Integrating Physical Samples into a Scientific Internet of Things

Kerstin Lehnert, Columbia University

The 'Internet of Things' is a term that refers to “uniquely identifiable objects (things) and their virtual representations in an Internet-like structure” (Wikipedia). Samples collected from our natural environment are largely unconnected from digital data infrastructures because they are lacking a uniform virtual representation (digital metadata profile) that can be accessed via a persistent and resolvable unique identifier. This hinders the discovery, access, and citation of the samples on the internet.

New internet-based capabilities have been developed over the past few years for the registration and unique identification of samples that make it possible to overcome these problems. For Earth Science samples, services for the registration and unique identification of samples and the long-term preservation of sample metadata are provided by the System for Earth Sample Registration SESAR (www.geosamples.org). SESAR developed the International Geo Sample Number, or IGSN, as a unique identifier for samples and specimens collected from our natural environment. Use of the IGSN now allows to establish links between the samples (i.e. their virtual representation), data acquired on these samples, and the publications that report these data. Samples can be linked to a dataset by including IGSNs in the metadata of a dataset DOI when the dataset is registered with the DOI® system for unique identification. Links between IGSNs, dataset DOIs, and publication DOIs will in the future allow to search for and find data acquired on a specific sample across the entire literature.

Keywords: data discovery and access, unique identifiers, sample collections, Earth Science samples, citation of objects