1. Sampling is fundamental to Science

The process of sampling, observing and analyzing physical samples is not unique to the geosciences. Physical sampling (taking specimens) is a fundamental strategy in many natural sciences, typically to support ex-situ observations in laboratories with the goal of characterizing real-world entities or populations. This is often a highly skilled specialization in its own right that is consistently under-documented and accredited within the research ecosystem.

2. Samples to Data to Publication

Observations and measurements are made on individual samples. Tracking samples and subsamples through diverse lab environments can become a non-trivial task. Combining all resultant measurements can be very challenging.

There may then be a sequence of publications based on research on an individual sample, undertaken by multiple, potentially unrelated research programs over many years. It can often be a laborious, near impossible undertaking to reliably collate and de-dupe multiple studies in order to perform systematic reviews or meta-analyses.

3. Uniquely Identifying our Samples

Unique, persistent web-compatible identifiers for physical objects and related sampling features are required to ensure their unambiguous identification and connection to related datasets through web identifiers. Identifier systems have been established within specific domains (e.g., bio, geo, hydro) or different sectors (e.g., museums, government agencies, universities). Cross referencing between a sample identifier and other identifiers for related research objects (papers, datasets) is also essential.

4. IGSN: a unique Identifier System in the Geosciences

The International Geo Sample Number (IGSN) in the geosciences, has already been used for rock, fossil, mineral, soil, regolith, fluid, plant and synthetic materials. Registration of IGSNs in the geosciences is governed through a central system that ensures they are globally unique. The handle.net global resolver system resolves each registered IGSN to a landing page, which contains a more detailed sample description. The IGSN description metadata schema includes elements that allow linking between the registered sample and other related resources (e.g., documents, physical samples/collections and data sets). The core IGSN metadata has been aligned with DataCite and IGSN has been included as a relatedIdentifierType option in the DataCite Metadata Schema 4.0. See www.igsn.org for more information.

5. Linking Identifier Systems

To enable the unique identification of all samples on Planet Earth and of data derived from them, the next step is to ensure IGSNs can either be integrated with comparable alternate sample identifier systems in other domains/sectors, or introduced into those domains that do not have available or sustainable or governed systems.

6. The Internet of Samples for Planet Earth

A registry of persistent identifier systems for all physical samples would allow users to choose which system best suits their needs. Such a registry may facilitate unifying best practice to enable consistent referencing of physical samples and of the methods used to link digital data to its sources. IGSNs could be extended to other domains, but additional methodologies of sample collection, curation and processing may need to be considered.