Rescuing Legacy Seismic Data FAIR'ly

Lorraine J. Hwang*[1], Tim Ahern[2], Cynthia J. Ebinger[3], William L. Ellsworth[4], Garrett G. Euler[5], Emile A. Okal[6], Paul G. Okubo[7], and William R. Walter[8]

The Earth's interconnected and dynamical systems operate on a spectrum of time scales from millions of years to fractions of a second. Although the evolution of the Earth and its deep interior are beyond the time span of human observations, understanding of many natural phenomena operating on human time scales has benefited from direct scientific observation. Continuous processes and those that are repeated over time shape the environment we live in. As we are faced with unprecedented changes to climate, understanding the deeper patterns and trends in natural systems through time has taken on new importance (Research Data Alliance, 2019). The call to reuse data is driven not only by economics but also by the recognition of their uniqueness (observations of natural systems are not repeatable) and scientific value in enhancing current understandings as well as potential new discoveries especially in the era of big data. These data are part of the historical record and our scientific heritage (American Geophysical Union, 2019) not only in explicitly recording earth observations but implicitly recording, and thus providing the evidence that addresses the manner in which science was conducted.

Recorded observations of ground motion began with the advent of the instrumental era of seismology in the late 1800s. Data began to be systematically collected in the early part of the twentieth century, evolving into today's system of regional and worldwide networks. The digital and Internet revolution in the late part of the twentieth and early twenty-first century ushered in the current era of continuously recorded, digital data. Federated repositories quickly followed, easing the barrier of access to high-quality, digital time series data. With these data, the scientific community has developed new techniques and continues to discover new phenomena that reveal themselves as seismic moment, enhancing our understanding of the relationships among diverse Earth systems. However, digital data only capture a fraction of the observations in the instrumental era (Ishii, 2018). A better picture of phenomena such as the earthquake cycle, magmatic systems, and the solid earth's response to climate change can be drawn by incorporating more observations from the historic record. The emerging use of machine learning in the geosciences shows future promise in finding signal and patterns—extracting previously unrecognized features in big data (Kong *et al.*, 2018). Analog seismograms are an untapped resource.

To be available to modern research techniques, the analog data must be accessible in digital form. These data are often kept in remote locations and not under archival condition leading to the physical deterioration and, hence, loss of some records. Adding to the sense of urgency is the lack of institutional will to maintain these resources and the retirement of a generation of scientists who used the data from these instruments. This knowledge represents significant institutional memory of how to properly interpret and analyze these data, as well as about the body of the resource itself that must be captured.

Preservation efforts are underway at a limited number of institutions worldwide. Digital preservation efforts of historical Italian and European seismograms through the Istituto Nazionale di Geofisica e Vulcanologia SISMOS and EuroSeismos (Ferrari and Pino, 2003) projects, Japan (Murotani *et al.*, 2020), and HRV (Ishii *et al.*, 2015) provide leadership in best practices in conservation, imaging, and vectorization (see Data and Resources for links to these data). These efforts, however, have proceeded with little community discussion on standards and how the collections can explicitly meet **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable (FAIR) data principles. (Wilkinson *et al.*, 2016). That these data be FAIR provides guidance for data management and stewardship in the modern digital ecosystem. Guidelines must also be supplemented with domain requirements and leverage existing standards and infrastructure. Seismology has a rich history of sharing and has established good data management practices in the federation of its natively digital data. However, it is generally acknowledged

1. University of California, Davis, Davis, California, U.S.A.; 2. IRIS DMC, Seattle, Washington, U.S.A.; 3. Department of Earth and Environmental Sciences, Tulane University, New Orleans, Louisiana, U.S.A.; 4. Department of Geophysics, Stanford University, Stanford, California, U.S.A.; 5. Los Alamos National Laboratory, Los Alamos, New Mexico, U.S.A.; 6. Department of Earth and Planetary Sciences, Northwestern University, Evanston, Illinois, U.S.A.; 7. U.S. Geological Survey, Hawaiian Volcano Observatory (retired), Hilo, Hawaii, U.S.A.; 8. Lawrence Livermore National Laboratory, Livermore, California, U.S.A.

*Corresponding author: ljhwang@ucdavis.edu

Cite this article as Hwang, L. J., T. Ahern, C. J. Ebinger, W. L. Ellsworth, G. G. Euler, E. A. Okal, P. G. Okubo, and W. R. Walter (2020). Rescuing Legacy Seismic Data FAIR'ly, *Seismol. Res. Lett.* **91**, 1339–1340, doi: 10.1785/0220200027.

that analog data, for example, paper, micro (fische, film, chip), and frequency modulation tapes represent a different challenge. Hence, organization and discovery are necessary processes toward making these data FAIR compliant.

In response to this pressing need, a workshop was held 18–19 September 2019 in Albuquerque, New Mexico, on *Securing Legacy Data to Enable Future Discoveries*. This international workshop addressed two broad needs of the community in discussing the framework for preservation of longitudinal seismic data and creating interdisciplinary connections to facilitate its management, use, and access. Participants identified eight community needs to advance these goals:

Analog Holdings Catalog. Create an inventory of analog seismic data holdings to identify current resources, connect potential users to resources, and aid in metadata discovery.

Publications Database. Create a database of research publications that use analog data as a resource to other researchers, inspire new studies, and to demonstrate the need to preserve these data.

Metadata Standards. Begin work on creating FAIR compliant metadata standards to enable federated discovery and access.

Pilot Project. Identify existing repositories to pilot federated data search and access utilizing proposed metadata standards, and retrieval of multiple data and metadata types.

Future Research. Identify strategies to enable future research through open source and standardization of both data and software and targeted campaigns with specific research objectives.

New Technologies. Identify enabling technologies to reduce human intervention in the end-to-end process of creating research-ready, time series data.

Other Communities. Attract a broader scientific community to apply seismological data in nontraditional research domains and communities with similar needs in preserving analog time series data.

Outreach. Create a larger community of users through outreach at all career levels.

The vision outlined previously is to create a community and the infrastructure necessary to enhance the preservation, access, and usage of analog seismic data. Knowledge of this resource is essential in describing current holdings and identifying the metadata necessary to find data and make them available to modern research techniques. Lowering the barriers to usage includes easing access and creating the tools necessary to transform the data to digital forms, whether as scanned images or a digital time series, in making them accessible to modern seismic analysis methods. Building a community of users includes inspiring early career researchers through an National Science Foundation Research Experiences for Undergraduates program, leading special sessions at professional meetings on research uses, volunteering to be an Editor of a special issue, or teaching workshops on tools and methods. Only good stewardship by the community will secure these primary observations for future generations and preserve our scientific heritage. We invite you to join the effort to safeguard this resource and make it FAIR for current and future generations of earth scientists.

Data and Resources

Information on obtaining historical seismograms from the EuroSeismos and SISMOS project can be found using the following link: http://sismos.ingv.it/. Seismograms from, and more information about, the HRV Seismogram Archival Project can be found at http://www.seismology.harvard.edu/HRV/archive.html. A list of scanned images of three stations in the Wakayama, Japan, microearthquake network can be accessed at http://wwweic.eri.u-tokyo.ac.jp/; additional seismograms archived at the Earthquake Research Institute, University of Tokyo, are in the process of being scanned. All websites were last accessed January 2020.

Acknowledgments

National Science Foundation (NSF-1917159) provided funding for the *2019 Securing Legacy Data to Enable Future Discoveries Workshop* and partial support for Lorraine J. Hwang.

References

American Geophysical Union (2019). *Supporting Data as a World Heritage*, https://www.agu.org/Share-and-Advocate/Share/Policymakers/Position-Statements/Position_Data (last accessed December 2019).

Ferrari, G., and N. A. Pino (2003). EUROSEISMOS 2002-2003 a project for saving and studying historical seismograms in the Euro-mediterranean area sic, *2003 EGS-AGU-EGU Joint Assembly*, 5274.

Ishii, M. (2018). Seismogram legacy data fundamental science goals, *Presentation to the National Academy of Sciences, Committee on Seismology and Geodynamics*, Washington, D.C, 5 June 2018.

Ishii, M., H. Ishii, B. Bernier, and E. Bulat (2015). Efforts to recover and digitize analog seismograms from Harvard-Adam Dziewonski observatory, *Seismol. Res. Lett.* **86,** 255–261.

Kong, Q., D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, and P. Gerstoft (2018). Machine learning in seismology: Turning data into insights, *Seismol. Res. Lett.* **90,** 3–14, doi: 10.1785/0220180259.

Murotani, S., K. Satake, H. Tsuruoka, H. Miyake, T. Sato, T. Hashimoto, and H. Kanamori (2020). A database of digitized and analog seismograms of historical earthquakes in Japan, *Seismol. Res. Lett.*, doi: 10.1785/0220190287.

Research Data Alliance (2019). *Data Rescue Interest Group*, https://www.rd-alliance.org/groups/data-rescue.html (last accessed December 2019).

Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship, *Scientif. Data* **3,** 160018, doi: 10.1038/sdata.2016.18.

Manuscript received 21 January 2020
Published online 11 March 2020