

DATA MANAGEMENT PLAN

During FY 2017, the Sustainable Intensification Innovation Lab established the [SIIL Dataverse](#) to store and curate all SIIL subaward datasets and serve as a data repository and access hub for the Sustainable Intensification community. All datasets collected using SIIL funds must be handled as outlined in this Data Management Plan. Within the SIIL Dataverse, a unique Dataverse has been created for each consortium and subaward. “[Dataverse](#) is an open source web application to share, preserve, cite, explore, and analyze research data. It facilitates making data available to others, and allows you to replicate others' work more easily. Researchers, data authors, publishers, data distributors, and affiliated institutions all receive academic credit and web visibility.”

The SIIL requires each subaward and consortium PI to report all planned and in-progress datasets in the Data Management module on the [SIIL Reporting Hub](#). However, only complete datasets are required for upload to the SIIL Dataverse. In addition to the dataset, the data package must also include metadata and other supporting documents including codebooks that document columns in the dataset, questionnaires used to collect the data, informed consent forms, reports, and any other contextual documents.

Datasets must be machine readable. This includes structured data in formats such as Excel (XML), JSON, KML, CSV, and SHP. Data that is not machine readable such as unstructured multi-media, photos, videos, emails, and narrative reports (e.g. Word, PDF) should be submitted to the [Development Experience Clearinghouse \(DEC\)](#). Datasets uploaded to the SIIL Dataverse can be given an embargo period, delaying the public release of the dataset. Though there is no official requirement for the length of the embargo period, USAID strongly recommends the embargo period lasts no longer than 12 months.

Each data package uploaded to the SIIL Dataverse will be given a DOI, which the PI will report in the Reporting Hub Data Management module for the corresponding dataset. Complete datasets that have been uploaded on the SIIL Dataverse will also be required to be registered by the PI to [the Data Development Library \(DDL\)](#) in a timely manner - within 30 days after it has been used to produce an intellectual work or is of sufficient quality to produce an intellectual work. The SIIL will track the submission of datasets by PIs (Principal Investigator) to the DDL by requesting a pipeline report from the DDL Data Stewards (i.e. abrenes@usaid.gov or opendata@usaid.gov).

Each complete dataset will be required to, at a minimum, include the following as part of the Dataverse dataset: codebooks; metadata; data dictionaries; forms, templates, and data gathering tools; explanations of redactions, when applicable (e.g. anonymization, removal/redaction/masking of personally identifiable information); notes on data quality, data limitations, or data context; and data gathering methodologies, dates, points of contact, geolocation(s).

Data packages that include Personally Identifiable Information (PII), as defined by the USG in the [OMB Memorandum M-07-1616](#), is required to include a consent form signed by the individual – including a copy of the template in the data package and all PII must be sufficiently masked so that direct (e.g. name, national ID number, personal address, telephone number, email, biometric data) or indirect (e.g. geographic data, birthday, race, gender, education levels) identification of the individual is prevented (or so difficult that it would be nearly impossible to us to identify an individual).