FAIR Samples for Open Astromaterials Science. S. Ramdeen[1] and K.A. Lehnert, Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA. [1] sramdeen@ldeo.columbia.edu

Introduction: Material samples, collected and curated with substantial investments of public funds, are critical resources for research in the Earth & Planetary Sciences. Today's Open Science paradigm requires these samples to be Findable, Accessible, Interoperable, and Reusable (FAIR) [1] so that they remain valuable for future research. The IGSN is a sample identifier that enables FAIR samples, addressing the two fundamental requirements: 1. Persistent and globally unique identification of objects. 2. Consistent and machine-actionable metadata describing the object.

SESAR, the System for Earth Sample Registration, provides services for the registration of samples with IGSN. Since 2016, SESAR curators have worked with staff from the Johnson Space Center to map SESAR sample metadata fields to the Lunar Curation Database in order to facilitate the registration of NASA's Apollo Lunar samples [2] and to explore registration of meteorite specimens with IGSN. In this presentation we will review the value of IGSN and how it may be used for the astromaterials captured during sample return missions, and the outcomes of our lunar sample metadata mapping.

IGSN: The IGSN is a persistent and globally unique identifier, designed specifically for material samples. It uses the handle system to link ("resolve") to a web page ("landing page") with the information about the sample. The IGSN is operated and governed by an international non-profit organization, the IGSN e.V., that ensures the uniqueness of the IGSN. Its members act as agents, who register IGSNs on behalf of their users at the central IGSN registry.

Benefits of the IGSN

- The IGSN makes it possible to track a sample from 'birth to cradle' (collection to archive) and on its path through many analytical labs, and allows linking to subsamples at any level of hierarchy.
- The IGSN allows one to unambiguously cite a sample, enabling previously impossible linking of samples to data and publications, and the gathering of metrics about sample use to give credit to curators and repositories.
- Use of the IGSN makes it possible to link and integrate sample-based observations across data systems, and paves the road toward advanced data mining of sample-based data.

Current members of the IGSN e.V. include national geological surveys and organizations and universities in the US, Europe, Australia, Africa and Asia. As of September 2021, 9.8 million samples have been registered with IGSN. Current users include: the Smithsonian Institution, the US Department of Energy, US core repositories, and the International Continental Drilling Program. The IGSN is recommended by publishers as a Best Practice for sample citation. In September 2021, the IGSN entered a partnership with DataCite [3] to ensure sustainable and reliable services for a growing, multi-disciplinary membership.

Relevance to Sample Return Missions

Astromaterial samples will be some of the most expensive samples to ever be collected and will be considered irreplaceable. It will be important to track actions taken on these samples such as subsampling, analytical measurements, and publication of data to preserve the provenance. IGSN is designed to provide such support. For example, capturing relationships between a sample and subsample (parent-child), documenting samples which have been exhausted (destructive sampling), and as mentioned providing a connection between analytical and research papers which relate to the sample by providing a unique, persistent, resolvable identifier which be used regardless of where or how the research is conducted.

Lunar Sample Mapping to SESAR metadata requirements: In 2016, curators and staff from NASA and SESAR reviewed the SESAR metadata fields for use in registering Lunar samples with IGSN [2]. The main findings from this effort were that controlled vocabularies in SESAR would need to be expanded to support the unique needs of astromaterials. This includes adding Moon/planetary object types, lunar rock classifications, space related sample collection methods, and expanding age ranges to include billions of years (Ga). Changes are also necessary in SESAR's location metadata fields to support non Earth coordinate systems and datums. In order to make these adjustments, community input is required. SESAR will need a list of community approved terms for the fields which need modification.

Conclusions: The IGSN is a well-established, globally used identifier system for samples that can be applied to astromaterials specimens, both previously collected and those being captured as part of the sample return missions.

References:

[1] Wilkinson, M. Dumontier, M. Aalbersberg, I. et al. (2016) *Sci Data 3*, 160018. https://doi.org/10.1038/sdata.2016.18. [2] Todd N.S. (1996) *AGU Fall Meeting Abstracts, 2016*, Abstract# PA51B-2270. [3] Buys M. and Lehnert K. A. (2021) DataCite Blog, https://doi.org/10.5438/thhf-kx17.