

iSAMPLES AND SAMPLING NATURE: MAXIMIZING THE VALUE OF TODAY'S SAMPLES FOR TOMORROW'S SCIENCE. R. L. Walls¹, J. Anderson², H. Cui³, N. Davies⁴, J. Deck⁵, S. Kansa⁶, C. Meyer⁷, S. Ramdeen⁸, S. Richard⁹, R. Snyder⁷, D. Vieglais¹⁰, and K. Lehnert⁸, ¹Bio5 Institute, University of Arizona, rwalls@email.arizona.edu, ²New York University, ³University of Arizona, ⁴Gump Pacific Research Station, University of California, Berkeley, ⁵Berkeley Natural History Museums, University of California, Berkeley, ⁶Open Context, ⁷National Museum of Natural History, Smithsonian Institution, ⁸Lamont Doherty Earth Observatory, Columbia University, ⁹USGIN Foundation Inc., ¹⁰University of Kansas

Introduction: Material samples are the most basic form of data about the natural environment. Their collection and study underpin knowledge that is fundamental for both basic science and for building a resilient society that meets sustainable development goals locally and globally. Unfortunately, harnessing today's samples and sample-based data for tomorrow's scientific research and societal applications is severely limited, largely because samples and associated data are often difficult to find, access, integrate, and reuse. Samples and their associated data are rarely FAIR [1] and are often siloed within the discipline or project for which they were collected. This presentation describes two related projects that aim to enhance the natural science value chain encompassing sampling events, processing of samples for scientific analyses in multiple laboratories, and storage of the resulting materials and data in diverse and distributed collections, databases, and repositories. Both the Sampling Nature Research Coordination Network (RCN) and the Internet of Samples (iSamples) are built on the premise that core elements of sample data management are discipline agnostic and that sample-related data can be valuable beyond the purpose for which they were collected.

Sampling Nature: The Sampling Nature RCN aims to accelerate sample-based scientific research involving diverse stakeholders across broad societal benefit areas including health, energy, and sustainability. The RCN brings together researchers from multiple disciplines through a series of workshops and outreach activities, as well as a website and discussion forum. Four working groups focusing on standards, policy, training materials, and long-tail research collections will contribute to RCN deliverables. By supporting the development, promotion, and implementation of community-driven metadata standards and data policies across domains, Sampling Nature RCN tackles provenance, compliance, and auditing challenges that limit the reuse of material samples. By engaging multiple stakeholders, including local and Indigenous communities, government agencies, and industry, Sampling Nature RCN seeks to maximize broader impacts of material samples and to disseminate best practices in sample and data stewardship.

iSamples: To coordinate community-building efforts with technical development, Sampling Nature is

working closely with the iSamples project [2]. iSamples is a standards-based collaboration to uniquely, consistently, and conveniently identify material samples, record core metadata about them, and link them to other samples, data, and research products such as datasets and published literature. iSamples extends existing infrastructure and best practices in data stewardship to render a cross-domain cyberinfrastructure that enables transdisciplinary research, discovery, and reuse of material samples in 21st century natural science. iSamples has two primary components: iSamples-Central is a central discovery and resolution service (search interface on the web and API) for any community that wishes to participate; iSamples-in-a-Box delivers distributed infrastructure early in the data production chain with an emphasis on the needs of specific research domains. An iSamples-in-a-Box instance is a stand-alone system that enables creation of identifiers and associated metadata, retrieval of the sample information, updates to the sample metadata, sample identifier resolution, and discovery of samples. iSamples-Central is designed as a permanent Internet service that preserves and indexes sample metadata to ensure reliable discovery and retrieval. It provides a gateway between iSamples-in-a-Box instances and identifier authorities to ensure that remote iSamples-in-a-Box content is fully synchronized with the relevant authorities. iSamples infrastructure is identifier agnostic, although initial use cases focus on globally unique, persistent, and resolvable, identifiers including International Geo-Sample Numbers (IGSN) [4] and Archival Resource Keys (ARK) [5]

Conclusion: iSamples and Sampling Nature already have substantial buy-in from the life sciences, Earth-Sciences, and archaeo- and anthropological research communities. Extension to extraterrestrial samples is a natural fit that could benefit the Astromaterials Data Management community.

Acknowledgments: U.S. National Science Foundation Award numbers: Sampling Nature: [2129268](#), iSamples: [2004839](#), [2004562](#), [2004642](#), and [2004815](#).

References: [1] Wilkinson M., Dumontier M., Aalbersberg I. *et al.* *Sci Data* 3, 160018. [2] Davies N., Deck J., Kansa E. C., *et al.* (2021) *GigaScience* 10, giab028. [3] <https://www.igsn.org/> [4] <https://arks.org/>