

Preserving Science for the Ages—USGS Data Rescue

Introduction

The U.S. Geological Survey (USGS) is a steward for over 130 years of rich, diverse natural science and information resources. We document one-of-a-kind observations of natural phenomena and cultural impacts on our changing world. In order for society to deal with national and global trends, the USGS must enable access and use of legacy, inaccessible information by including these data in our digital archives and databases.

The USGS has conducted scientific assessments on the quality and quantity of the Nation's water resources, provided access to geospatial and natural resource data, and conducted multi-purpose natural science studies. All of these have generated records that need to be accessible and integrated in order to be examined for new information and interpretations that were never intended by the original collector.

The Federal Records Act of 1950 mandates that the USGS preserve Federal records containing evidence of the agency's organization, functions, policies, decisions, procedures, and essential transactions. At the USGS, the goal of Open Government¹ is to improve and increase access to scientific information. Therefore, it is incumbent upon the USGS to preserve, make available, and provide accountability for the data that it creates from our scientific projects.

Data Rescue activities provide access to valuable and critical data assets and make the USGS's rich contributions available for researchers in the U.S. and international science communities. Since its inception in late Fiscal Year (FY) 2006, the Data Rescue Program has received more than 300 requests from scientists across the USGS totaling over 15 million dollars. Through FY 2011, the program has been able to provide assistance of 1.3 million dollars, or roughly 9 percent of the known need. The costs to rescue data has varied from 3,000 dollars to 65,000 dollars per project. Many USGS science records have permanent or long-term value to the Nation and the world on the basis of the substantial impact the scientific investigations have had on public policies or decisions and on the basis of datasets that are one-of-a-kind collections—information that represents observations and events that will never be repeated.

¹http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment/

A few examples from the Data Rescue Program illustrate USGS legacy products that have been preserved and are now more widely accessible, serving scientific needs today and available for the future.

Historical Images of Alaskan Volcanoes

Using Data Rescue funds, the Alaska Volcano Observatory has scanned and digitally archived over 6,000 35 millimeter slides and prints from the library and files of the USGS, the University of Alaska Fairbanks Geophysical Institute, the Alaska Division of Geological and Geophysical Surveys, and other geoscientific organizations. About 1,000 images have been uploaded to public image databases, making them accessible to the public and the research community. These images provide an invaluable time line of conditions at Alaskan volcanoes and include views of dynamic glacial and volcanic landscapes over multiple decades. To access a sampling of these images, please visit the Alaska Volcano Observatory Web site image database at <http://www.avo.alaska.edu/images/>.



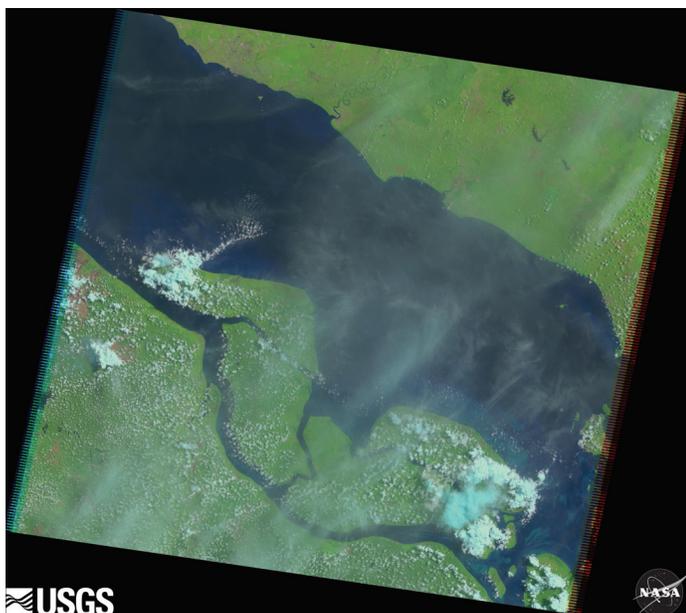
Aerial view of Snowy Mountain Dome, Katmai National Park and Preserve, photographed by USGS volcanologist Wes Hildreth. Before this rescue project, the Alaska Volcano Observatory database featured few historical images of this volcano. Scientists can now compare this image with modern and future views to assess changes in the landscape as a result of eruptions or changing climate.

Diagnostic Records for Wildlife Disease

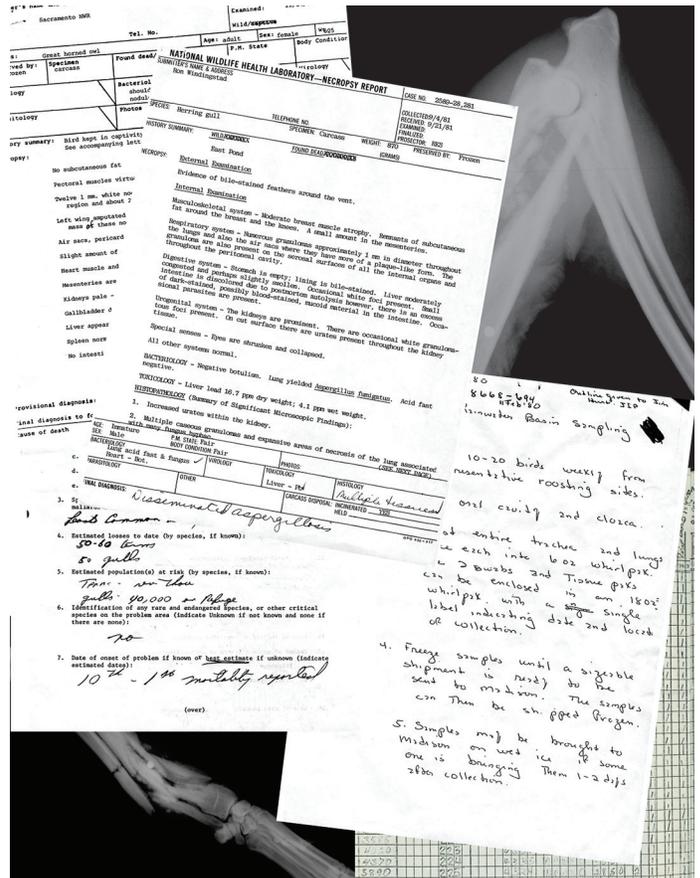
The USGS National Wildlife Health Center provides a national clearinghouse for information on wildlife morbidity and mortality, and houses an extensive collection of diagnostic information pertaining to wildlife disease. Since wildlife, domestic animal, and human diseases are closely intertwined, documenting disease in wildlife not only provides an 'early warning system' for possible human or domestic animal outbreaks, but helps to ascertain ecosystem health at local, regional, and national levels. Much of the textual information before 1990 existed in hardcopy form only. With the help of Data Rescue funding, over 88,000 original documents, such as handwritten and typewritten pages, radiographs, and photographs have been scanned and saved into Portable Document Format documents, representing 23,030 individual diagnostic submissions. Data reviewers have incorporated 9,452 new or updated records into the Center's wildlife mortality database providing easier, consolidated access to critical information. Additional information can be found on the National Wildlife Health Center Web site at <http://www.nwhc.usgs.gov/>.

Expansion of USGS Landsat Archive

The Landsat Mission has been acquiring images from around the world since 1972. In the past four decades, data have been downlinked to satellite ground stations across the globe. The USGS estimates that over 5 million images reside in international archives, and there is a major effort underway to bring all these data back to the USGS. In response to this endeavor, the USGS has received over 500,000 images in the last 2 years from its international partners. These digital files come in all types of formats and on all types of media. Funding for software development and archive storage has come from the Data Rescue Program.



Malaysia is often cloud covered, so obtaining cloud-free satellite imagery is difficult. However, with the use of Data Rescue funds, the



Sample documents from various diagnostic cases examined at the National Wildlife Health Center that have been scanned and added to the clearinghouse.



USGS Landsat archive was able to capture clearer views, as seen above, taken by Landsat 5 Thematic Mapper. (Image date: June 2000.)

In addition, some data stored at the USGS Earth Resources Observation and Science (EROS) Center had been received from an outside source many years ago and could not be read or processed into a useful data product. The Data Rescue Program funded development so that the USGS now has 145,000 of some of the oldest scenes in the archive (some which have never before been seen, to our knowledge) available for viewing and analysis. For more information on the Landsat Program visit <http://landsat.usgs.gov/>.

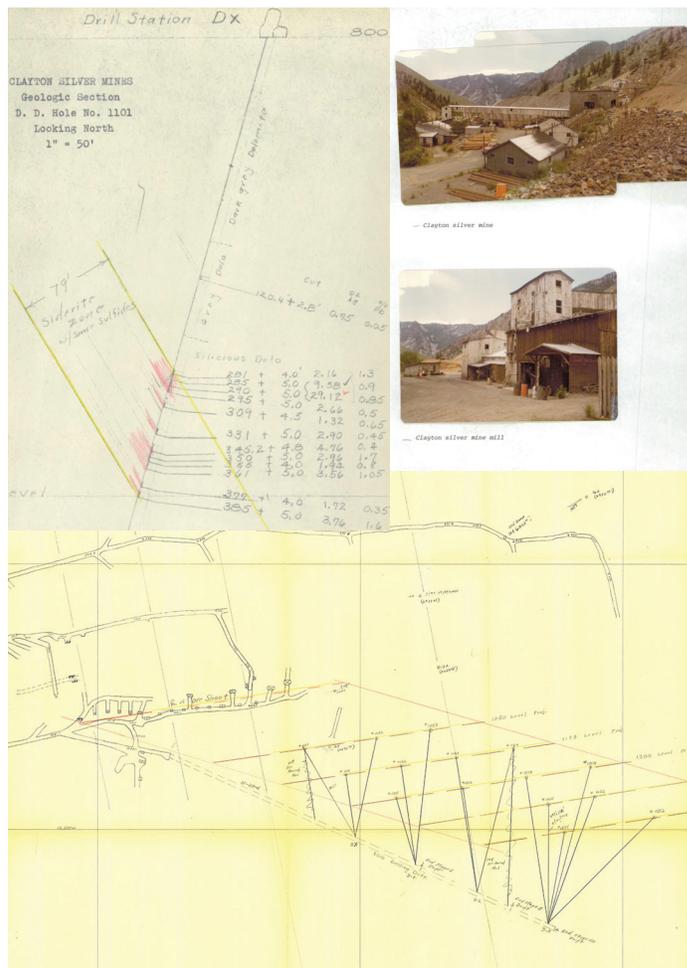
Historical Files from Federal Government Mineral Exploration-Assistance Programs, 1950–1974

With the use of Data Rescue funding more than 5,000 original historical minerals exploration-assistance docket (paper, carbon paper copies, copies of maps, graph copies, blue-line copies of maps, linen or mylar maps, and other fragile and sometimes poor quality media) of the 1950–70 era were scanned and are now available to the public. These dockets provide access to unique or difficult-to-recreate data that are invaluable to land and resource management organizations and the minerals industry. In FY 2011, there were over 1.2 million online requests for dockets from the Web site created with Data Rescue funds. For more information see the Historical Files from Federal Government Mineral Exploration-Assistance Programs, 1950–1974 Web site at <http://minerals.usgs.gov/dockets/>.

Water Resources Records Available only in USGS Archives

Most USGS Water Science Center offices have little-known, but highly valuable sets of historic water resources records in their paper archives. A demonstration project, funded by Data Rescue funds at the Texas Water Science Center, improved the visibility of these archives by digitizing more than 1,500 photographs, field measurements, reports, charts, maps, and other information from over 100 years of data collection at a single streamflow station. The documents include information beneficial to scientific efforts including hydrology, biology, climate change and geomorphology, plus the records assist genealogical investigations, historical society studies, and even USGS historians. Follow-on data rescue efforts digitized the complete collections of more than 1,800 streamflow-sediment-related documents and nearly 350 pump tests for wells in aquifers. The latest rescue efforts focused on digitizing a large collection of detailed supporting documentation of post-flood analyses including field notes, photographs, surveyed channel cross sections, and newspaper articles about individual flood events at USGS streamflow stations.

A template is available to facilitate working with other agencies to rescue additional records to complement the modern data they fund USGS to collect. A USGS fact sheet (<http://txpub.usgs.gov/archives/FactSheet.pdf>) highlights the digital data now available from our historical paper archives and directs users to a comprehensive Web site (<http://tx.usgs.gov/archives/>), which describes the benefits of the newly digitized data holdings and allows users to search a number of categories, filter by dates, and download documents of interest.



Examples of historical mineral exploration documents that have been scanned and are now available online.



Streamgage documents being made available through Data Rescue projects.

Historical Files from USGS North American Bird Phenology Program (BPP), 1880–1970

With the assistance of Data Rescue funding, more than a million original historical bird migration records documenting arrival and departure dates for over 870 species across North America were electronically scanned. In an innovative project to curate the data and make them publicly available, the scanned records were uploaded to the Internet, where volunteers worldwide transcribed the records and added them to a database for analysis. This dataset is unique in its coverage, temporal breadth, comprehensiveness, and comparative value to modern measures of climate change. Its rescue, digitization, and distribution provide unprecedented insight into bird migratory behaviors, including their responses to climate change. For more information, see the Web site for the North American Bird Phenology Program at <http://www.pwrc.usgs.gov/bpp/>.



Original historical bird migration records.

The USGS is the Nation's and the world's leading natural science and information agency—an authoritative source of scientific datasets and the keeper of information. Scientific inquiries within and outside the USGS will benefit from the enhanced accessibility of our decades of observational data and analysis. The USGS has a substantial role to play in providing the information and tools needed to address pressing societal issues and to help ensure the sustained health, welfare, and prosperity of the Nation. The Data Rescue Program demonstrates the long-term commitment to protecting and preserving USGS' scientific information for the Nation and the world for as long into the future as possible.

For more information on Data Rescue activities, contact:

Carol Wippich
Data Rescue Project Manager
Administrative and Enterprise Information
U.S. Geological Survey
12201 Sunrise Valley Drive, MS 807
Reston, VA 20192

Nancy Sternberg
Chief, Office of Information and Investment Management
Administrative and Enterprise Information
U.S. Geological Survey
12201 Sunrise Valley Drive, MS 807
Reston, VA 20192