ID Assignment Checklist Data Type Example 1

Data Type: Static dataset (no longer being updated, but still associated with published research results)

Use Case: You are ready to publish your findings and need to get an ID for your CSV file or for a table generated from your dataset that provides a visualization of your results so that the data can be associated with your paper. 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. (Journals and professional societies may have requirements or guidelines.)	The most commonly used ID schemes for data that are intended for publication are: ➢ Digital Object Identifiers (DOI) ➢ Archival Resource Keys (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.	The ID schemes for these kinds of data often use the DataCite Metadata Schema which requires both a DOI identifier and a "resourceTypeGeneral" that could be, for example: "Dataset," "DataPaper," "Image," or "Computational Notebook." See *DataCite-Metadata Kernel*, v4.4, pg. 48 for definitions of each type. Other schemas commonly used include Dublin Core and Dublin Kernel.
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?	For this use case, we assume there is no need to assign an identifier and describe anything other than a single-level data object.

➢ Or used to reflect relationships among disparate data entities?	If it would be useful to provide identifying information about the static dataset that generated the table or other visualization, however, the ID schemas mentioned above include means to do so. The DataCite schema used for creating DOIs, for instance, uses a "RelatedItem" property which would allow for the DOI of the static dataset [*DataCite-Metadata Kernel*, v4.4, pg. 9]. The ARK ID scheme allows for both subsetting of a data object and versioning. See ARK Identifiers FAQ.
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:	DOIs are probably the most commonly used ID scheme for this kind of data type and use case, and are accepted by most data repositories and archives. ARKs are also used quite widely for a number of types of data objects ranging from scientific records to datasets to archeological artifacts.
➢ Do the data repositories or archives where you plan to store your data for the longer-term support these IDs?	You can find both general and domain-specific data repositories and archives to find out which identifiers are accepted by going to: ➢ Scientific data's Data Repository Guidance (from nature.com) ➢ DataCite's Repository Finder ➢ FAIRsharing registry
➢ Do the publishers that you intend to use accept these IDs?	Most publishers accept DOIs; from the DOI Factsheet, DOIs are "currently used by well over 5,000 assigners, e.g., publishers, science data centres, movie studios, etc." See https://www.doi.org/factsheets/DOIKeyFacts.html As of September 2020, an estimated 8.2 billion ARKs have been created by more than 725 registration agents. They are used as permalinks (and identification) for (among other publishing entities): ➢ the Data Citation index (linked to the Web of Science)

	➢ Wikipedia articles ➢ Wikidata records ➢ Internet Archive collections ➢ ORCID researcher profiles ➢ See an interactive map of registration organizations for ARKs at ARK Identifiers FAQ
➢ Does the ID registration agent provide training or help with the description (metadata) requirements or other technical questions?	Training / help in meeting the metadata requirements for DOIs and ARKs can often be found by checking with the research support staff at your institution, from DataCite and CrossRef (commonly used registration agents for DOIs in North America) documentation, ARK Alliance Resources, and tutorials and other learning resources from training catalogues such as ESIP's Data Management Training Clearinghouse.
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?	When an application is made to a registration agency to create a DOI or an ARK (e.g., by your library), you'll need to provide the following information (See more complete definitions from the DataCite Metadata Scheme, pgs. 12 – 17): ➢ **Creator**: Your name or name of the primary creator of the data object ➢ **Title**: What it's called ➢ **Publisher**: Organization making the data publicly available ➢ **Publication Year** ➢ **Resource Type** (general description of the data type using a controlled vocabulary) Although expressed in XML, you can see examples of what kind of information should be included for a dataset; a polygon; a dataset with GeoLocation; and a data paper to name a few.

➢ Include the contextual information most important for others to reuse your data & more easily interoperable	Other information that may make your data more reusable and interoperable would be: ➢ **Language** in which the text/documentation is expressed ➢ **Size and format** of the file(s) (if they are intended to be downloaded) ➢ **GeoLocation** information is important to understand the context of visualization or for reuse ➢ **Rights/license for reuse** (This information may be dictated by your institution, repository or publisher.)