Physical Samples and Digital Libraries

Unmil P. Karadkar
School of Information
The University of Texas at Austin
1616 Guadalupe St. Ste. 5.202
Austin, TX 78701-1213 USA
+1-512-471-9292
unmil@ischool.utexas.edu

Kerstin Lehnert
Lamont-Doherty Earth Observatory
Columbia University
61 Route 9W - PO Box 1000,
Palisades, NY 10964-8000 USA
+1-845-365-8506
lehnert@ldeo.columbia.edu

Chris Lenhardt
Renaissance Computing Institute
University of North Carolina
at Chapel Hill
100 Europa Drive Suite 540
Chapel Hill, NC 27517 USA
+1-919-445-0480
clenhardt@renci.org

ABSTRACT

Research in disciplines such as the earth and biological sciences depends on the availability of representative physical samples that have been collected at substantial cost and effort and some are irreplaceable. The EarthCube iSamples (Internet of Samples in the Earth Sciences) Research Coordination Network (RCN), funded by the National Science Foundation, aims to connect physical samples and sample collections across the Earth Sciences with digital data infrastructures to revolutionize their utility in the support of science. The goal of this workshop is to attract a broad audience comprising of earth scientists and other scientists working with physical samples, data curators, and computer and information scientists to learn from each other about the requirements of physical as well as digital sample and collection management.

Keywords

Internet of Samples in the Earth Sciences; iSampleS, Cyberinfrastructure; EarthCube

1. INTRODUCTION

Research in the many scientific disciplines depends on the availability of representative samples collected from our natural environment and even 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 infrastructures 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, its second workshop in Chapel Hill, NC, 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, Scientific Collections International).

The proposed workshop will focus on designing a distributed data infrastructure 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 of domain scientists, data curators, and computer and information scientists to learn from each other about the requirements of physical and digital sample and collection management. Attendees will address the issues and challenges in the creation, development and maintenance of collection management systems, citation of specimens, persistent identifiers, system architectures, administration, user interfaces, requirements engineering, evaluation models, and policy implications for digital collections.

2. OBJECTIVES

The goal of this workshop is to attract an 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 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. A secondary goal is to focus the attention of the digital libraries community on the tremendous opportunities for research in this space and for collaborating with researchers in relevant science domains.

3. TOPICAL OUTLINE

The workshop will invite submissions on a broad range of topics at the intersection of physical samples in the sciences and digital libraries that facilitate the data collection, management, and use of samples. Topics include but are not limited to:

- physical sample collection curation
- information behavior and needs, user modeling
- evaluation of existing environments
- theoretical models
- system architectures
- social-technical perspectives on digital methods for sample management
- policies and workflows
- data analytics and visualization
- user interfaces
- cyberinfrastructure architectures, applications, and deployments
- distributed data management information systems
- impact of digital libraries on earth science sample collections

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

Copyright is held by the owner/author(s).

JCDL '16, June 19-23, 2016, Newark, NJ, USA

ACM 978-1-4503-4229-2/16/06.

<http://dx.doi.org/10.1145/2910896.2926736>

- linked data and its applications
- personal information management of physical samples
- retrieval and browsing of samples
- scientific data curation, citation and scholarly publication
- social networks, virtual organizations and networked information
- sample collection archiving and preservation

4. FORMAT – FULL DAY WORKSHOP

We intend to host to a typical JCDL workshop, which includes position papers, presentations of ongoing work or results, and planning activities for building a community of scholars focusing on the creation, management, and use of scientific sample collections. The workshop will allow for plenty of discussion time between presentations as well as informally, during breaks. The workshop will invite posters and demonstrations (2 pages in the ACM proceedings format), position papers as well as works-in-progress papers (4 pages) and full papers, reporting on mature research (6 pages in the ACM proceedings format).

5. AUDIENCE

The expected audience is researchers and practitioners who are active in the management of scientific collection data as well as who have an interest in working at the intersection of earth science data and digital libraries. Thus, this workshop aims to bring together communities from multiple disciplines to develop a shared understanding of the needs to further scholarship and practice in this cross-disciplinary area.

Our intention is to engage 20-30 attendees approximately balanced between domain scientists and Computer/Information Scientists. A previous workshop accepted at JCDL did not attract a quorum of participants. With this experience, we have broadened the theme of the workshop to include all physical samples (not just earth samples, as we did last year), reached out to other communities and organizations, especially those geographically proximal to the conference location, and plan to advertise our workshop to cognate communities such as DataONE, and Research Data Alliance. We have also invited Chris Lenhardt, an Environmental Data Science Domain scientist at the Renaissance Computing Institute to serve as a workshop co-chair. Historically, the JCDL community has had a small but committed group of scholars working in the Earth Science domain, such as the DLESE project staff. We expect that the additional time afforded this year by the earlier JCDL notification will be beneficial for attracting participants.

6. IDENTIFICATION AND SELECTION

In order to attract attendees from a broad background, we will advertise this workshop to the geoscience and information science communities. Several SIG bulletin boards and mailing lists such as ACM SIGCHI, ASIS&T, TPD, American Geophysical Union, EarthCube, Research Data Alliance, and DATAONE. We will also reach out to groups such as DLESE that have focused on the Earth Sciences within the digital libraries discipline. In addition, we invite the JCDL 2016 organizing committee to suggest other

venues that will broaden dissemination of our workshop announcement.

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.

Preliminary schedule:

May 15: Paper submissions due

May 22: Notify authors of acceptance

June 22, 23: Workshop at JCDL

The dates above may be tweaked to align with JCDL early registration deadline.

7. ORGANIZER INFORMATION

Unmil Karadkar (unmil@ischool.utexas.edu) is an Assistant Professor in the School of Information at The University of Texas at Austin. He situates his work at the intersection of digital libraries, human-computer interaction, and visualization. He studies data practices of researchers with an eye toward identifying unmet information needs. Based on an understanding of these needs, he designs software to support their evolving practices and evaluates the impact of this software on their work.

Kerstin Lehnert (lehnert@ldeo.columbia.edu) is a Senior Research Scientist at Columbia University's Lamont-Doherty Earth Observatory. She is the principal investigator of the NSF-funded EarthCube Research Coordination Network project: Earthcube RCN: iSampleS: The Internet of Samples in the Earth Sciences, and has a long-term record designing data systems and digital infrastructure for samples and collections. She is Director of the Interdisciplinary Earth Data Alliance (IEDA), an NSF-funded data facility, chair of the EarthCube Leadership Council and president of the Earth & Space Science Informatics Focus Group of the American Geophysical Union

Chris Lenhardt (clenhardt@renci.org) is a Domain Scientist in Environmental Data Science and Systems at Renaissance Computing Institute, NC. He has worked on the management of digital scientific data throughout his career. Prior to coming to RENC, he was the manager of NASA's Distributed Active Archive Center (DAAC) for Biogeochemical Data at Oak Ridge National Laboratory. His current research interests include the iSamples effort, agile data curation, and sustainable scientific software. He is part of the DataNet Federation Consortium (DFC) team at UNC. Mr. Lenhardt was part of the original SESAR (System for Earth Sample Registration) team. He also served as the president of the ESIP (Federation of Earth Science Information Partners) and is the Chair of NOAA's Data Access and Archiving Requirements Working Group (DAARWG).

8. ACKNOWLEDGEMENTS

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