GeoResJ 6 (2015) 108–114

Contents lists available at ScienceDirect

GeoResJ

journal homepage: www.elsevier.com/locate/GRJ

Rescue of long-tail data from the ocean bottom to the Moon: IEDA Data Rescue Mini-Awards

Leslie Hsu [a,*], Kerstin A. Lehnert [a], Andrew Goodwillie [a], John W. Delano [b], James B. Gill [c], Maurice A. Tivey [d], Vicki L. Ferrini [a], Suzanne M. Carbotte [a], Robert A. Arko [a]

[a] Geoinformatics Center, Lamont-Doherty Earth Observatory, Columbia University, 61 Route 9W, Palisades, NY 10964, USA
[b] Department of Chemistry, University at Albany (SUNY), Albany, NY 12222, USA
[c] Department of Earth and Planetary Sciences, University of California, Santa Cruz, 1156 High St., Santa Cruz, CA 95064, USA
[d] Woods Hole Oceanographic Institution, 266 Woods Hole Rd., MS #22, Woods Hole, MA 02543, USA

ARTICLE INFO

Article history:
Received 12 October 2014
Revised 19 February 2015
Accepted 20 February 2015

Keywords:
Data rescue
Persistent identifiers
Long-tail data
Lunar samples
Analytical geochemistry
Ocean-bottom magnetics

ABSTRACT

Over the course of a scientific career, a large fraction of the data collected by scientific investigators turns into data at risk of becoming inaccessible to future science. Although a part of the investigators' data is made available in manuscripts and databases, other data may remain unpublished, non-digital, on degrading or near obsolete digital media, or inadequately documented for reuse. In 2013, Integrated Earth Data Applications (IEDA) provided data rescue mini-awards to three Earth science investigators. IEDA's user communities in geochemistry, petrology, geochronology, and marine geophysics collect long-tail data, defined as data produced by individuals and small teams for specific projects, tending to be of small volume and initially for use only by these teams, thus being less likely to be easily transferred or reused. Long-tail data are at greater risk of omission from the scientific record. The awarded projects topics were (1) Geochemical and Geochronological data on volcanic rocks from the Fiji, Izu-Bonin-Mariana arc, and Endeavor segments of the global mid-ocean ridge, (2) High-Resolution, Near-bottom Magnetic Field Data, and (3) Geochemistry of Lunar Glasses. IEDA worked closely with the awardees to create a plan for the data rescue, resulting in the registration of hundreds of samples and the entry of dozens of data and documentation files into IEDA data systems. The data were made openly accessible and citable by assigning persistent identifiers for samples and files. The mini-award program proved that a relatively small incentive combined with data facility guidance can motivate investigators to accomplish significant data rescue.

1. Introduction

Today, most data held by active Earth scientists are data at risk because they are in formats that do not permit full electronic access to the information they contain. Examples include non-digital data, data on obsolete or near-obsolete digital media, insufficiently-described data, and even digitally-acquired data that cannot be ingested into managed databases because they lack adequate formatting or metadata [26]. Other examples of data at risk are data in proprietary formats and data stored in old file formats that are not compatible with newer software versions. As investigators approach retirement, this data at risk increases its probability of exclusion from the scientific record. The loss of such data would be unfortunate for several reasons, including lost opportunities for synthesizing and reusing data, the lost monetary value of the funded research, inaccessibility of original sites, and irreproducibility of previous studies. Data at risk can be rescued by digitization, format migration, treating damaged storage media, adding metadata, or any action taken to make data accessible in the long term.

Long-tail data is defined as data produced by individuals and small teams for specific projects, that "tend to be small in volume, local in character, intended for use only by these teams, and are less likely to be structured in ways that allow data to be transferred easily between teams or individuals" [51]. However, even if initially intended for a specific research question, synthesis and reuse of the data, including for purposes other than the original intent, is

---

* Corresponding author. Tel.: +1 (845) 365 8484.
E-mail addresses: lhsu@ldeo.columbia.edu (L. Hsu), lehnert@ldeo.columbia.edu (K.A. Lehnert), andrewg@ldeo.columbia.edu (A. Goodwillie), jdelano@albany.edu (J.W. Delano), gillord@ucsc.edu (J.B. Gill), mtivey@whoi.edu (M.A. Tivey), ferrini@ldeo.columbia.edu (V.L. Ferrini), carbotte@ldeo.columbia.edu (S.M. Carbotte), arko@ldeo.columbia.edu (R.A. Arko).

http://dx.doi.org/10.1016/j.grj.2015.02.012
2214-2428/

one of the great benefits of rescuing long-tail data. Long-tail data are also particularly at risk because they have less-established documentation standards, formats, and community-trusted repositories than that required of data collected in large-scale data collection campaigns such as satellite data and network seismic data. Long-tail data are highly diverse in data type, collection method, and processing method. For this reason, domain-specific and community-guided repositories are well-suited for serving high-quality, trusted data that is suitable for reuse, because they understand the data and their scientific meaning and application, and are thus responsive to community requirements and concerns (e.g. [22]). A major component of rescuing long-tail data entails ensuring that adequate metadata are documented for proper reuse of the data.

The chemical analysis of geologic samples is the basis for scientists' understanding of the Earth and Moon's composition and evolution. These physical samples and the data derived from them also require curation. The need for unambiguous sample identification and proper documentation of sample metadata describing physical samples motivated the development of the International GeoSample Number (IGSN) (e.g. [23]). Its 9-digit alphanumeric codes resolve in a sample registry, SESAR (System for Earth Sample Registration, www.geosamples.org) to a metadata profile that includes description, collection, location, and archive information. The IGSN code does not replace the investigator-given sample name, but is designed to be a unique code that is user-friendly for humans, short enough to be used on sample labels and in data and publication tables, and to link the full sample metadata profile to resources such as publications and websites. The best practice is to assign IGSNs immediately upon specimen collection in the field, such as onboard the ship as the dredge samples are being described and cataloged, or at the outcrop when collecting a land-based sample. Then, the IGSN remains associated with the sample as it is sent to laboratories, subsampled, analyzed, and used in any publication based on the sample or the data derived from the sample.

Data rescue mini-awards were established in 2013 by Integrated Earth Data Applications (IEDA, http://www.iedadata.org/) to preserve valuable long-tail datasets that are in danger of being lost by degradation or investigator retirement. IEDA is a U.S. National Science Foundation-funded facility to support the solid Earth sciences. The data rescue competition was announced via e-mail lists and newsletters serving the IEDA community which includes geochemistry, petrology, geochronology, and marine geology researchers. The applications were judged by the IEDA User Committee, a group of scientists and users familiar with IEDA tools and databases whose primary responsibility is to evaluate usability, performance, and utility of IEDA data systems. Awards were selected on the basis of highest impact on future research by quality, size, rarity, and unique location or data type. Data from the funded projects would be made openly accessible to the community by inclusion in the IEDA data collections. Each successful proposal received $7000 to support proper compilation, documentation, and transfer of data.

IEDA was well-positioned to conduct a multi-discipline data rescue effort because it operates diverse community-driven databases and tools including data repositories (e.g. the EarthChem Library, Marine Geoscience Data System, and USAP Data Center), registries (e.g. the System for Earth Sample Registration, SESAR), and global syntheses (e.g. PetDB – The Petrological Database and the Global Multi-Resolution Topography (GMRT) synthesis). IEDA assigns persistent identifiers to files (DOIs – Digital Object Identifiers) and samples (IGSNs – International GeoSample Numbers) in order to promote unambiguous and citable identification and access to data and metadata. The IEDA team also includes staff who have built relationships within their domain communities. This paper summarizes the three data rescue projects funded in 2013, their challenges for data rescue, and lessons learned.

2. The 2013 IEDA data rescue projects

Each of the three projects was selected because of the importance of the data, the relevance to IEDA data systems, and completeness of the data rescue plan as outlined in the proposal. Two projects dealt with data collected on marine research cruises, and the third with samples retrieved from the moon. All projects involved data collected over several decades, which were at risk of disappearing from the research community if not properly documented and made available before retirement of the investigator.

Guidance by the IEDA facility removed the large barrier of uncertainty on how to begin and proceed with data rescue. Each project investigator interacted with IEDA staff who had domain training, and together they determined the best route for producing usable data products. Consideration of data from an initial evaluation and planning phase was completed by an in-person visit, phone call, or email conversation. Together, the investigator and IEDA staff agreed on a list of data products, and periodic check-ins occurred until the publication of the final products.

2.1. Project 1: Sample curation: Geochemical analyses and sample metadata from Fiji, Izu, and Endeavour

Samples dredged and grabbed from the ocean floor from research ships or with ROVs (Remotely-Operated Underwater Vehicles) yield data about the geology and chemical composition of rocks formed in active tectonic areas. The chemical composition of a sample, together with its collection location, form the basis of scientific arguments about the formation of the Earth's crust (e.g. [53,30]), for example in the Lau Basin and the Fiji Arc (e.g. [14,15]). Sample metadata and geochemical data from four decades of research on intra-oceanic arcs and spreading centers were compiled and organized by James Gill of the University of California, Santa Cruz. Three study areas were targeted: the Endeavour segment of the Juan de Fuca ridge (Fig. 1), the Fiji Arc, and the Izu Arc.

The identification, management, and curation of these physical geological samples are of critical importance for the proper interpretation of their results, and for possible reuse in future scientific studies. Once they are collected from the ocean floor, samples are often kept in the private collections of individual scientists. The samples are prepared and then analyzed for different elemental and isotopic abundances. Some samples are shared or split for analysis in several laboratories.

Gill and colleague Erin Todd recount an instance where inadequate sample management led to questionable scientific conclusions:

A scientifically important paper in 2001 (by Woodhead et al. [52]) argued that, contrary to near-universal belief, Hf is transported from the subducting slab into the overlying mantle and eventually to arc volcanoes. The key measurement was the one backarc basalt called "PPTUW" in this paper.

Subsequent efforts to confirm the observation ran into problems. The apparently-same sample was variously called PPTU, PPTUW/5, PPTUW-1, and TVZ19 in four other papers that reported its major element, trace elements, and other isotopes. None of those papers gave its latitude and longitude or cruise metadata, but figures in two of the papers showed its location at ∼34°S in the backarc basin. Many emails and phone calls to many people unearthed that the sample had been collected on one of five SIO cruises in 1985–86 that were all part of the PPTUW Expedition. The fifth and last of those cruises went to the area shown on maps, and metadata for this cruise were found at

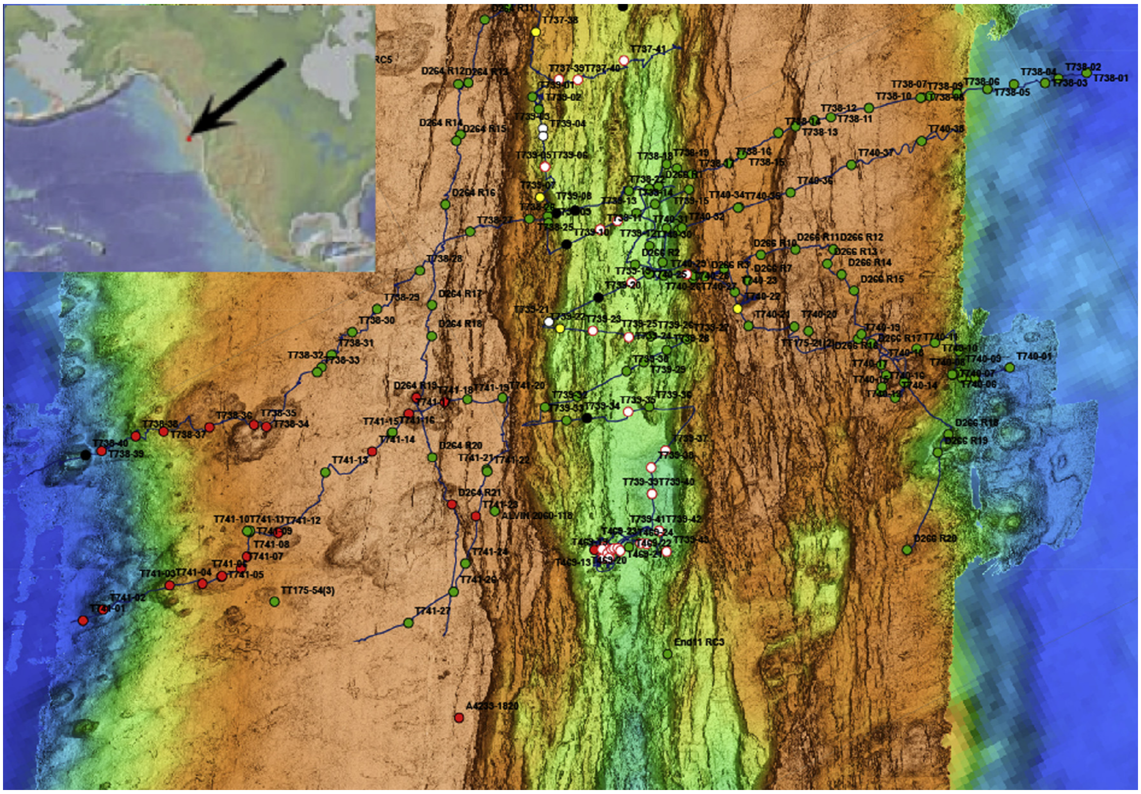

Fig. 1. Map of the Endeavour segment of the Juan de Fuca Ridge spreading center, with sample locations and detailed topography. For full view, see Gill [16]: Endeavour segment ISS (Integrated Study Site) high-resolution map of the location of analyzed basalt samples. Integrated Earth Data Applications (IEDA). http://dx.doi.org/10.1594/IEDA/100407. Inset map made with GeoMapApp (www.geomapapp.org).

SIO. One basalt sample from the cruise – Dredge 41 – was found in the lab of the deceased SIO professor who had been its Chief Scientist (Harmon Craig), and Hf isotope data for it were published in a paper [whose results are now in IEDA with IGSN numbers] [50]. However, the new Hf isotope result differed from the one in the 2001 paper and contradicted its conclusion. We think that the sample in the other papers was not collected where it is shown on the maps but, instead, was Dredge 40 from PPTUW Cruise 5 in deeper water where similar basalts were later found by Todd et al. [50] The new information led to a new, more nuanced interpretation of magma genesis during the rifting stage of backarc basin development.

Had all the data referenced the same IGSN number linked to correct metadata for the sample, all this would have been clear at the outset and weeks of effort could have been avoided.

Gill's data rescue project assigned IGSNs to his collection of samples from Fiji, Izu, and the Endeavour segment. Some of the sample metadata had been reported in papers and in the GEOROC and PetDB databases, but the data rescue award allowed the metadata to be checked for completeness, and additional database fields such as latitude, longitude, rock classification, and cruise ID were filled in. Completeness of metadata greatly aids in search and discovery of samples. Gill also compiled the geochemical data for the three areas from multiple publications and established links from the geochemical data to the sample registry and published papers. Additionally, Gill digitized some data that were formerly available only in a book chapter [19] and supplied these in a standard IEDA EarthChem geochemical format [17]. Other information provided by Gill included sample location maps overlaid on high-resolution bathymetry (Fig. 1), and supplementary explanations of analytical techniques used in the Gill lab to aid investigators when reproducing or comparing results.

The data rescue effort increased discoverability and accessibility by linking existing related datasets and samples together and homogenizing data that were entered by different people and through different interfaces. Much effort was spent compiling information on all of the samples, deciding with IEDA staff the appropriate level of documentation, matching samples to IGSNs and publications, and making links between IEDA systems and publications. The data rescue project brought together the regional compilations in the EarthChem Library, specified metadata to describe the resource in a way aligned with the DataCite standard [1,31], assigned a persistent DOI (digital object identifier), and made it discoverable via the EarthChem Library interface. Some datasets were assigned release dates in the future, allowing manuscripts to be published before release of the data (e.g. [18]). All products are gathered on a featured dataset page in IEDA EarthChem, http://www.earthchem.org/featured/gill.

2.2. Project 2: Standardizing 35 years of evolving technology: Near-bottom magnetic data rescue

High-resolution near-bottom magnetometer data provide valuable insight into the formation, evolution, and structure of oceanic crust, the tectonic evolution of ocean basins, as well as a detailed history of the Earth's magnetic field and its polarity reversals (e.g. [32–34]). Whereas conventional marine magnetic data obtained at the sea surface is filtered by the depth of the ocean, magnetic data obtained near the seafloor eliminates this filtering effect resulting in a better definition of the source magnetism. Such near-bottom magnetic field data can be obtained either by an instrument package towed behind a ship ("Deep Tow"), or by human-occupied vehicles, remotely-operated vehicles, or autonomous vehicles. Deep Tow magnetic data from thirty-five years of research was compiled and documented by Maurice Tivey at Woods Hole Oceanographic Institution (WHOI).

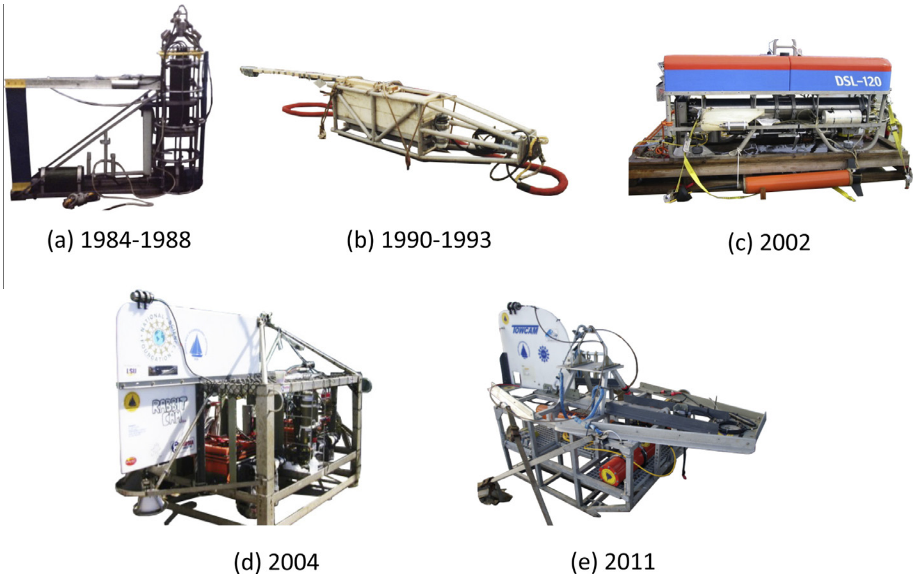
(a) 1984-1988
(b) 1990-1993
(c) 2002
(d) 2004
(e) 2011

Fig. 2. Magnetometer instruments from 1984 to 2011. (a) University of Washington (UW) Deep Tow Magnetometer used from 1984 through 1988. Data recorded on an internal "Memodyne" cassette tape inside electronics housing. (b) UW RealTime Deep Tow Magnetometer, 1990–1993. (c) Woods Hole Oceanographic Institution (WHOI) DSL-120 deep-towed sidescan sonar sled with two magnetometer sensors, 2002. Data sent in real-time up a fiber-optic tow cable. (d) Internally recording WHOI 3-axis vector magnetometer mounted onto TowCam sled, 2004. (e) WHOI RealTime Deep Tow Magnetometer, data sent digitally via the CTD conducting cable and logged on the ship, 2011.

The instruments used to measure magnetic field properties have evolved dramatically over the past thirty years (Fig. 2). The evolution of the sensor package and navigational instruments makes data compilation and alignment over the decades challenging. In addition, storage media over the past three decades has evolved from memodyne cassette tape and 5¼″ floppy disks to the now absent Jaz drives, to CDs and DVDs and USB drives. Tivey revisited original data from each of these types of media, and fortunately had working machines that could read all of these old types of storage media.

For the Deep Tow Magnetometer data, Tivey worked with IEDA staff to determine which files would be the most useful to archive, i.e., which stage of the merged data would be the most useful for future users interested in reprocessing. This included consideration of file formats accessible across a wide range of platforms. In addition to the magnetic sensor, the depth and altitude and ship and towed package navigation data are essential for transforming the raw data into processed, usable values. Tivey compiled the raw magnetic sensor data, the merged magnetic, depth and altitude data, and the processed projected data files used in analysis, as well as the navigation information needed to locate the data track lines. He also constructed documentation files with a standardized set of metadata including date, research vessel, cruise ID, chief scientist, location, instrument, sensor, PI, file list, a full description of column descriptors and units, and a list of the processing steps including any inconsistencies in the data or procedure. An overview file for each dataset contains a description of the equipment, a general outline of the data processing flow, a description of the format, and links to Matlab processing scripts.

The data are curated in the IEDA Marine Geoscience Data System (MGDS, http://www.marine-geo.org), a data repository that holds a wide range of marine geoscience data, including bathymetry, gravity, magnetics, seismic, and image data. The system is designed around collections representing both field expeditions and post-field compilations, with metadata including investigators, cruise or field program ID, geographic location, time of data collection, data types and data-specific metadata. Currently, MGDS serves more than 560,000 data files from over 2500 field programs. Tivey's contributed data are registered in the MGDS associated with the original cruise IDs (e.g. A112-11, http://www.marine-geo.org/tools/search/entry.php?id=A112-11) and each dataset has been published with a DOI. These data are now available to anyone, and the well-documented raw data can be reprocessed with new techniques if necessary [35–49]. Links to all products are gathered at the URL, http://www.iedadata.org/featured/tivey.

2.3. Project 3: Digitizing images and geochemical data from Apollo lunar samples

An enduring legacy of the Apollo program is the lunar sample collection that is currently maintained and curated at Johnson Space Center in Houston, Texas, USA. The samples, obtained at high cost and great risk during the six Apollo Missions are the only samples that have been returned by astronauts from the surface of another planetary body. However, despite the fastidious care and effort with which the lunar samples are stored, handled, distributed and accounted for, and the wealth of published and unpublished data on these samples, there is currently no searchable digital database for the geochemistry of lunar samples. John Delano (University at Albany, NY) has initiated work on such a database by compiling the major and trace element data of lunar volcanic and impact glasses returned by the Apollo 14, 15, 16, and 17 missions.

The goal of the data rescue project was to assemble geochemical analyses obtained for two unique types of lunar samples – lunar volcanic glasses (LVG) and lunar impact glasses (LIG). Due to the abundance of impact-generated craters and basins, LIGs were expected to be found on the lunar surface, but the LVGs were not expected, and were the first samples that confirmed the presence of pyroclastic volcanic activity on the lunar surface [28]. Chemical analyses of the LVGs have resulted in the identification of ∼25 chemically distinct varieties that demonstrate that the Moon's mantle is internally heterogeneous [5]. Isotopic dating of individual LVG samples has demonstrated that these pyroclastic eruptions occurred during the 3200–3800 My time-interval when extensive basaltic volcanism was also occurring [27]. It is generally agreed that the LVGs provide our best constraints on the chemical composition and mineralogy of the Moon's mantle. The LIGs provide scientists with information about the wide compositional range of upper crustal materials on the Moon, as well as information on the history of meteorite bombardment during the last 3900 My.

Before the rescue project began, information on these samples existed in papers but was not readily available to other investigators in digital tabular format. Delano began publishing the chemical analyses of glasses in lunar soil samples in the 1970s (e.g. [3–5]). At the time that the analyses were published, page length constraints in paper journals often allowed only for “representative samples” or plots, instead of full data tables. The Lunar Compendium [25], a rich source of information on lunar samples, also contained detailed descriptions and photos, but only example or average chemical compositions in PDF tables.

Delano worked with IEDA staff to make all of the resources digital and accessible, and to add previously inaccessible thin section photos to the data. In his initial visit to the team, he brought the original large-format photographs of the analyzed thin sections with tracing paper overlays for a discussion about the best way to archive them. He scanned the photographs and digitally labeled the individual glasses, saving the files to PDF format (Fig. 3). Thus each analysis can be tracked to the individual glass bead that was measured. IEDA EarthChem data templates [29] were used to document both the analytical values and the sample and procedural metadata, such as laboratory, instrument, date, and sample lunar geocoordinates (selenographic coordinates). The templates ensure metadata completeness and capture the data needed for inclusion in geochemical syntheses of well-used data models such as PetDB [24]. Once in a synthesis database, data is searchable by location and composition, providing a much higher level of discoverability and accessibility.

The completed templates and associated images were made available through the IEDA EarthChem Library, organized by sample and glass type, with DOIs as persistent, citable identifiers [6–13], (http://www.earthchem.org/featured/delano). Users can access the geochemistry and photos online, identify samples for reanalysis, and request the samples through Johnson Space Center. As a result of this data rescue by Delano, the full datasets and related metadata for thousands of lunar glasses have been made electronically and openly accessible for the first time.

3. Common themes addressed by the IEDA data rescue projects

Without concerted effort, useful research data is at risk of remaining inaccessible and being permanently lost to future generations. Each of the three data rescue projects described here had unique needs and challenges but the final outcomes for all projects addressed some common themes for rescuing data-at-risk:

(a) In each project, accessibility of the data was increased. Data were digitized, converted from obsolete formats, and placed with similar domain data so that they could be discovered and accessed. Data that were previously locked behind publisher pay walls were made open-access through the IEDA website. Disparate data from multiple field campaigns and data systems were compiled for easier access to related data.

(b) Additional documentation was added to the datasets in order to make the data reusable. When only the original investigator holds the necessary metadata for reuse, the data is often sparsely documented. But when placed in a public repository, much more explanation is required by the data facility. The three projects took advantage of the investigators’ decades of experience to help determine what documentation was necessary for appropriate evaluation and reuse of the data. Examples of additional documentation

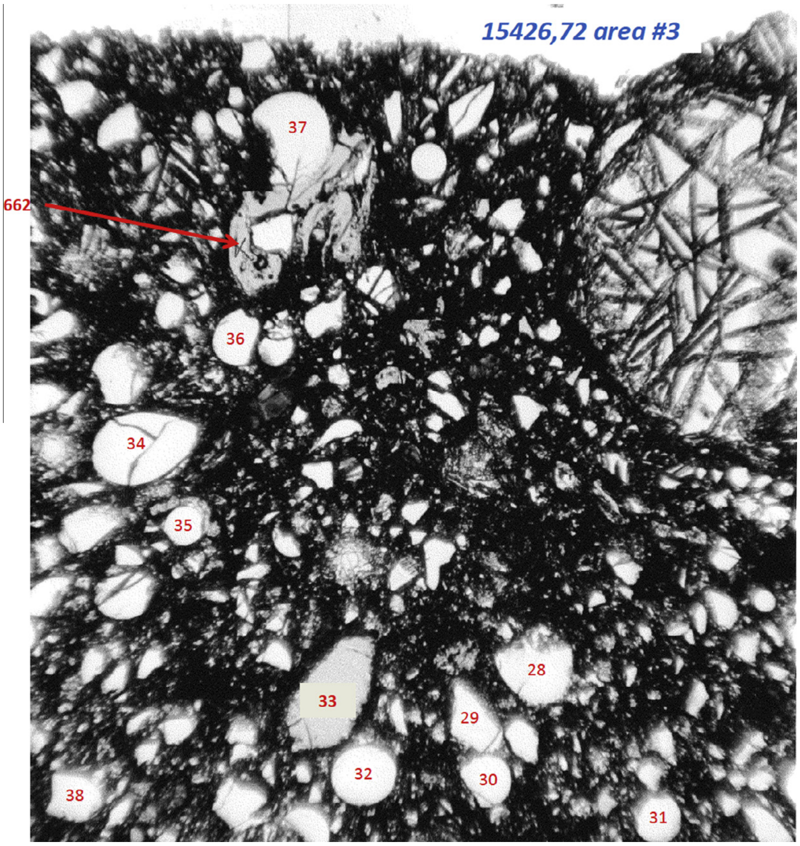

Fig. 3. Scanned and labeled image of a thin section for lunar sample 15426,72.

supplied include descriptions of laboratory analytical procedures and descriptions of procedures for transforming raw to processed data. Data formats were explained and acronyms expanded so that users with different domain training can better manipulate the data. Adequate documentation of datasets for reuse requires time, energy, and is aided by a data facility to guide the data producer.

(c) Data from decades of research were standardized into common formats applicable to the data type. Standardization of research data is hard to achieve during the course of a research project, because procedures, storage formats, and software all change with time. In these data rescue projects, data were standardized into community-approved formats such as the SESAR sample metadata profile and the IEDA EarthChem analytical geochemistry data templates which comply with the DataCite standard. Standardization of data documentation and formats is of key importance for the purpose of making data reusable.

(d) Persistent unique identifiers were assigned to the data files and samples thus increasing long-term accessibility and citability. The three projects spanned the spectrum of sample and file management: the magnetic data project had no samples, the lunar project had an elaborate curation system for the samples via NASA, and the Fiji, Izu, and Endeavour project had the normal path of individual investigator curation. However, all projects benefited from the assignment of persistent unique identifiers. Persistent and unique identifiers such as IGSNs for samples and DOIs for files are integral to maintaining accurate links to data over time (e.g. [21]). Citation of resources is an important metric of scientific productivity, and inclusion of the persistent and unique identifier ensures that the cited resource can be accessed long into the future. In addition to assigning DOIs, IEDA provides example citation syntax on its dataset pages, encouraging the data users to cite the datasets in publications and to give appropriate credit to the data producers.

A debated issue is the level of domain knowledge required when dealing with data management and archiving of scientific data (e.g. [20]). Librarians argue that they have competently archived resources for centuries without domain knowledge of all of the different fields. On the other hand, to archive scientific data for reuse, domain training is helpful to determine what metadata is required or missing. Domain expertise can efficiently identify the important metadata and make suggestions for improving the reusability of the data. A mix of skills from researchers, librarians, and IT professionals is necessary for proper data rescue [2].

4. Lessons learned

The 2013 IEDA mini-awards marked the first round of its data rescue competition and both the awardees and IEDA learned valuable lessons about the process, summarized below.

- A small funding award appears to motivate investigators to embark on large data rescue missions that otherwise would not be tackled. In this case, an award of $7000 per project, sometimes split between multiple investigators, enabled the documentation of hundreds of samples, analyses and data files.
- Good communication between the investigator and the data facility is essential throughout the data rescue effort and helps to reduce the overwhelming nature of large data rescue projects. Some projects may only require one initial meeting to agree on the tasks required, but others of more complexity may require multiple calls or emails per week to address questions as they come up.
- Awardees noted the importance of taking control of one's own research legacy because only they have the knowledge to assess the completeness, validity, and reusability of the archived data. For example, gaps in previously-published data were filled by the data rescue projects.
- Awardees found that it was most efficient to first rescue the most recent (and most completely documented) data, gradually working back to the oldest and most challenging data.
- Since the nature and type of long tail-data are vast, there is no established workflow for long-tail data rescue. The specific needs of the rescue project were determined by both the investigator and the domain data facility.

The three inaugural data rescue projects funded by IEDA have rescued long-tail data at risk by digitizing analog data and data stored on near-obsolete media, by sufficiently describing data to allow reuse, by creating file formats that permit full electronic access, and by promoting citation of datasets through the assignment of persistent identifiers. The success of the data rescue projects prompted IEDA to establish a second round of mini-awards, further rescuing data at risk.

Acknowledgments

This work was funded by the IEDA cooperative agreement, U.S. NSF Award 0950477. The authors would like to thank Megan Carter and three anonymous reviewers for comments that improved the manuscript.

References

[1] Brase J. DataCite – a global registration agency for research data. In: Cooperation and promotion of information resources in science and technology. COINFO'09. Fourth international conference on, IEEE; 2009. p. 257–61. http://dx.doi.org/10.1109/COINFO.2009.66.
[2] Davidson J, Jones S, Molloy L, Kejser UB. Emerging good practice in managing research data and research information within UK Universities. Procedia Comput Sci 2014;33:215–22. http://dx.doi.org/10.1016/j.procs.2014.06.035.
[3] Delano JW. Apollo 15 green glass: chemistry and possible origin. Proceedings of the 10th Lunar Planetary Science Conference. Geochim Cosmochim Acta 1979;1(Suppl. 11):275–300.
[4] Delano JW, Livi K. Lunar volcanic glasses and their constraints on mare petrogenesis. Geochim Cosmochim Acta 1981;45:2137–49.
[5] Delano JW. Pristine lunar glasses – criteria, data, and implications. J Geophys Res-Solid 1986;91:D201–13.
[6] Delano JW. Electron microprobe analyses of Apollo 15 green and yellow glasses from Lunar sample 15426,72. Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/100410.
[7] Delano JW. Electron microprobe analyses of Apollo 17 volcanic orange glasses from Lunar sample 74220,128. Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/100411.
[8] Delano JW. Maps showing locations of glasses analyzed from Apollo 16, sample 60014. Integrated Earth Data Applications (IEDA) 2014. http://dx.doi.org/10.1594/IEDA/100455.
[9] Delano JW. Maps showing locations of glasses analyzed from Apollo 16, sample 64001. Integrated Earth Data Applications (IEDA) 2014. http://dx.doi.org/10.1594/IEDA/100456.
[10] Delano JW. Maps showing locations of glasses analyzed from lunar meteorite ALHA 81005, 8. Integrated Earth Data Applications (IEDA) 2014. http://dx.doi.org/10.1594/IEDA/100457.
[11] Delano JW. Electron microprobe analyses of Apollo 14 high-Ti glasses from Lunar samples 14301, 14307, 14313, 14049. Integrated Earth Data Applications (IEDA) 2014. http://dx.doi.org/10.1594/IEDA/100429.
[12] Delano JW. Electron microprobe analyses of Apollo 16 glasses from Lunar sample 64001. Integrated Earth Data Applications (IEDA) 2014. http://dx.doi.org/10.1594/IEDA/100430.
[13] Delano JW. Electron microprobe analyses of Apollo 16 glasses from Lunar sample 60014. Integrated Earth Data Applications (IEDA) 2014. http://dx.doi.org/10.1594/IEDA/100431.
[14] Gill JB. Composition and age of Lau Basin and Ridge volcanic rocks: implications for evolution of an interarc basin and remnant arc. Bull Geol Soc Am 1976;87:1384–95. http://dx.doi.org/10.1130/0016-7606(1978)89<1119:CAAOLB>2.0.CO;2.
[15] Gill JB. Sr–Pb–Nd isotopic evidence that both MORB and OIB sources contribute to oceanic island arc magmas in Fiji, Earth Planet. Sci Lett 1984;68:443–58. http://dx.doi.org/10.1016/0012-821X(84)90129-8.

[16] Gill JB. Endeavour segment ISS (Integrated Study Site) high-resolution map of the location of analyzed basalt samples. Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/100407.

[17] Gill JB. Fiji tonalites. Integrated Earth Data Applications (IEDA) 2014. http://dx.doi.org/10.1594/IEDA/100418.

[18] Gill JB. Endeavour segment ISS (Integrated Study Site) basalt geochemical analyses. Integrated Earth Data Applications (IEDA) 2015. http://dx.doi.org/10.1594/IEDA/100408.

[19] Gill JB, Stork A. Miocene low-K dacites and trondhjemites of Fiji. In: Barker F, editor. Trondhjemites, dacites, and related rocks. Elsevier; 1979. p. 629–49.

[20] Heidorn PB. The emerging role of libraries in data curation and e-science. J Lib Administration 2011;51:662–72.

[21] Klump J, Lehnert K. Use-cases and services for unique identifiers in the geological sciences, EGU General Assembly 2013, held 7–12 April, in Vienna, Austria, id. EGU2013-11809; 2013.

[22] Lehnert K, Hsu L. Data publication: the role of community-based, disciplinary repositories, EGU General Assembly 2013, held 7–12 April, in Vienna, Austria; 2013. p. 13218.

[23] Lehnert K, Klump J. Facilitating research in mantle petrology with geoinformatics. S. 9IKCA00250, Copernicus Society, Frankfurt [online]; 2008. Available from: http://www.cosis.net/abstracts/9IKC/00250/9IKC-A-00250-1.pdf.

[24] Lehnert K, Su Y, Langmuir CH, Sarbas B, Nohl U. A global geochemical database structure for rocks. Geochem Geophys Geosyst 2000;1(5):1.

[25] Meyer C. Lunar Sample Compendium; 2004. Available from: http://curator.jsc.nasa.gov/lunar/lsc/15425.pdf, accessed on 11 September, 2014.

[26] Murillo AP. Data at risk initiative: examining and facilitating the scientific process in relation to endangered data. Data Sci J 2014;12:207–19.

[27] Podosek FA, Huneke JC. Argon in Apollo 15 green glass spherules (15426) – Ar-40–Ar-39 age and trapped argon. Earth Planet Sci Lett 1973;19:413–21.

[28] Reid AM, Ridley WI, Harmon RS, Jakes P. Major element chemistry of glasses in Apollo-14 soil 14156. Geochim Cosmochim Acta 1973;37:695–9.

[29] Rivera T, Lehnert K, Hsu L, Johansson A. Overcoming challenges to making data re-usable: the example of geochemical databases, Abstract ED53B-0780 presented at 2011 Fall Meeting, AGU, San Francisco, Calif., 5–9 Dec; 2011.

[30] Rubin KH, Sinton JM. Inferences on mid-ocean ridge thermal and magmatic structure from MORB compositions. Earth Planet Sci Lett 2007;260:257–76. http://dx.doi.org/10.1016/j.epsl.2007.05.035.

[31] Starr J, Gastl A. isCitedBy: A metadata scheme for DataCite. California Digital Library. D-Lib Magazine, 17(1/2); 2011. doi: 10.1045/january2011-starr.

[32] Tivey MA, Johnson HP. The central anomaly magnetic high: implications for ocean crust construction and evolution. J Geophys Res 1987;92(B12):12685–94. http://dx.doi.org/10.1029/JB092iB12p12685.

[33] Tivey MA, Sager WW, Lee S-M, Tominaga M. Origin of the Pacific Jurassic Quiet Zone. Geology 2006;34:789–92. http://dx.doi.org/10.1130/G22894.1.

[34] Tivey MA, Dyment J. The magnetic signature of hydrothermal systems in slow-spreading environments. In: Rona P, Devey C, Dyment J, Murton B, editors. AGU monograph on diversity of hydrothermal systems on slow spreading ridges, vol. 188; 2010. p. 43–66.

[35] Tivey MA. Raw Deep Tow magnetic field data, Endeavour Ridge segment, Juan de Fuca ridge, cruise A118–41 (1988, Tivey). Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/320790.

[36] Tivey MA. Processed magnetic field and bathymetry data, Endeavour Ridge segment, Juan de Fuca ridge, cruise A118–41 (1988, Tivey). Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/320796.

[37] Tivey MA. Raw Deep Tow magnetic field data, Endeavour Ridge segment, Juan de Fuca ridge, cruise A112–11 (1984, Tivey). Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/320853.

[38] Tivey MA. Processed Deep Tow magnetic field and bathymetry data, Endeavour Ridge segment, Juan de Fuca ridge, cruise A112–11 (1984, Tivey). Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/320858.

[39] Tivey MA. Raw Deep Tow magnetic field data, West Endeavour Ridge segment, Juan de Fuca ridge, cruise TT-196 (1986, Tivey). Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/320866.

[40] Tivey MA. Processed Deep Tow magnetic field and bathymetry data, West Endeavour Ridge segment, Juan de Fuca ridge, cruise TT-196 (1986, Tivey). Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/320870.

[41] Tivey MA. Raw Deep Tow magnetic field data, Northern Symmetrical (Cobb) Ridge, Juan de Fuca, cruise PGC89005 (1989, Tivey). Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/320876.

[42] Tivey MA. Processed Deep Tow magnetic field and bathymetry data, Northern Symmetrical (Cobb) Ridge segment, Juan de Fuca ridge, cruise PGC89005 (1989, Tivey). Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/320880.

[43] Tivey MA. Raw Deep Tow magnetic field data, Northern Endeavour Ridge, Juan de Fuca, cruise A125–10 (1990, Tivey). Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/320886.

[44] Tivey MA. Processed Deep Tow magnetic field and bathymetry data, Northern Endeavour Ridge segment, Juan de Fuca ridge, cruise A125–10 (1990, Tivey). Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/320891.

[45] Tivey MA. Raw Deep Tow magnetic field data, Cleft and Vance segments, Juan de Fuca ridge, VENTS cruise DI-91-01 (1991, Tivey). Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/320903.

[46] Tivey MA. Processed Deep Tow magnetic field and bathymetry data, Cleft and Vance segments, Juan de Fuca ridge, cruise DI-91-01 (1991, Tivey). Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/320910.

[47] Tivey MA. Raw Deep Tow magnetic field data, Pigafetta Basin, Western Pacific Ocean, cruise TUNE08WT (1992, Tivey). Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/320922.

[48] Tivey MA. Raw Deep Tow magnetic field data, Pigafetta Basin, Western Pacific Ocean, cruise TUNE08WT (1992, Tivey). Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/320924.

[49] Tivey MA. Processed Deep Tow magnetic field and bathymetry data, Pigafetta Basin, Western Pacific Ocean, cruise TUNE08WT (1992, Tivey). Integrated Earth Data Applications (IEDA) 2013. http://dx.doi.org/10.1594/IEDA/320928.

[50] Todd E, Gill JB, Wysoczanski RJ, Hergt J, Wright IC, Leybourne MI, Mortimer N. Hf isotopic evidence for small-scale heterogeneity in the mode of mantle wedge enrichment: Southern Havre Trough and South Fiji Basin back arcs. Geochem Geophys Geosyst 2011;12:Q09011. http://dx.doi.org/10.1029/2011GC003683.

[51] Wallis JC, Rolando E, Borgman CL. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. PLoS One 2013;8(7):e67332. http://dx.doi.org/10.1371/journal.pone.0067332.

[52] Woodhead JD, Hergt JM, Davidson JP, Eggins SM. Hafnium isotope evidence for 'conservative' element mobility during subduction processes. Earth Planet Sci Lett 2001;192(3):331–46. http://dx.doi.org/10.1016/S0012-821X(01)00453-8.

[53] Workman RK, Hart SR. Major and trace element composition of the depleted MORB mantle (DMM). Earth Planet Sci Lett 2005;231:53–72. http://dx.doi.org/10.1016/j.epsl.2004.12.005.