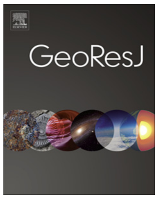
GeoResJ 6 (2015) 213–220

Contents lists available at ScienceDirect

GeoResJ

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



Preservation challenges for geological data at state geological surveys

Sarah Ramdeen

School of Information and Library Science, University of North Carolina at Chapel Hill, CB 3360 Manning Hall, Chapel Hill, NC 27599, USA

ARTICLE INFO

Article history:
Received 11 October 2014
Revised 28 April 2015
Accepted 28 April 2015
Available online 19 May 2015

Keywords:
Data rescue
State geological surveys
Data preservation
Core and sample libraries
Geological data repositories
Curation

ABSTRACT

State geological surveys are home to legacy geological data that holds value in the present. Early legislation of geological surveys often included requirements that state surveys have a museum or cabinet to house their physical collections. These collections currently include data such as cores, cuttings, thin sections and fossils. State geological surveys maintain these collections to support scientific research that has value to those in government, industry, academia and the public. Survey collections and other similar science data collections, are in danger of being lost due to various risks such as poor curation, few access points, lack of funding, and space considerations. Efforts to preserve these collections have increased, beginning with a National Research Council report in 2002 highlighting this plight, and the founding of the National Geological and Geophysical Data Preservation Program by the United States Geological Survey (USGS) in 2005. Currently, programs like EarthCube address this problem by focusing on cyberinfrastructure needs that will ease discovery and access to specimen datasets. Even with these efforts, there is still much work to be done.

Increasing preservation and ease of access requires training in data curation and preservation as well as a better understanding of the users of geological data. This paper will introduce geological collections, provide examples of preservation challenges surrounding these types of collections, and suggest future research directions. This includes collaborations with library and information scientists, archivists, museums curators, as well as cross training of domain scientists. Future management systems for these collections should provide increased discovery and access to geological data.

1. Introduction

In 2008, Western Michigan University's Michigan Geological Repository for Research and Education (MGRRE) and the state's geological repository, acquired 500,000 feet of rock core from the Mosaic company. The company owned a potash mine from which the cores were drilled. They no longer wanted to store them and offered the cores as a donation to the university [54]. Two administrators at the MGRRE, realizing the research value of these materials, drove their own vehicle to pick up the 4000 boxes of cores. "It took four pick-up loads to bring all the material down to Kalamazoo" [54]. These samples were later used to verify the quality of amount of potash (a mineral used in fertilizers) in a rediscovered mineral deposit in West Michigan [54]. This discovery is valued at $65 billion dollars and has a major impact on the local economy. It will lower the costs of farming in the Midwest where farmers must pay to import potash from Mexico, Canada, and Russia. A new mine will create construction jobs as well as full time jobs at the site [54]. These 'unwanted' samples have become a major resource for the state of Michigan, and it was fortuitous that MGRRE saw the value in them as legacy data and had the opportunity (and resources) to preserve them.

On January 17th, 2001, a natural gas explosion occurred in downtown Hutchinson, Kansas. Two local businesses burned down as a result. Two days later, another leak occurred under a mobile home, and two people were killed. As a safety precaution, the city was evacuated. Ultimately, residents were not able to return until March [33]. During the intervening months, KGas, the local gas company, collaborated with the Kansas Geological Survey to investigate the leaks. "Everyone involved in the crisis came to quickly value the geologic data and samples the Kansas Geological Survey had collected and archived for decades" [1, p. 14]. Among the materials used were a collection of cores drilled in the 1960s by the U.S. Atomic Energy Commission (AEC). The AEC was investigating the possibility of nuclear storage in Kansas [12]. The Kansas Geological Survey had maintained these legacy data as part of their repository. This reuse, use beyond their original purpose, helped the investigators better understand how the natural gas was leaking from a nearby underground storage facility ([12], p.16). As the NRC [34] summarized, "having immediate access to

---

E-mail address: ramdeen@email.unc.edu

http://dx.doi.org/10.1016/j.grj.2015.04.002

critical geoscience data and information played a crucial role in facilitating rapid response to a local crisis" [34, p. 1].

The examples above demonstrate the importance of geological collections, their continued maintenance, and their potential for reuse. Close examination of current practices can lead to more sustainable preservation and better access to these collections.

2. Geological collections

2.1. Geological data

In some subdomains of geology, physical specimens are key to research. Scientists gather data from these items, analyze these data and produce scientific outcomes. These physical objects become data once they have been used in research, along with their associated metadata and descriptions. This metadata and documentation is also used to enable discovery and access for reuse as well as to capture geological information. There is a transition from a rock being just a rock, to it now representing scientific knowledge with this connection to the documentation. If this connection is lost, the value as data becomes lost and the physical item just becomes a rock again.

Such physical geological data include items such as rock core and cuttings, thin sections and fossils, as illustrated in Fig. 1. Most physical geological materials, when properly maintained, can be stored for future access without the risk of major sample degradation. For example, a properly stored and curated core sample from a well drilled in 1907 can produce new knowledge today and in the future. The data that these samples hold can be reused, reanalyzed, potentially using previously unavailable technology, and contribute to studies beyond the scope of the original project. The materials in these collections may be (1) examples of earlier observations or results, (2) standards, kept for the base of future comparisons, (3) resources for research into geological issues, (4) collections of rare or valuable items, (5) resources used for education and training future geologists, and (6) proactively collected materials for future use [36].

In a recent White House memorandum, Holdren [25] states "scientific collections provide an essential base for developing scientific evidence and are an important resource for scientific research, education, and resource management. Scientific collections represent records of our past and investment in our future". It is important to maintain collections of scientific data not just for new research but to confirm previous work. As geologist and historian Jackson [28] explains, "a fundamental tenet in science is the need for viable checking and reproducibility of results. Re-analyses may not be undertaken for some time after the original research, but require preservation of the original material worked on in order to be of any value" (p. 423). Raw data, which may include physical samples, may be used to conduct reliability and validity checks on the work being produced. Heidorn [22] stresses the idea that science is based on theories and theories are created based on replicable data. If the data are inaccessible and the theory cannot be replicated, scientific results would be unsubstantiated. "The availability of the data behind experiments helps to insure scientific integrity by keeping the process open to external evaluation" [22, p. 286].

There are many ways to categorize data, some of which may not be mutually exclusive, e.g. big, small, dark, legacy, etc. Legacy data are part of what Heidorn [22] termed the 'long tail of science data'. Heidorn [22] suggests the long tail of science data represents smaller individual collections, which never get inventoried and live in drawers or closets. These may also be categorized as dark data collections [52]. Suggest that these types of collections are similar to those covered by the term small science. Small science includes specialized datasets collected by individual and small teams of scientists rather than large groups. These larger groups collect "big data from big science [which] are intended for sharing among big teams" [52, p. 3].

The examples in the introduction demonstrate the value of geological collections. However, long term management and storage has not always been factored into the data collection process. Differences in management might depend on the intended use, the focus of metadata, and other institutional variances. This may lead to valuable collections being abandoned or left deteriorating (see Fig. 2) at the end of a project. This is not due to neglect, or lack of care, but a lack of resources and focus. State geological surveys face a variety of preservation challenges in relation to their geological data collections. Many facilities would like to have full maintenance for their samples, however these organizations do not have the proper resources to do so or lack a standard procedure for curation. Resources includes staffing, funding, and space.

The NRC [34] provides a number of examples of potential loss of geological data collections. For example, in 2002, cores collected by the Tennessee Valley Authority and the Department of Energy were being stored outside, in the elements. The cores are from such important locations as the Clinch River Breeder Reactor site and the Oak Ridge Reservation. Exposure to air and humidity can cause boxes to decay, hand written labels to be lost, and for minerals to decay (see Fig. 2, bottom right for an example of pyrite to oxidizing). When minerals decay, they no longer represent what the rocks and minerals represented in situ. When metadata on boxes becomes unreadable, or when samples change, their scientific value may be lost or diminished.

Long tail data are important as they are "a breeding ground for new ideas and never before attempted science" [22, p. 282]. When they are inaccessible, these sets of data may be lost to the public beyond the finished publication. In his 2014 testimony, Gooding explains that state geological surveys get many of the items in their collections from donations [27]. These donations come from a wide range of individuals including scientists from "coal, oil and gas, mining, highway construction, and environmental investigations; construction projects; quarry operators; university research; and federal and state projects" ([27], p.3). Each has their own method of documentation, data collection, and curation. This can lead to complicated hybrid collections at the state survey level that, owing to their complicated curation schemes and lack of standardization, and may become lost.

Concerns for physical items also includes concern for their digital surrogates. In order to discover and access these geological collections, adequate metadata, records and other text based materials are needed. These may be found in paper records, but are increasingly being digitized or digital born. Without this documentation, various aspects of scientific information contained in physical geological materials may be lost.

2.2. Evolution from museums to libraries

The origins of museums and geological collections are closely linked. Geology uses analytical or comparative ways of 'knowing'; research in geology involves deconstructing strata "into elements, in order to make classifications, or to better understand (and regulate) technical processes" [40, p. 113]. Pickstone [40] calls it a 'museological science' because "geology and mineralogy [are] also, in part, sciences of collections" (p. 117). When geology was still developing as a scientific field in the 1800's, the role of the curator became very important in managing geological collections [30,49]. These managers were often expert geologists [30]. Museums recognized the need to have someone manage and curate a collection to not only prevent it from falling into disrepair but also because of