

Physical Samples, Digital Collections

Unmil P. Karadkar¹, Kerstin Lehnert², Chris Lenhardt³, Sarah Ramdeen⁴

¹School of Information, The University of Texas at Austin

²Lamond-Doherty Earth observatory, Columbia University

³Renaissance Computing Institute, University of North Carolina at Chapel Hill

⁴School of Information and Library Science, University of North Carolina at Chapel Hill

Abstract

Research in the earth and biological sciences depends on the availability of representative, often irreplaceable physical samples that have been collected at substantial cost and effort. The NSF-funded EarthCube iSampleS (Internet of Samples in the Earth Sciences) research coordination network aims to connect physical samples and sample collections with digital data infrastructures to revolutionize their utility in the support of science. The goal of this workshop is to cross-pollinate the expertise of a broad audience comprising of biologists, earth scientists, and data curators who work with physical samples and information scientists. Prior workshops, hosted in conjunction with JCDL 2016 and ASIS&T Annual Meeting 2016 attracted participation from researchers and practitioners in the USA, Australia, Germany, and the Netherlands. In conducting this workshop at the iConference we aim to strengthen connections between sample-related communities in Asia-Pacific and the USA.

Keywords: internet of samples in the earth sciences; iSampleS; Cyberinfrastructure; physical samples; research data management

Citation: Karadkar, U. P., Lehnert, K., Lenhardt, C., & Ramdeen, S. (2017). Physical Samples, Digital Collections. In *iConference 2017 Proceedings* (pp. 922-924). <https://doi.org/10.9776/17405>

Copyright: Copyright is held by the authors.

Contact: unmil@ischool.utexas.edu

1 Purpose and Intended Audience

Research in the many scientific disciplines depends on the availability of representative samples collected from our natural environment, extraterrestrial bodies, or those generated in experiments. In the earth sciences, these physical samples serve as fundamental references for generating new knowledge about the earth and the entire universe, contribute to a deeper understanding of the processes that created and shaped it, assess the availability of natural resources, and measure the risk of natural hazards. Many samples have been collected at great cost and with substantial difficulty, are rare or unique, and irreplaceable. The EarthCube Research Coordination Network (RCN) iSamples (Internet of Samples in the Earth Sciences) aims to advance the use of innovative cyberinfrastructure to connect physical samples and sample collections across the Earth Sciences with digital data cyberinfrastructures to revolutionize their utility in the support of science. The goal of this RCN is to dramatically improve the discovery, access, sharing, analysis, and curation of physical samples and the data generated by their study for the benefit of science and society as part of the EarthCube program. The RCN hosted its first workshop in Austin, TX, in 2015, its second workshop in Chapel Hill, NC in 2016, and is coordinating its efforts with other EarthCube RCNs, such as the Earth-Centered Communication for Cyberinfrastructure (EC3), as well as national and international efforts to enhance access to scientific collections in general (e.g. CODATA Task Group, Research Data Alliance, Scientific Collections International).

The proposed workshop will focus on issues related to designing a distributed data cyberinfrastructure required to make samples easily accessible, to ensure persistent access to relevant sample metadata, and to allow unambiguous linking of the physical objects to the digital data. This workshop is intended to attract a broad audience comprising domain scientists, data curators, as well as computer and

information scientists to facilitate knowledge sharing about the requirements of physical sample and collection management. Attendees will address the issues and challenges in the creation, development and maintenance of computer-mediated collection management systems, citation of specimens, persistent identifiers, system architectures, administration, user interfaces, requirements engineering, evaluation models, and policy implications for digital collections.

Our two past workshops, held in conjunction with the ACM/IEEE Joint Conference on Digital Libraries 2016 (Newark, NJ) and ASIS&T Annual Meeting 2016 (CopenHagen, Denmark) attracted participants from US-based, European, and Australian communities. We expect that the location of the iConference will attract participants from the Asia-Pacific, enabling us to learn from the scholars and practitioners in this part of the world.

2 Proposed Format

The workshop will balance the presentations from scholarly and practitioner efforts in developing digital data management infrastructures for physical sample collections. We envision that the workshop will include two panels—one consisting of practitioner presenters and another of scholars engaged in developing the cyberinfrastructure for supporting the earth sciences. The panelists will introduce their efforts and then engage in discussion of the issues involved with audience participation. The workshop will begin and end with keynote talks and include a poster session that will enable several participants to showcase their advances while engaging with a broad audience.

Like past workshops, we plan to advertise our workshop to the target communities via a variety of community mailing lists such as ASIST, DataCite, DataONE, SIGIR, ACM SIGCHI, IEEE-TPDL, American Geophysical Union, Earth Science Information Partners (ESIP), EarthCube, and Research Data Alliance (RDA). A small program committee consisting of Computer Scientists and Earth Scientists will judge the merit and relevance of the papers as well as ensure that the presentations cover a breadth of topics in order to engage an audience with diverse interests.

The organizers will invite submissions focusing on various aspects of this large problem space at the intersection of physical samples and advances in digital collection creation, management, and use. Broad areas include social and technical approaches; policies and workflows; as well as issues involved in the curation, publication, and citation of data sets. Topics include but are not limited to: physical sample collection curation, theoretical models and system architectures for collection management, evaluation of existing environments, social-technical perspectives on digital methods for sample management, policies and workflows, personal information management in physical sample collections, sample collection archiving and preservation.

3 Goals and Outcomes

The goal of this workshop is to attract an international, inter-disciplinary community of researchers, curators, and practitioners who are interested in studying the issues involved in the management of samples, sample collections, and sample-based data in the field, in the lab, in repositories, in computer-mediated data systems and scientific publications. The intention is both to assemble the existing community as well as invite those with emerging interests in this area. The location of the iConference will enable a secondary goal of bringing together Asia-Pacific and North American communities together to focus on the tremendous opportunities for research in this space and for fostering collaboration between these communities

Outcomes of the workshop will include more exposure to the opportunities in collaborative research between the information science and earth science communities as well as improved sharing of advances made by geographically distributed existing participants in this community.

Workshop presentation slides will be published on a public-facing website. We are currently exploring possibilities for publishing the accepted papers via the IEEE-TCDL bulletin. Following the

workshop, we will issue an open call for submitting significant reports related to the workshop topics via a special issue of the International Journal on Digital Libraries.

4 Relevance to the iConference

As the premier meeting of the 60+ strong iSchool community, the iConference brings together the worldwide community of information studies scholars and students. This workshop will present compelling information challenges faced by the earth science community in the realm of physical samples to this gathering of information scientists, enhancing our awareness of the information challenges faced in the conduct of earth science as well as the management of information infrastructure that is critical for research in areas such as climate science, biology, global water resources, to name a few.

5 Duration

We would like to host a full-day workshop. The conference page includes contradictory information about whether this can be accommodated. If a full-day workshop is not possible, we will be happy to host a half-day workshop and the proposed format in section 3 reflects a half-day schedule.

6 Attendance

Based on responses to our past workshop, we expect to attract 5 to 8 submissions for presentations and about 5 to 10 submissions for posters. In addition, given the nature of the iConference workshops, we will be happy to accommodate between 20 and 30 attendees in order to broaden the community as well as invite a lively conversation among those with emerging interests at the intersection of earth and information sciences.

7 Special Requirements

No unusual requirements. We will require a projector and an internet connection for participants who may demonstrate a live software. If possible, we will require poster boards and space to showcase posters.

8 Acknowledgements

The iSamplES Research Coordination Network is funded by the National Science Foundation through grant number: 1440351.