ID Assignment Checklist Data Type Example 3

Data Type: Field Collection dataset (used as the basis for subsequent and comparative analyses); e.g., physical soil samples from which physical, chemical & biological analyses will be generated.

Use Case: You're involved in a cross- disciplinary environmental research project that will build upon and integrate data from different domains to better understand a broader ecosystem. You want to reduce the time devoted to integrating the various data types into an efficient workflow for the project. You do have the right to register identifiers for this data.

Ask the research support staff at your institution what are the standard ID schemes used by your community or data domain.	International GeoSample Number (IGSN) (most commonly used for physical sample data) ***– the focus of this checklist example.*** Another ID often used is Archival Resource Key (ARK)
Check to make sure that the ID schemes used by the community are flexible enough to use for the types of datasets you and your collaborators produce.	Sample types that can be registered at SESAR and assigned IGSNs include: Biology, Ice, Liquid (aqueous & organic), Mineral, Organic Material, Particulate, Plant Structure, Rock, Sediment, Soil, Synthetic, Tephra (as of August 2021).
Check data similar to yours to see whether the ID schemes have been successfully applied to the use cases that you anticipate for your data. For example, can the IDs be used At varying levels of granularity?	In this use case, i.e., for the data from a physical sample, IGSNs can be assigned at the parent (or object) level and the child (or sub-object) level. Example Parent IGSN: IENTLCOLD ECA Coldstream 1 MH Example Siblings of this parent IGSN: o IENTL00UX Coldstream C1 B1 o IENTL00UY Coldstream C1 B2 o IENTL00UZ Coldstream C1 B3 , etc. Extensions to IGSNs are being proposed which would allow IGSNs to include information about:

Or extended to reflect the relationships among disparate data entities?	Collection to which a parent/child sample belongs Geolocation(s) info for the location from which samples were collected Event(s) when collection/parent/child samples were collected Method(s) used to collect parent/child samples
Ask if the infrastructure and services of the allocation agent that provides and registers the IDs has sustained organizational & community support. For ex., see below:	Allocating Agent quite commonly used in North America for those researchers unaffiliated with another allocation agent: System for Earth Sample Registration (SESAR) http://www.geosamples.org For other allocating agents around the globe, see: https://igsn.github.io/agents/
Do the data repositories or archives where you plan to store your data for the longer term support these IDs?	For this use case, our assumption is that the data repository or archive to be considered would be the one where you plan to store the analytical data you derive from the sample dataset. Repositories /archives to consider that accept IGSNs might be: Various Earth Science data archives such as ESS-DIVE Various state geological surveys, college & university repositories See SESAR full list of archives SESAR at: https://app.geosamples.org/reference/advanced_list.php?picklist=1&srv=archive
➢ Do the publishers that you intend to use accept these IDs?	"Publishers can link IGSNs to their sample metadata profile pages so readers can see all sample metadata. (AGU, GSA, Wiley, Elsevier, Copernicus) " from SESAR website at: https://www.geosamples.org/resources/researchers
➢ Does the ID registration agent (e.g. SESAR) provide training or help with the description (metadata) requirements or other technical questions?	SESAR minimal metadata requirements for pre-sample registration & printing of sample labels are noted (Object type , Sample name); SESAR recommendations for full sample registration metadata elements are intended to enable search & organization capabilities (Classification, Collection Method and Geolocation (Lat/Lon), Current Archive, Current Archive Contact, and Collecting scientists). For more information, see: https://www.geosamples.org/resources/help#1605908227292-5-5

	See also SESAR Metadata and Vocabulary Guides at: https://www.geosamples.org/resources/help/#vocabularies
Pull together the important information required by the ID scheme to describe & make your data more FAIR (Findable… Accessible… Interoperable…Reusable) ➢ What's important for finding your data in a catalog or archive and making it accessible over time?	Descriptive information strongly recommended, if not required to make a physical sample more readily searchable in a physical sample catalog include[1]: ➢ IGSN and Parent IGSN (where applicable) ➢ Sample Name (must be unique and project-specific) ➢ Chief Scientist / Collector ➢ Sample Type Field including: ○ Object type(e.g., Individual sample, core, site) ○ Material (e.g., Liquid-aqueous, rock, soil, biology) ○ Sampled feature (primary physiographic feature sample collected from) ➢ Location Information ➢ Date ➢ Collection Method Description ➢ Project
➢ Include the contextual information most important for others to reuse your data & more easily interoperable	Other descriptive information that could help facilitate the reuse of physical sample data will vary depending upon the sample type. SESAR provides metadata templates and guides in their SESAR Batch Registration Quick Guide.

[1] See DOI: http://doi.org/10.5334/dsj-2021-011, pg. 3 for more information on this recommendation.