tara M. Bell, Erin E. Boydston, Richard A. Erickson, A. Lance Everette, Erika Lentz, Jeff Peters, Brian E. Reichert, Sarah Nagorsen, Jason T. Sherba, Richard P. Signell, Mark T. Wiltermuth, and John A. Young

## Abstract

The U.S. Geological Survey Community for Data Integration annually funds small projects focusing on data integration for interdisciplinary research, innovative data management, and demonstration of new technologies. This report provides a summary of the 11 projects funded in fiscal year 2017, outlining their goals, activities, and outputs.

## Introduction

The U.S. Geological Survey (USGS) Community for Data Integration (CDI) annually funds small projects focusing on data integration for interdisciplinary research, innovative data management, and demonstration of new technologies. Since 2010, the CDI has funded more than 70 projects. The CDI supports projects that

- focus on targeted efforts that yield near-term benefits to Earth and biological science;
- leverage existing capabilities and data;
- implement and demonstrate innovative solutions (for example, methodologies, tools, or integration concepts) that could be used or replicated by others at scales from project to enterprise;
- preserve, expose, and improve access to Earth and biological science data, models, and other outputs; and
- develop, organize, and share knowledge and best practices in data integration.

This report provides a summary of the 11 projects funded in fiscal year 2017, outlining their goals, activities, and outputs.

## Community for Data Integration Funded Projects, Fiscal Year 2017

The 11 projects funded in fiscal year 2017 illustrate the breadth of USGS research. Topics covered include landslides, biodiversity surveys, high-throughput computing, legacy data, coastal change, geologic field photos, tsunami hazards, computing in the cloud, species monitoring tools, mobile apps for field data collection, and vegetation metrics.

## An Interactive Web-Based Application for Earthquake-Triggered Ground Failure Inventories

*Principal investigator.*—Kate E. Allstadt (USGS)

*Coinvestigators and partners.*—Anna Nowicki Jessee (Indiana University), Robert G. Schmitt (USGS), Hakan Tanyas (Faculty of Geo-Information Science and Earth Observation), Eric M. Thompson (USGS), and Jing Zhu (Tufts University)

Inventories of landslides and liquefaction triggered by major earthquakes (fig. 1) are key research tools that can be used to develop and test hazard models. To eliminate redundant effort, we created a centralized and interactive repository of ground failure inventories that currently hosts 32 inventories generated by USGS and non-USGS authors and designed a pipeline for adding more as they become available. The repository consists of (1) a ScienceBase community page where the data are available for download and (2) an accompanying web application that allows users to browse and visualize the available datasets. We anticipate that easier access to these key datasets will accelerate progress in earthquake-triggered ground failure research, ultimately benefiting the mission of the USGS to apply hazard science to protect the Nation's safety, security, and economic well-being. Our solution also provides a template for other USGS groups seeking to create customized interactive tools using existing resources.

*Related links.*—

- See more at <https://www.sciencebase.gov/catalog/item/58b5c886e4b01ccd54fde368>
- ScienceBase community page: <https://doi.org/10.5066/F7H70DB4>
- Earthquake-Triggered Ground-Failure web application: <https://doi.org/10.5066/F7D799CT>
- USGS Data Series 1064: <https://doi.org/10.3133/ds1064>
- Code to generate bulk metadata: <https://code.usgs.gov/usgs/landslides-metadata>



**Figure 1.** Shallow landslides (the unvegetated areas) triggered by the 2016 Kaikoura Earthquake in New Zealand. Photograph by Kate Allstadt, U.S. Geological Survey (from Allstadt and others, 2018).



## Automating the Use of Citizen Scientists' Biodiversity Surveys in iNaturalist to Facilitate Early Detection of Species Responses to Climate Change

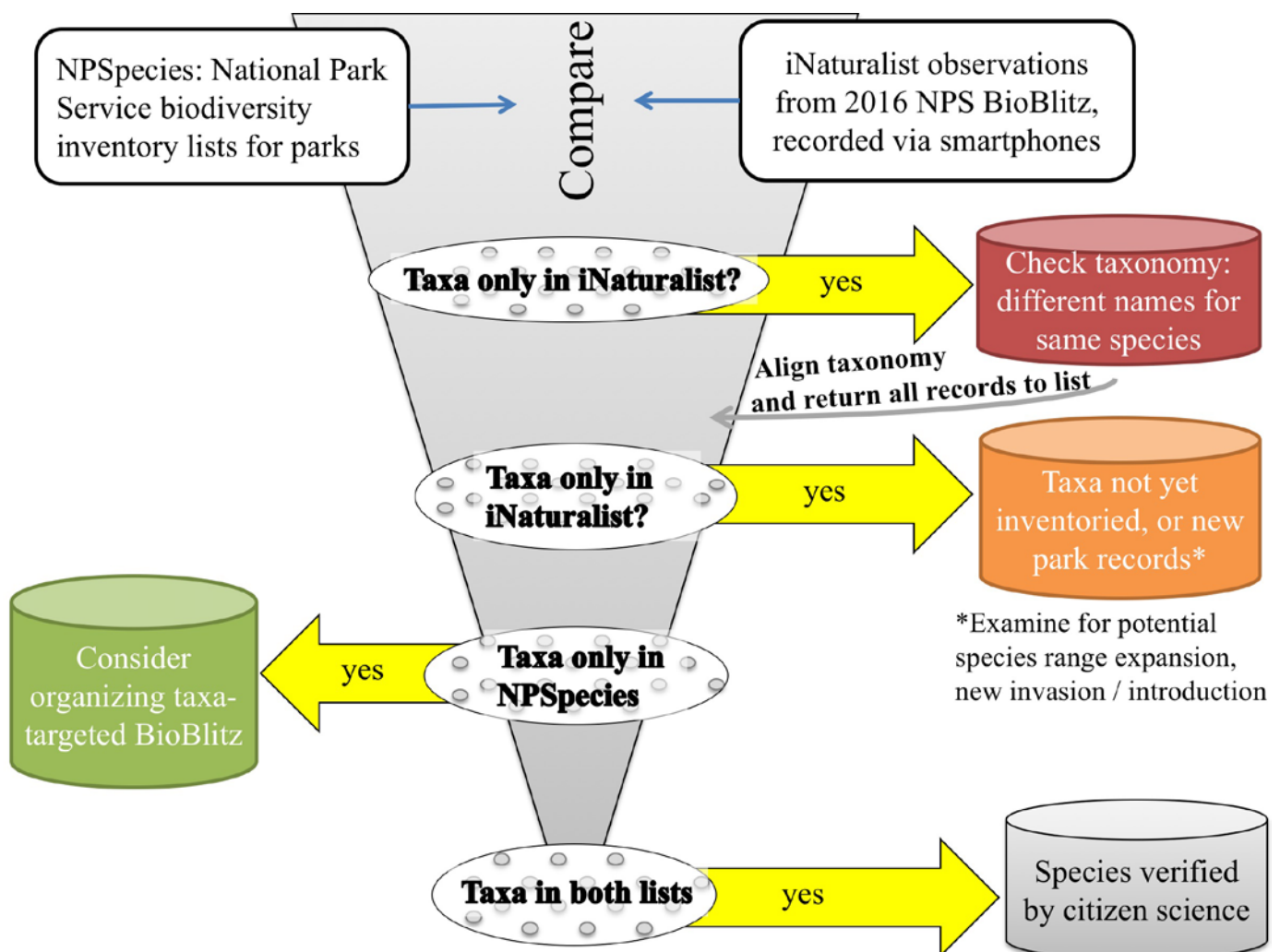
*Principal investigator.*—Erin Boydston (USGS)

*Coinvestigators and partners.*—Vijay Barve (Florida Museum of Natural History), Jenny Briggs (USGS), Lena Lee (National Park Service), and Toni Lyn Morelli (USGS)

A BioBlitz is a field survey method for finding and documenting as many species as possible in a specific area over a short period. The National Park Service and National Geographic Society hosted the largest BioBlitz survey ever in 2016; people in more than 120 national parks used the iNaturalist app on mobile devices to document organisms they observed. Resulting records have Global Positioning System (GPS) coordinates, include biological accuracy assessments, and provide an unprecedented snapshot of biodiversity nationwide. Additional processing and analysis would make these data available to inform conservation and management decisions. This project developed a process to integrate iNaturalist data with existing species lists from national parks (fig. 2). This cross-referencing of citizen science data with species inventories helps to update taxonomic and species occurrence information and may detect species range shifts or new invasions. Further, this process can serve as the basis for incorporating other online databases and to increase engagement of the public in biodiversity stewardship.

*Related link.*—

- See more at <https://www.sciencebase.gov/catalog/item/58b5cd07e4b01ccd54fde37f>



**Figure 2.** How TaxaTaxi compares online biodiversity databases to find species in common and unique to each database. From Boydston and others (2017). [NPS, National Park Service]

## Flocks of a Feather Dock Together—Using Docker and HTCondor to Link High-Throughput Computing Across the U.S. Geological Survey

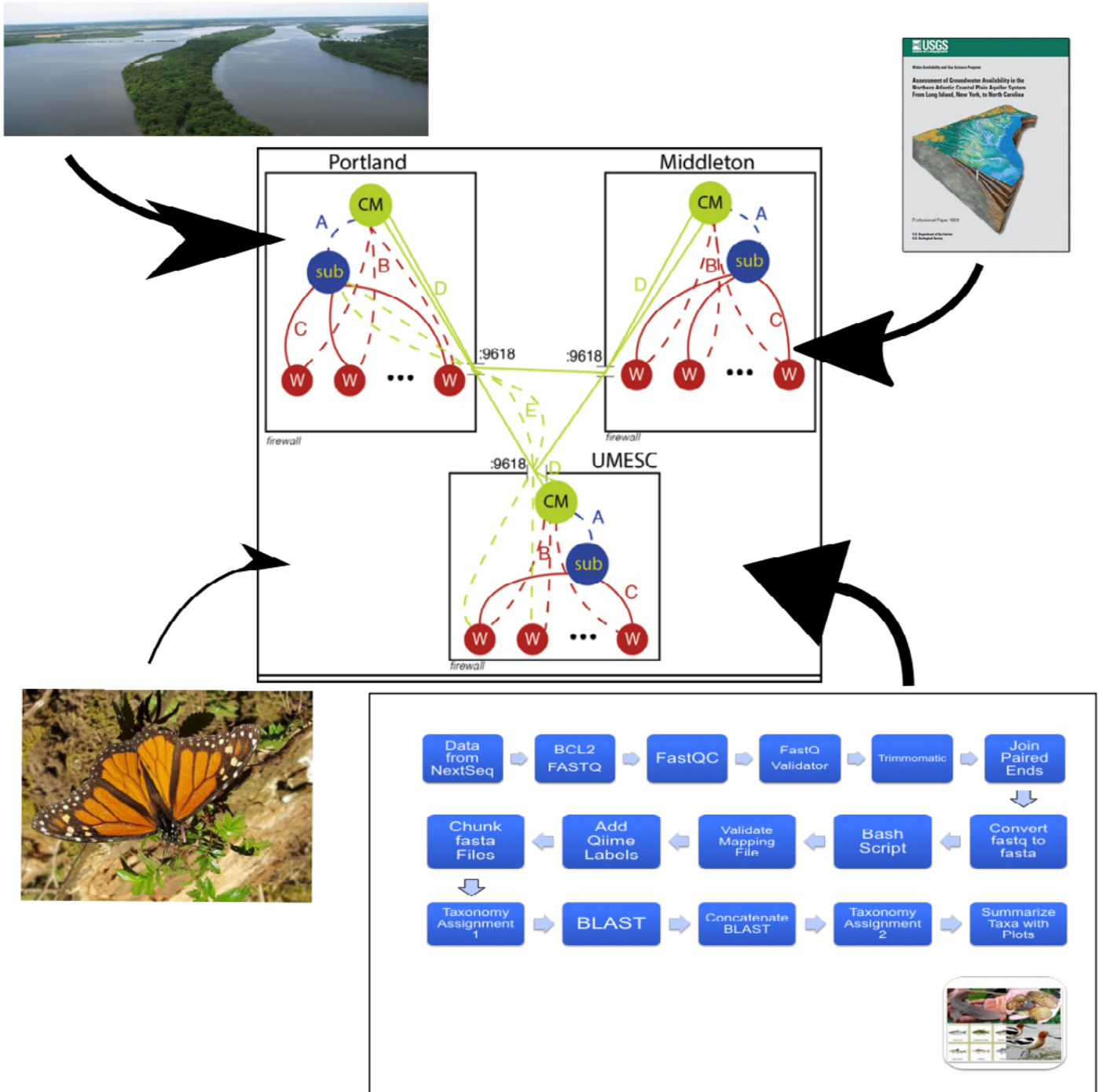
*Principal investigator.*—Richard Erickson (USGS)

*Co-investigators and partners.*—Sunnie (Grace) McCalla (USGS), Michael N. Fienen (USGS), Jon J. Amberg (USGS), Janice Gordon (USGS), Paul E. Exter (USGS), Melvin Bower (USGS), and Jonathan M. Knudson (USGS)

USGS scientists often face computationally intensive tasks that require high-throughput computing capabilities. Several USGS facilities use HTCondor to run their computational pools but are not necessarily connected to the larger USGS pool. This project demonstrated how to connect HTCondor pools by flocking, or coordinating, within the USGS (fig. 3). In addition to flocking the Upper Midwest Environmental Science Center and the Wisconsin Water Science Center, we have flocked with the USGS Advanced Research Computing Yeti supercomputing cluster and other water science centers. We also developed tutorials on how to sandbox code using Docker within the USGS environment for use with high-throughput computing. A main accomplishment of the project was working through technical hurdles in the existing computing environment. Our results from this project helped the USGS Advanced Computing Cooperative flock HTCondor pools together and move towards an operational implementation.

*Related links.*—

- See more at <https://www.sciencebase.gov/catalog/item/58b5db5ce4b01ccd54fde3e4>
- HTCondor tutorials: [https://my.usgs.gov/bitbucket/projects/CDI/repos/hunting\\_invasive\\_species\\_with\\_htcondor/](https://my.usgs.gov/bitbucket/projects/CDI/repos/hunting_invasive_species_with_htcondor/)



**Figure 3.** Schematic of flocking with HTCondor between U.S. Geological Survey (USGS) centers. During the past year, this project demonstrated how USGS scientists can use HTCondor across centers by flocking. Current uses of HTCondor span mission areas ranging from hydrological modeling to biological monitoring and processing of environmental deoxyribonucleic acid (eDNA) bioinformatics data. From Erickson and others (2017). [CM, central manager; A, B, C, and D, various connections; sub, submit node; 9618, a single port that is the only one needs to be opened through the firewall to enable flocking capabilities; W, worker node; UMESC, Upper Midwest Environmental Science Center]

## U.S. Geological Survey Data at Risk—Expanding Legacy Data Inventory and Preservation Strategies

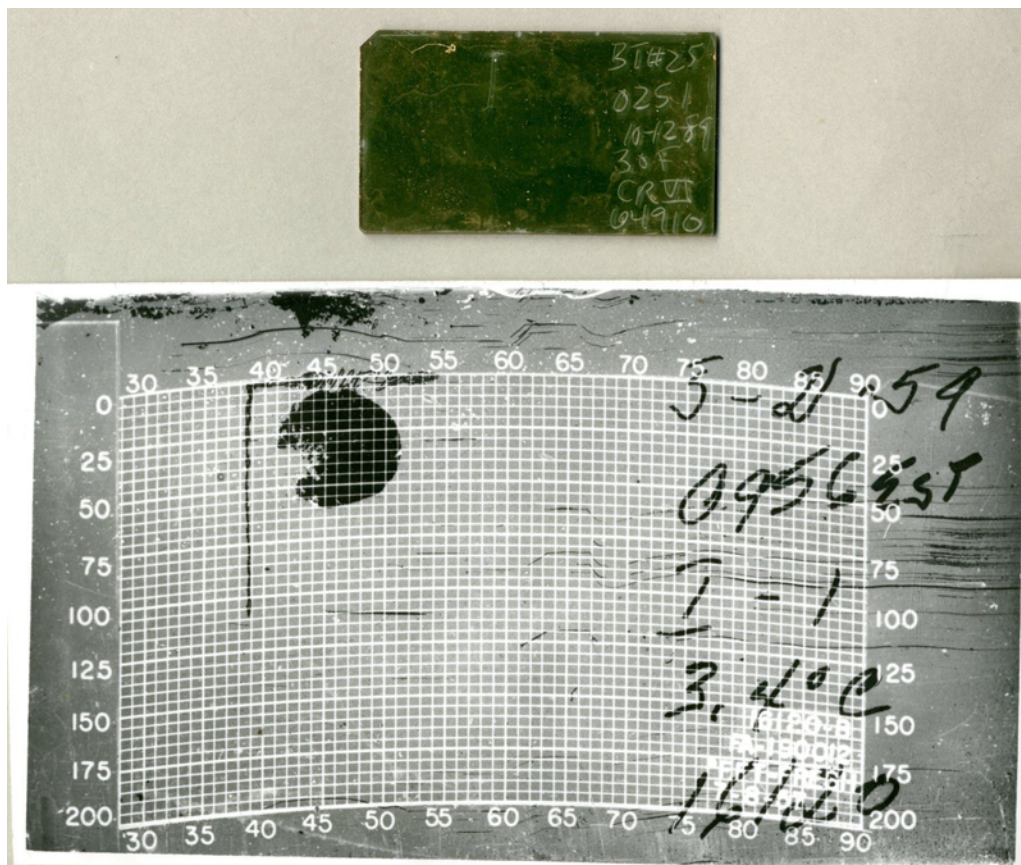
*Principal investigator.*—Lance Everett (USGS)

*Co-investigator.*—Tara Bell (USGS)

The Data at Risk (DaR) project built on previous work by refining the prioritization algorithm for the Legacy Data Inventory Reporting System (LDIRS) to better assess temporal, geographic, and taxonomic extents of datasets at risk. We continued to expand the LDIRS inventory (fig. 4) by working with two USGS career scientists and two USGS science centers (the Great Lakes Science Center and the Upper Midwest Environmental Science Center). During the past 2 funding periods, the DaR project has selected 13 high priority preservation projects to validate best practices for preserving and publishing USGS legacy data. Each project is summarized in a case study that describes methods and lessons learned. One of our highest priorities this year was promoting awareness of data in the inventory and communicating methods with other groups. We fostered relationships with other like-minded USGS groups, such as the USGS Step-Up program, which used the LDIRS prioritization reports to select the North American Bat Banding Program data for their fiscal year 2018 preservation work, a project first started with CDI funding in 2014.

*Related links.*—

- See more at <https://www.sciencebase.gov/catalog/item/58b5ddc3e4b01ccd54fde3fa>
- “Case Study: Great Lakes Science Center (GLSC) Legacy Data Inventory”: <https://www.sciencebase.gov/catalog/item/5abe9272e4b081f61ac14dd2>
- “Case Study: Upper Midwest Science Center (UMESC) Legacy Data Inventory”: <https://www.sciencebase.gov/catalog/item/5ad63d41e4b0e2c2dd23f456>
- USGS data release “Migrating Bird Survey Data Along the San Pedro River and its Tributaries, Southeastern Arizona, 1989-1994”: <https://doi.org/10.5066/F7K35SMD>



**Figure 4.** Bathythermogram of one of the types of legacy data inventoried and rescued in the Legacy Data Inventory Reporting System.



## Empowering Decision Makers—A Dynamic Web Interface for Running Bayesian Networks

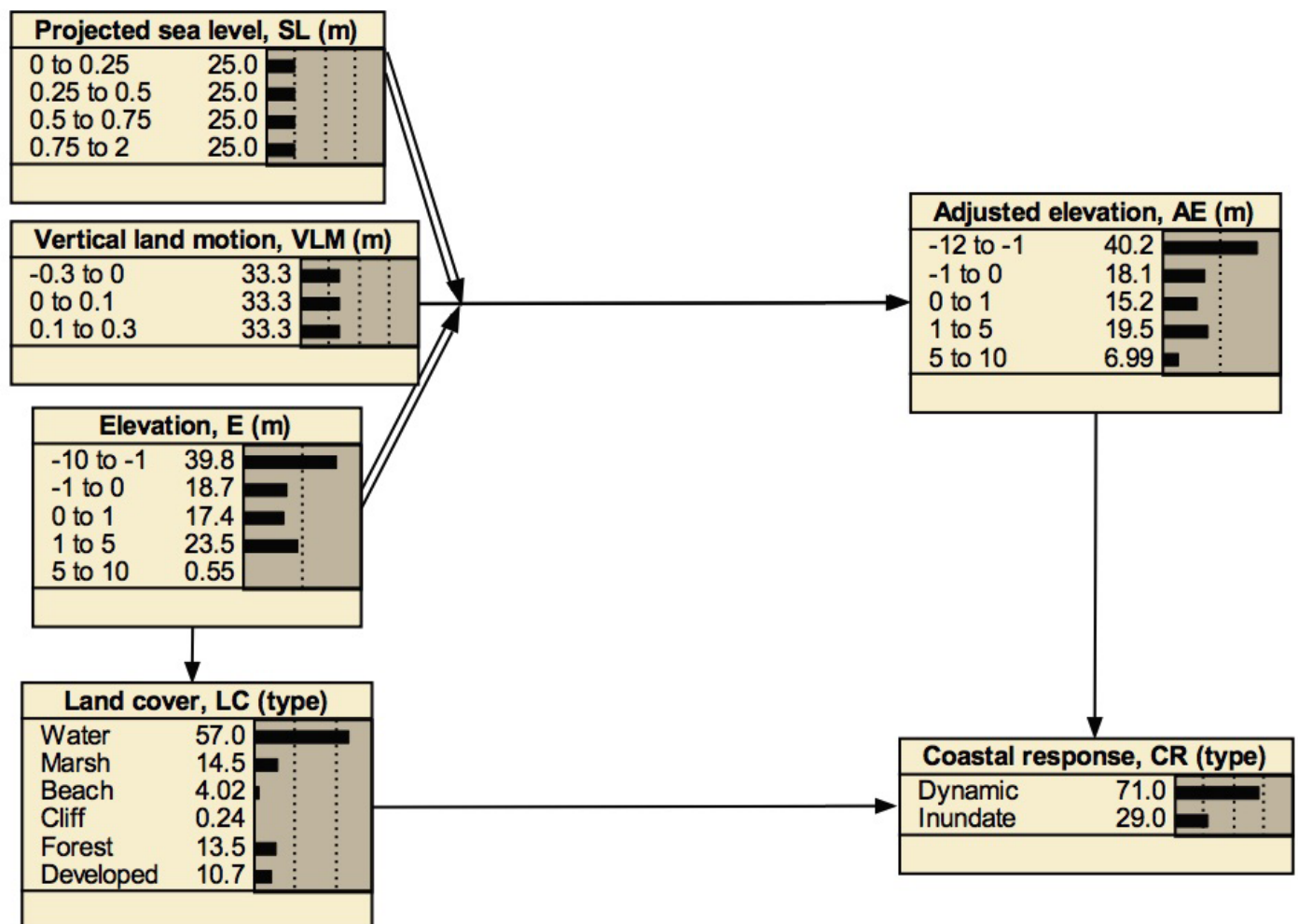
*Principal investigator.*—Erika Lentz (USGS)

*Co-investigators and partners.*—Ben Gutierrez (USGS), Michelle D. Staudinger (USGS), Nathaniel G. Plant (USGS), Sara L. Zeigler (USGS), Emily J. Sturdivant (USGS), Emily Himmelstoss (USGS), and Michael N. Fienen (USGS)

USGS scientists are at the forefront of research that is critical for decision making, particularly through the development of models (Bayesian networks[BNs]; fig. 5) that forecast coastal change. The utility of these tools outside the scientific community has been constrained because they rely on expensive technical software and a moderate understanding of statistical analyses. We proposed to convert one of our models from proprietary to freely available open-source software, resulting in a portable interactive web interface. The resulting product will serve as a prototype to demonstrate how interdisciplinary USGS science and models can be transformed into an approachable format for decision makers. We present lessons learned to date through the ongoing project planning and process of filling staffing requirements. Our report, available at our ScienceBase page, summarizes lessons learned and next steps in three categories: (1) processing, storage, and hosting; (2) staffing requirements and needs; and (3) the development process.

*Related link.*—

- See more at <https://www.sciencebase.gov/catalog/item/58b5e351e4b01ccd54fde418>



**Figure 5.** Bayesian network to predict coastal change as currently configured in proprietary software. Ongoing work in fiscal year 2018 and 2019 will develop an interface and back end code wherein a user can query different ranges in parameters and explore how coastal response scenarios change. Modified from Lentz and others (2015, 2016). [m, meter]

## Web Mapping Application for a Historical Geologic Field Photo Collection

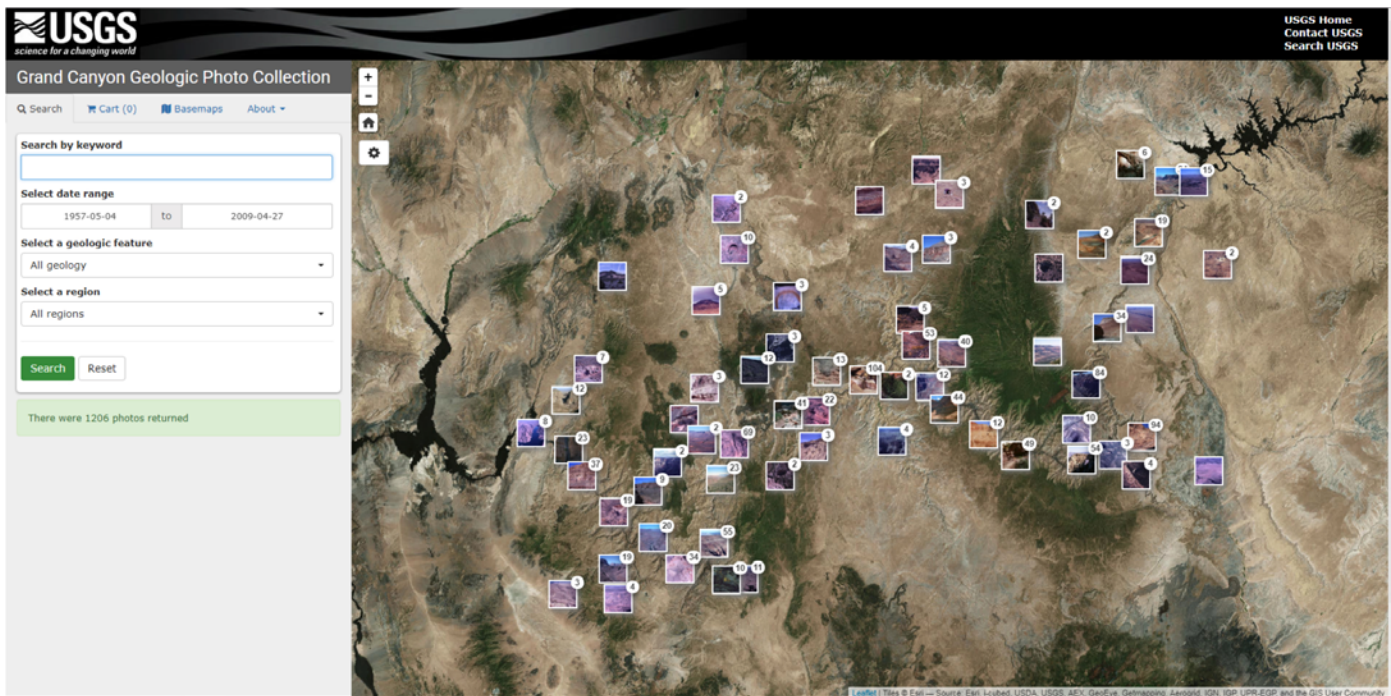
*Principal investigator.*—Sarah Nagorsen (USGS)

*Co-investigators and partners.*—Jason T. Sherba (USGS), Drew A. Ignizio (USGS), and Christopher E. Soulard (USGS)

Geotagged photographs have become a useful medium for recording, analyzing, and communicating Earth science phenomena. Despite their utility, many field photographs are not published or preserved in a spatial or accessible format—often-times because of confusion about photograph metadata, a lack of stability, or user customization in free photo sharing platforms. After receiving a request to release about 1,210 geotagged geological field photographs of the Grand Canyon region, we set out to publish and preserve the collection in the most robust (and expedient) manner possible (fig. 6). We leveraged and reworked existing metadata, JavaScript, and Python tools and developed a toolkit and proposed workflow to display the photograph collection in an interactive web mapping application populated from ScienceBase. The open-source tools generated from this project can be used as a pathway for making large collections of photographs available on ScienceBase and in web mapping applications. The proposed workflow could serve as a starting point for a USGS photograph publishing process.

*Related links.*—

- See more at <https://www.sciencebase.gov/catalog/item/58b5f81ee4b01ccd54fde479>
- Adobe Bridge javascript code to write metadata to photo EXIF: <https://my.usgs.gov/bitbucket/projects/CDI/repos/photometadata-webmap-cdi/browse/GeotagPhotoMetadata>
- Javascript and python code for web map application: <https://gitlab.cr.usgs.gov/wgeogdev/sb-photo-map>



**Figure 6.** Screen capture of the user interface for the Grand Canyon photo collection web map application. From Nagorsen and others (2017).

## Visualizing Community Exposure and Evacuation Potential to Tsunami Hazards Using an Interactive Tableau Dashboard

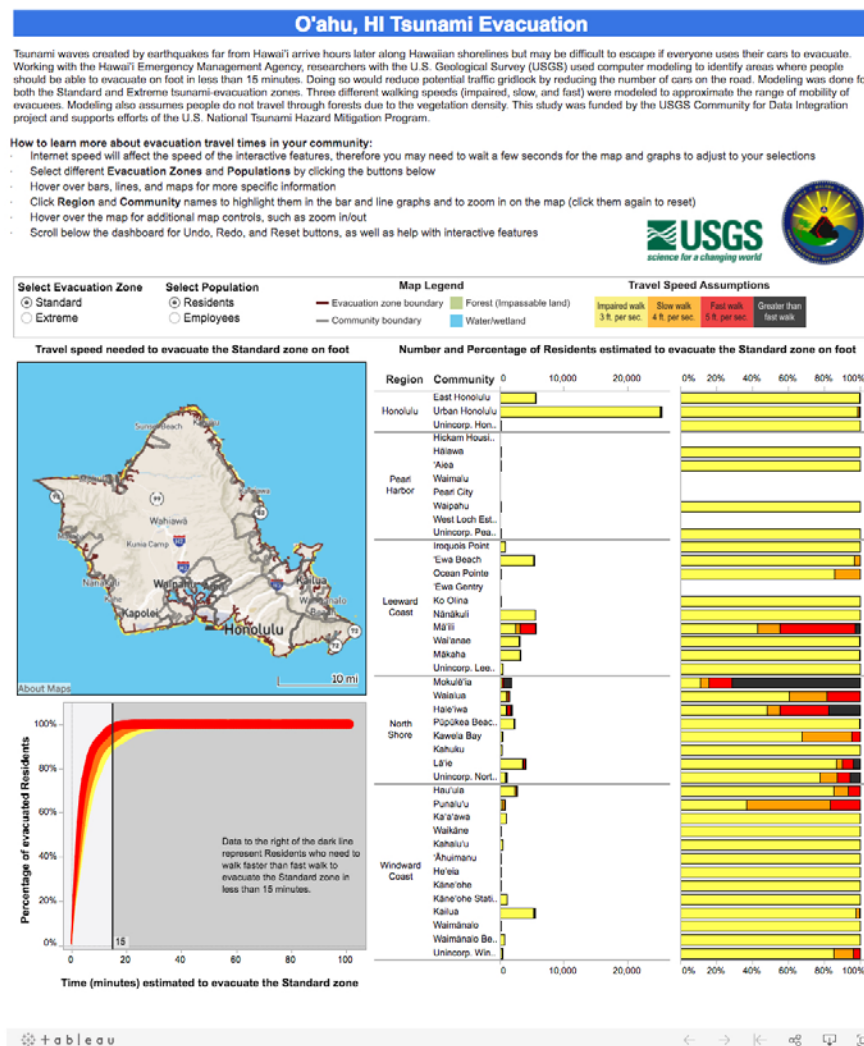
*Principal investigator.*—Jeff Peters (USGS)

*Co-investigators and partners.*—Kevin D. Henry (USGS) and Jamie L. Jones (USGS)

Risk reduction planning organizations across the United States rely on USGS science to inform community exposure and evacuation potential to natural hazards. Currently, USGS science is shared in published reports and journal articles that contain static maps, figures, and tables. Interactive graphics to visualize this science would allow interested parties to tailor the content, form, and appearance of a vulnerability analysis to best suit their specific planning needs. This project evaluated the process of disseminating hazard-exposure data using the Tableau software to provide interactive interpretation of maps and results, using the case study of community exposure and pedestrian evacuation for tsunami hazards on the island of Oahu, Hawaii (fig. 7). Tableau strengths include ease of use, reusability, data interoperability, and visual capabilities, whereas weaknesses include cost, slow online performance, lack of mobile friendliness, requirement of preformatted data, limited visualizations, and poor mapping functionality.

*Related links.*—

- See more at <https://www.sciencebase.gov/catalog/item/58b5f9d9e4b01ccd54fde47d>
- Dashboard: <https://geography.wr.usgs.gov/science/vulnerability/oahuEvacDashboard.html>
- Code repository: [https://my.usgs.gov/bitbucket/projects/CDI/repos/evac\\_dash/browse](https://my.usgs.gov/bitbucket/projects/CDI/repos/evac_dash/browse)
- Annotated presentation with cost/benefit analysis: [https://www.sciencebase.gov/catalog/file/get/58b5f9d9e4b01ccd54fde47d?name=OahuEvacTableau\\_CDI\\_Presentation.pptx](https://www.sciencebase.gov/catalog/file/get/58b5f9d9e4b01ccd54fde47d?name=OahuEvacTableau_CDI_Presentation.pptx)



**Figure 7.** Screen capture of the Oahu Tsunami Evacuation dashboard. <https://geography.wr.usgs.gov/science/vulnerability/oahuEvacDashboard.html>. Users can interact with the graphics by clicking and hovering over community names and the graphic to find more information. From Peters and others (2017).



## Exploring the U.S. Geological Survey Science Data Life Cycle in the Cloud

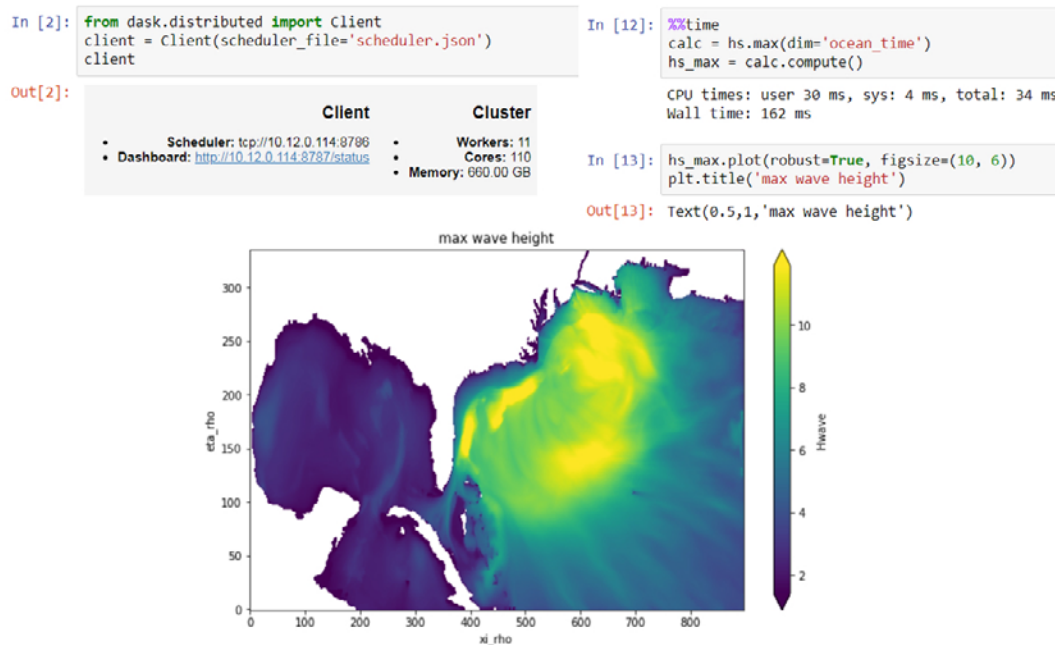
*Principal investigator.*—Richard P. Signell (USGS)

*Co-investigators and partners.*—Nadine Golden (USGS), Joseph D. Hughes (USGS), Andrew Stevens (USGS), Jeremy T. White (USGS), Nathaniel L. Booth (USGS), Andrew C. Ritchie (USGS), Ivan Suftin (USGS), Jonathan Warrick (USGS), Christopher R. Sherwood (USGS), and John C. Warner (USGS)

Traditionally in the USGS, data are processed and analyzed on local researcher computers and then moved to centralized, remote computers for preservation and publishing (for example, ScienceBase or the USGS Publications Warehouse). This approach requires each researcher to have the necessary hardware and software for processing and analysis and also to bring all external data required for the workflow over the internet to their local computer. To explore a more efficient and effective scientific workflow, we explored an alternate model: storing scientific data remotely and performing data analysis and visualization close to the data, using only a local web browser as an interface. Although this environment was not a good fit for the current policies of the USGS Cloud Hosting Solutions, we were able to demonstrate huge efficiency gains using these data-proximate scalable workflows both on the National Science Foundation's Extreme Science and Engineering Discovery Environment (XSEDE) Jetstream Cloud and the USGS Yeti High Performance Computing cluster (fig. 8).

*Related links.*—

- See more at <https://www.sciencebase.gov/catalog/item/58b5e10fe4b01ccd54fde40b>
- Repository on JupyterHub Docker Swarm service: <https://github.com/USGS-CMG/data-life-cycle-cloud-docker-jupyterhub>
- Repository on deploying the Thematic Real-time Environmental Distributed Data Services (THREDDS): <https://github.com/USGS-CMG/data-life-cycle-cloud-docker-thredds>
- Documentation on running THREDDS, pycsw (an OpenGeospatial Consortium Catalog Service for the web server written in python), ERDDAP, ncWMS (a web map service for geospatial data that are stored in Climate & Forest-compliant Network Common Data Form [NetCDF] files) and TerriaMap via Docker: <https://github.com/rsignell-usgs/docker-nginx-thredds-pycsw/blob/master/README.md>



**Figure 8.** Image of snapshots from a Jupyter notebook using a scalable computer cluster on the U.S. Geological Survey Yeti High Performance Computing facility to compute maximum wave height on an 80 gigabyte dataset. The user is interacting with Yeti in their browser on their laptop connected on the Trusted Internet Connections network. Because the processing is happening remotely, the local user does not need a fast internet connection or powerful local hardware. From Golden and others (2017).



## Developing Application Programming Interfaces to Support Enterprise Level Monitoring Using Existing Tools

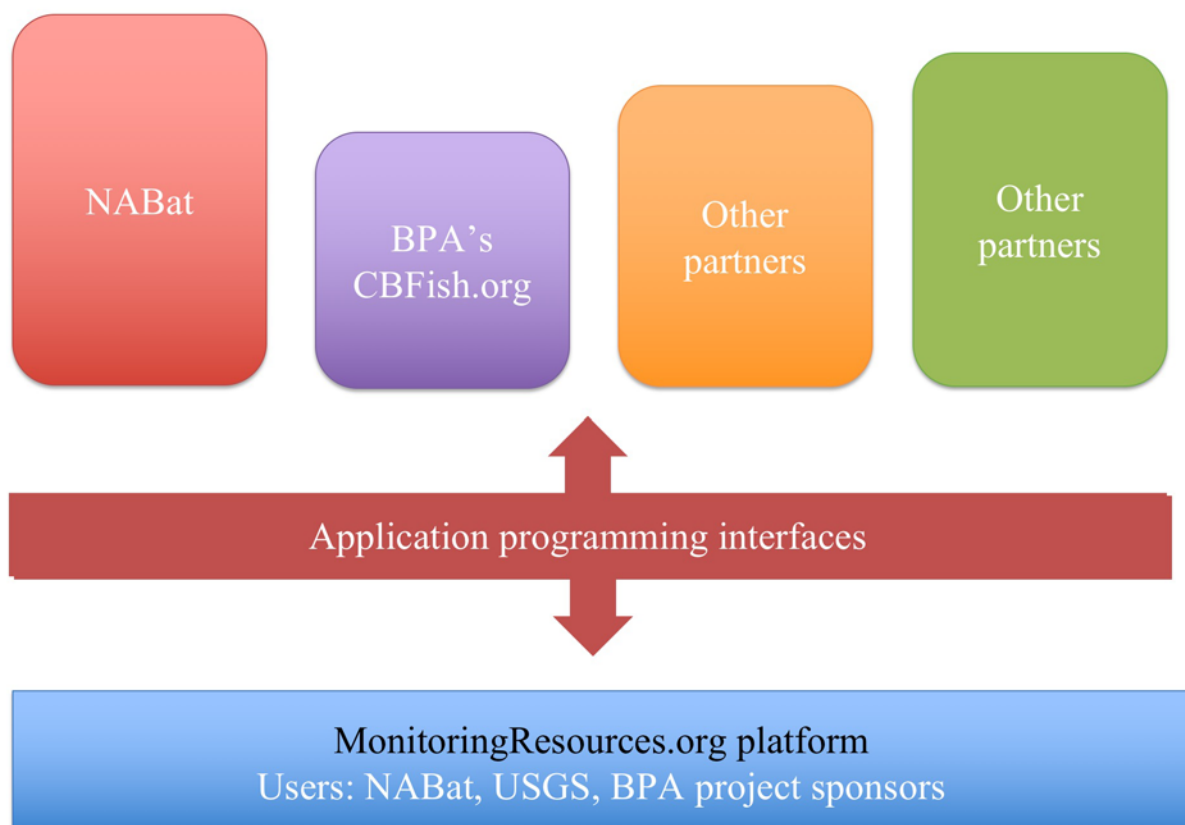
*Principal investigator.*—Brian Reichert (USGS)

*Co-investigators and partners.*—Jennifer Bayer (Pacific Northwest Aquatic Monitoring Partnership), Rebecca A. Scully (Pacific Northwest Aquatic Monitoring Partnership), Jake F. Weltzin (National Phenology Network), and Patricia Stevens (USGS)

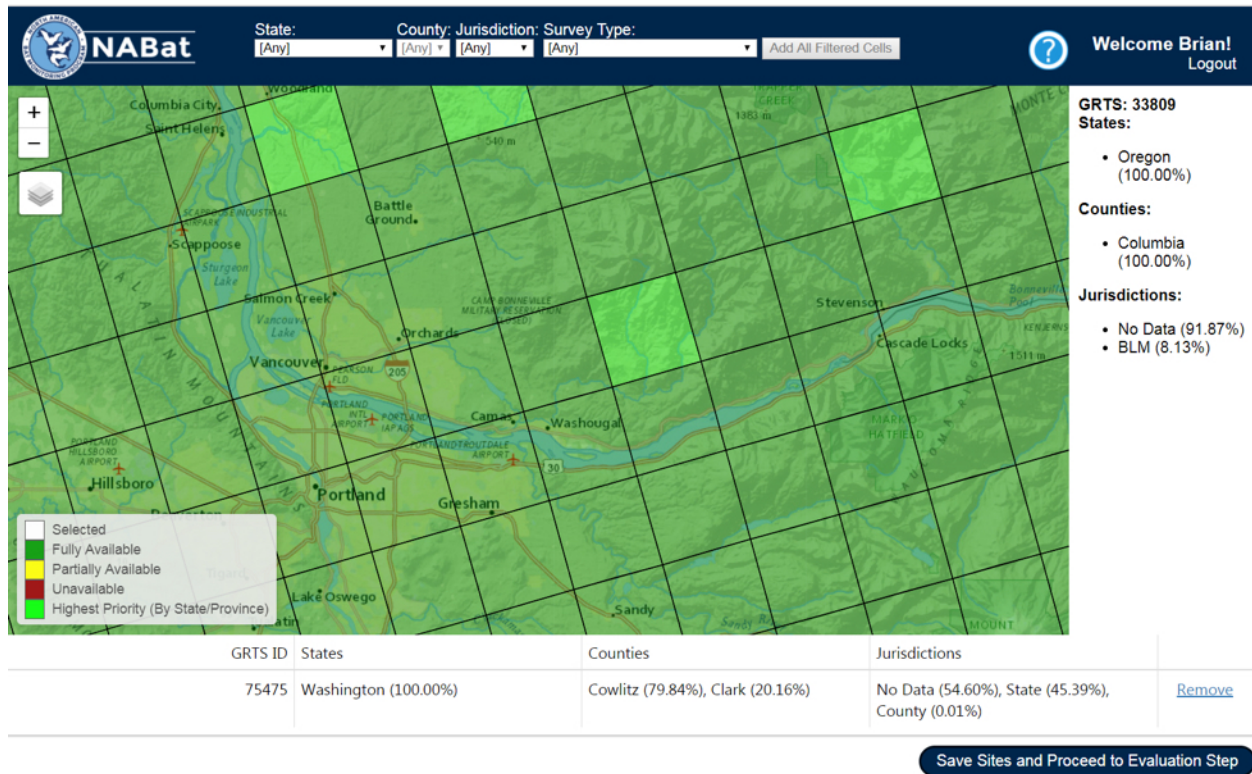
In this age of rapidly developing technology, scientific information such as monitoring data is constantly being gathered across large spatial scales. However, our ability to coordinate large-scale monitoring efforts depends on the development of tools that leverage and integrate multiple sources of data. North American bats are experiencing unparalleled population declines, so the North American Bat Monitoring Program (NABat), a multinational, multiagency coordinated monitoring program, was developed to better understand the status and trends of North American bats. Similar to other large-scale monitoring programs, the ultimate success of NABat relies on a unified web-based data system (fig. 9). Our project successfully developed a program interface between two separate web applications to facilitate data sharing, curation, and access to NABat data (fig. 10). This work expanded the capacity of [MonitoringResources.org](https://www.monitoringresources.org) to enable support of NABat. The additions to [MonitoringResources.org](https://www.monitoringresources.org) are extensible and replicable and can play a critical role in implementing other large-scale monitoring programs in the future.

*Related links.*—

- See more at <https://www.sciencebase.gov/catalog/item/58b5fbdce4b01ccd54fde482>
- Monitoring Resources application programming interface (API) and documentation: <https://mrv1.docs.apiary.io/#>



**Figure 9.** Diagram of that Application Programming Interfaces (APIs) allow the connection of information documented in [MonitoringResources.org](https://www.monitoringresources.org) to multiple applications. From Reichert and others (2017). [NABat, North American Bat Monitoring Program; BPA, Bonneville Power Administration; USGS, U.S. Geological Survey]



**Figure 10.** Screen capture of how a grid cell selection tool web application for the North American Bat Monitoring Program (NABat) uses newly developed application programming interfaces (APIs) to share documentation and design of data collection efforts helping organize, curate, and facilitate access to NABat data.

## Extending ScienceCache Mobile Application for Data Collection to Accommodate Broader Use Within the U.S. Geological Survey

*Principal investigator.*—Mark T. Wiltermuth (USGS)

*Co-investigators and partners.*—Tim Kern (USGS), John (Dell) L. Long (USGS), and Sanjay M. Advani (USGS).

ScienceCache was originally developed as a mobile device data collection application for a citizen science project (fig. 11). ScienceCache communicates with a centralized database that facilitates near-real-time use of collected data that enhances efficiency of data collection in the field. We improved ScienceCache by creating a flexible, reliable platform that reduces effort required to set up a survey and manage incoming data. Now, ScienceCache can be easily adapted for citizen science projects as well as restricted to specific users for private internal research. We improved scEdit, a web application interface, to allow for creation of more-complex data collection forms and survey routes to support scientific studies. Surveys are deployed to iOS and Android mobile devices through the ScienceCache mobile application. Using ScienceCache, any USGS scientist can create and deploy data collection forms to mobile devices. The improved ScienceCache is currently in Beta development.

*Related link.*—

- See more at <https://www.sciencebase.gov/catalog/item/58b5fd17e4b01ccd54fde487>

scEdit - ScienceCache Editor   Public Routes   My Routes   Add New Route   Sign Out

Route Name:

Route Version:

Description:

Warning:

Route Length:       Route Difficulty:

☒ Published

☒ Private

☒ Federal Staff Only

☒ Visible

☒ Allow Unplanned Observation

**Figure 11.** Screen capture of how the improved ScienceCache Editor allows U.S. Geological Survey scientists to easily create data collection forms for internal studies in a private mode with user management capabilities. From Wiltermuth and others (2017).

## Evaluation and Testing of Standardized Forest Vegetation Metrics Derived from Light Detection and Ranging Data

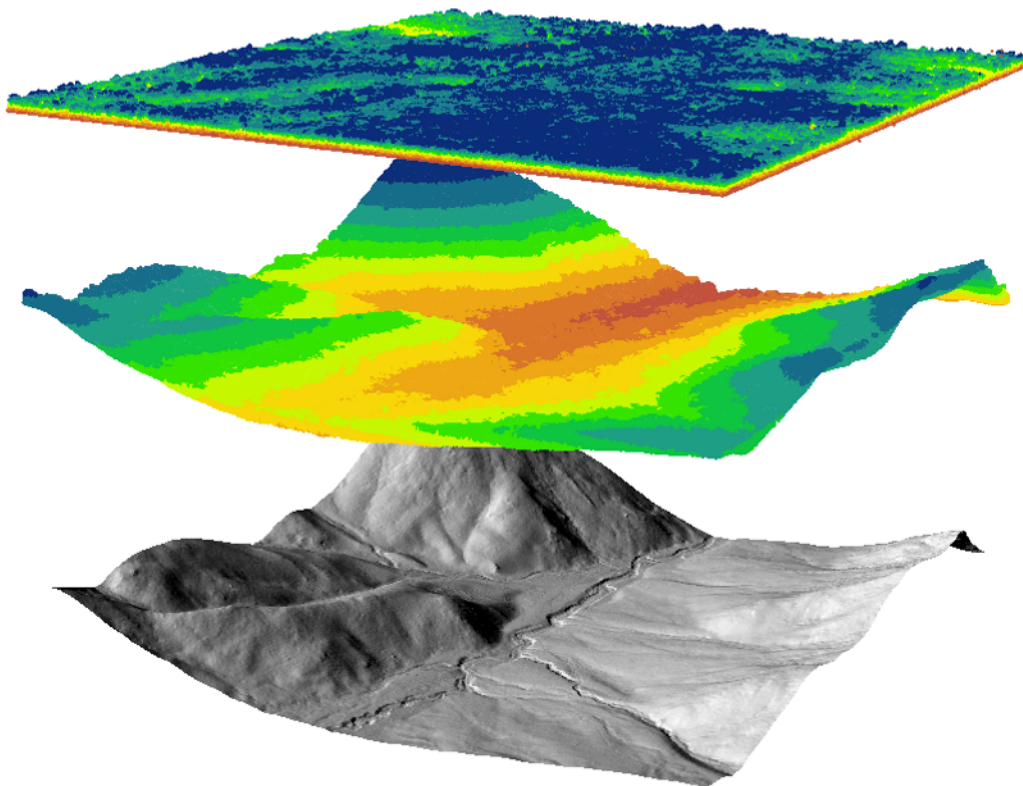
*Principal investigator.*—John A. Young (USGS)

*Co-investigators and partners.*—Monica Palaseanu-Lovejoy (USGS), Jeffrey J. Danielson (USGS), Hans Karl Heidemann (USGS), Kurtis Nelson (USGS), Nicholas Kruskamp (North Carolina State University), Jason Stoker (USGS), William Slocum (North Carolina State University), Cindy Thatcher (USGS), Dean B. Gesch (USGS), and Birgit Peterson (USGS)

Light detection and ranging data are initially collected as three-dimensional point clouds that map the interaction of the airborne laser with earth surface features (fig. 12). Generally, the product of interest has been high-resolution digital elevation models of the ground surface, generated by removing returns from vegetation and other above ground features. However, the full point cloud may be analyzed to characterize the three-dimensional nature of vegetation, including vegetation height variability, structure, and volume. These mapped attributes are useful for habitat studies, vegetation biomass and biomass change studies, and wildfire behavior models. In this project we developed methods to standardize vegetation metric geospatial product generation from USGS 3DEP point clouds using scripting in Python and R. We generated standardized metric products for two National Park pilot areas from five recent lidar acquisitions, and we developed a web front end to demonstrate product search, selection, and download of merged processed tiles at two spatial resolutions.

*Related link.*—

- See more at <https://www.sciencebase.gov/catalog/item/58b5ff4fe4b01ccd54fde490>



**Figure 12.** Image of the data layers involved in extracting height above ground vegetation data from 3DEP light detection and ranging data. From Young and others (2017).

## Summary

The Community for Data Integration funded projects described in this report develop, organize, and share knowledge and best practices in data integration. Products from the projects are intended to build the U.S. Geological Survey knowledge base and capacity for working with data. Fiscal year 2017 products included interactive web applications and their underlying code, code to compare species lists, high-throughput computing workflows, case studies of U.S. Geological Survey science center legacy data inventories, lessons for delivering scientific results to decision makers, documentation of data life cycle workflows in the cloud, application programming interfaces for monitoring tools, mobile application development, and an example of delivering new vegetation metrics derived from lidar products.

## References Cited

- Allstadt, K.E., Nowicki Jesse, A., Schmitt, R.G., Tanyas, H., Thompson, E.M., and Zhu, J., 2018, An interactive web-based application for earthquake-triggered ground failure inventories: ScienceBase release, accessed September 2018 at <https://www.sciencebase.gov/catalog/item/58b5c886e4b01ccd54fde368>.
- Boydston, E.E., Barve, V., Lee, L., Morelli, T.L., and Briggs, J.S., 2017, Automating the use of citizen scientists' biodiversity surveys in iNaturalist to facilitate early detection of species' responses to climate change: ScienceBase release, accessed September 2018 at <https://www.sciencebase.gov/catalog/item/58b5cd07e4b01ccd54fde37f>.
- Erickson, R.A., McCalla, S., Fienen, M.N., Amberg, J.J., Gordon, J., Exter, P.E., Bower, M., and Knudson, J.M., 2017, Flocks of a feather dock together—Using Docker and HTCondor to link high-throughput computing across the USGS: ScienceBase release, accessed September 2018 at <https://www.sciencebase.gov/catalog/item/58b5db5ce4b01ccd54fde3e4>.
- Golden, N., White, J.T., Booth, N.L., Ritchie, A.C., Signell, R.P., Suftin, I., Warrick, J., Sherwood, C.R., Warner, J.C., Hughes, J.D., and Stevens, A., 2017, Exploring the USGS science data life cycle in the cloud: ScienceBase release, accessed September 2018 at <https://www.sciencebase.gov/catalog/item/58b5e10fe4b01ccd54fde40b>.
- Lentz, E.E., Stippa, S.R., Thieler, E.R., Plant, N.G., Gesch, D.B., and Horton, R.M., 2015, Evaluating coastal landscape response to sea-level rise in the northeastern United States—Approach and methods (ver. 2.0, December 2015): U.S. Geological Survey Open-File Report 2014–1252, 26 p., <https://doi.org/10.3133/ofr20141252>.
- Lentz, E.E., Thieler, E.R., Plant, N.G., Stippa, S.R., Horton, R.M., and Gesch, D.B., 2016, Evaluation of dynamic coastal response to sea-level rise modifies inundation likelihood: *Nature Climate Change*, v. 6, p. 696–700. <https://doi.org/10.1038/nclimate2957>.
- Nagorsen, S.E., Sherba, J.T., Ignizio, D.A., and Soulard, C.E., 2017, Web mapping application for a historical geologic field photo collection: ScienceBase release, accessed September 2018 at <https://www.sciencebase.gov/catalog/item/58b5f81ee4b01ccd54fde479>.
- Peters, J., Jones, J.L., and Henry, K.D., 2017, Visualizing community exposure and evacuation potential to tsunami hazards using an interactive Tableau dashboard: ScienceBase release, accessed September 2018 at <https://www.sciencebase.gov/catalog/item/58b5f9d9e4b01ccd54fde47d>.
- Reichert, B.E., Stevens, P., Scully, R.A., Bayer, J.M., and Weltzin, J., 2017, Developing APIs to support enterprise level monitoring using existing tools: ScienceBase release, accessed September 2018 at <https://www.sciencebase.gov/catalog/item/58b5fbdce4b01ccd54fde482>.
- Wiltermuth, M.T., Kern, T., Long, J.L., and Advani, S.M., 2017, Extending ScienceCache—a mobile application for data collection—to accommodate broader use within USGS: ScienceBase release, accessed September 2018 at <https://www.sciencebase.gov/catalog/item/58b5fd17e4b01ccd54fde487>.
- Young, J.A., Palaseanu-Lovejoy, M., Danielson, J.J., Heidemann, H.K., Nelson, K., Kruskamp, N., Stoker, J., Slocum, W., Thatcher, C., Gesch, D.B., and Peterson, B., 2017, Evaluation and testing of standardized forest vegetation metrics derived from lidar data: ScienceBase release, accessed September 2018 at <https://www.sciencebase.gov/catalog/item/58b5ff4fe4b01ccd54fde490>.

