Feed the Future Innovation Lab for Peanut (Peanut Innovation Lab)

Data Management Plan
Version: 20180316

Peanut Innovation Lab Management Entity
The University of Georgia, Athens, Georgia

Feed the Future Innovation Lab for Peanut (Peanut Innovation Lab) Data Management Plan
Version: 20180316

The Feed the Future Innovation Lab for Peanut (Peanut Innovation Lab) adheres to the USAID's Open Data Policy and ADS Chapter 579 on USAID Development Data, and has established the following Data Management Plan (DMP) that applies to all research and publications supported by the program. Specific details regarding the data sets to be produced made available following this policy are listed in the Peanut Innovation Lab Data Management Table that is updated periodically.

While all efforts will be made to adhere to the Peanut Innovation Lab DMP, exceptions may be necessary in order to adhere to target country national and institutional requirements and restrictions. These will be duly noted in the Data Management Table.

Data Description

The research projects supported by the Peanut Innovation Lab will generate significant amounts of varied quantitative and qualitative data across a range of locations and countries, acquired via various means, and stored in various formats. The types of data will include experimental laboratory and field results (especially phenotypic evaluations of crop performance), genotypic data (e.g., SSR and SNP profiles), gene and genome sequences, household survey data, and clinical trial data. In addition, the program will generate databases, spreadsheets, develop and refine software, produce training materials, acquire digital images, videos and audio files, develop software models and produce project performance information. Emphasis is being placed to acquire all data electronically, and ultimately, all data will be stored in digital form, either in its original electronic format, or if necessary, converted into digital form by manual entry/scanning. As Peanut Innovation Lab is focused on peanut improvement, efforts are being made to use standard ontologies and nomenclatures for all phenotypic and genotypic data collection and storage. This will take advantage of existing efforts in crop ontology by the CGIAR Research Programs and the Crop Ontology community. As much as possible, critical metadata will be collected with all data, especially geo-references.

Intellectual Work

Includes all work products that document the implementation, monitoring and evaluation and results of USAID under this award. This includes, but is not limited to, program and communications materials, evaluations and assessments, information products, research and technical reports, progress and performance reports, other reports, articles and papers. It does not include financial or administrative information.

Peanut Innovation Lab Data Types

Phenotypic

- Varieties: on-station, on-farm performance (pest and disease ratings, maturity, yield, quality components)
- Populations (diversity, genetic): on-station, on-farm performance (pest and disease ratings, maturity, yield, quality components)

Genotypic

- Populations (diversity): SSR, SNP, gene/genome sequence
- Populations (genetic): SSR, SNP, gene/genome sequence, mapping genotypes

Survey

- Socio-economic (household, village, region, country)
- Mycotoxin levels
- Soil fertility
- Peanut production, processing, buying/selling, export

Weather

- Rain, temperature, solar (on-station, on-farm)

Dataset Description

A dataset is an organized collection of research data that is required to be publicly shared by USAID via loading in the Data Development Library (DDL). Supporting documentation for this dataset must also be loaded including, but not limited to, code books, data dictionaries, data gathering tools, notes on data quality, and explanations of redactions, surveys and research data.

In the event that this research data is being shared in a different publicly accessible research repository, the location of the repository and dataset may be loaded into the DDL instead.

The dataset is created by the project scientist and provided to the Peanut Innovation Lab Management Entity that will monitor and load the data in the DDL.

Data Organization and Management

All data produced by Peanut Innovation Lab research activities will be systematically managed, tracked, stored and published (as open source and/or in formal publications). Overall data management and curation will be the responsibility of the Peanut Innovation Lab Director and Assistant Director.

Peanut Innovation Lab is working with the Integrated Plant Breeding Platform (www.integratedbreeding.net) in the implementation of the Breeding Management System (BMS) for use in breeding programs and to handle breeding related data, with the Peanut Genome Initiative in developing PeanutBase (www.peanutbase.org) as a repository for genetic and genomic data, and with the CGIAR Excellence in Breeding initiative (excellenceinbreeding.org). Such integrated systems provide options for a central repository of all phenotypic and genotypic data generated by the various breeding programs supported by Peanut Innovation Lab.

During the active program period, Peanut Innovation Lab data will either be made available via one or more of the above data sites, the Peanut Innovation Lab website (ftfpeanutlab.caes.uga.edu) and/or other open-source data repository systems (e.g., Harvard University's Dataverse Network, www.thedata.org). However, all data will also be uploaded, if necessary, to a permanent, publically available site that meets the US Open Data Requirements. Specifications regarding the appropriate system(s) will be indicated in the Peanut Innovation Lab Data Management Table.

While the Peanut Innovation Lab website will provide global public access to data and information, the Peanut Innovation Lab will use project management software (e.g., Piestar DPx) to manage project tasks, data, information and reports. Access will be restricted to Peanut Innovation Lab staff and scientists, although ultimately, all information will be transferred to the public accessible site(s).

Prior to collecting any data, agreement will be reached by the Peanut Innovation Lab teams on standardized collection methodology and nomenclatures. The agreed methodologies will be included in the annual work plans for the projects. Results from such discussions will be made available across Peanut Innovation Lab to better inform all Peanut Innovation Lab researchers. Data collection, management and analysis will be a continuing topic at all Peanut Innovation Lab meetings, workshops and training sessions.

Data Access and Sharing

Ultimately, it is the goal of Peanut Innovation Lab to make all data produced by the program freely available to the global public. To do this effectively will require not only appropriate data access systems, but also data standards and metadata to allow the global public better utility and understanding of the data provided. The above description of Peanut Innovation Lab's efforts in using standardized data acquisition methodologies, ontologies and nomenclatures will be critical in meeting the requirements for global public use.

As Peanut Innovation Lab is a research program, it is expected that the scientists involved publish their results in refereed journals and/or conference and workshop proceedings. Authors will be encouraged to publish under the Creative Commons Attribution 4.0 Generic License (CC BY 4.0) or an equivalent license. The Peanut Innovation Lab would allow authors to pay using project funds reasonable fees required by a publisher to effect publication on these terms. This will be encouraged as it provides both credibility and visibility to the state-of-the-art research underway in the program. While this may result in some delay in releasing research data and results, such delays will be kept to a minimum and agreed upon up-front by the Peanut Innovation Lab community. It is proposed that all data supporting a publication should be made publicly available at the time of submission or at the time a manuscript is accepted for publication. Researchers will also be encouraged to publish in open source journals, and if this is not feasible, to make sure that an open source version of the article is made available to Peanut Innovation Lab. All publications, along with all data and supplementary information, will be made available to the general public via the Peanut Innovation Lab website and ultimately the DEC.

Data that is not targeted for publication would be made available no later than the agreed time period (e.g., one year) post completion of data acquisition, and all data would be made available at an agreed time point post-completion of the current Peanut Innovation Lab phase (January 2023). The Peanut Innovation Lab website will also be the source of other data (policy briefs, training materials, press releases, news stories, progress reports, etc.) produced by Peanut Innovation Lab research and activities.

Care will be taken to ensure that all data released is appropriately attributed to the contributor and made as anonymous as necessary. The sharing of research results will be consistent with all institutional policies governing intellectual property, copyright and the dissemination of research products.

Data Preservation

As the Peanut Innovation Lab Management Entity, UGA will maintain the Peanut Innovation Lab data system (website, data servers, etc.) according to the UGA Records Retention Policy, which currently requires all contracts, patents and agreements to be preserved indefinitely, and all accounting records for seven years after final financial report is submitted and the account is closed unless otherwise specified by the terms of contract (UGA OMB Circular A: 110-53). All other Peanut Innovation Lab data will be preserved for seven years after the close of the Peanut Innovation Lab program. All data that is housed on Cloud Servers outside of UGA will be backed up on UGA servers and ultimately transferred to the UGA Peanut Innovation Lab server at the end of the Peanut Innovation Lab program, or when the third-party server is no longer available. Where possible, all data will also be permanently archived in relevant disciplinary repositories (e.g., GenBank).

Peanut Innovation Lab Implementation Plan

Datasets and Intellectual Work to be produced during the next year will be indicated in the annual Peanut Innovation Lab work plans for each project.

The Peanut Innovation Lab [Principal Investigator/Scientist] for the project will be required to submit the appropriate dataset and supporting documentation, in conjunction with intellectual work, to the Management Entity for submission to the DDL and DEC as part of the annual reporting requirement.

Datasets must be submitted in a non-proprietary format such as CSV. Submissions using proprietary software such as Microsoft Excel are not permitted.

The Management Entity will confirm that the data does not include any proprietary or personally identifiable information.

The Management Entity will designate an access level for each dataset being "public, restricted public, or non-public". Note that data must be submitted to the DDL whether or not USAID chooses to release it publically.

The Management Entity will be responsible submitting the datasets, intellectual work and supporting documentation to the PRIL AOR for review.

The Management Entity will be responsible uploading the datasets, intellectual work and supporting documentation to the DDL and DEC.

Peanut Innovation Lab Data Management Table

See accompanying Excel file / Last Revised: 2/22/2018

Data Fields to be completed by the ME and PIs

Peanut Innovation Lab Project – Indicate which project/sub-award will generate the data

Dataset Type Generated by Project – Indicate type of dataset to be generated

1. Genomic sequence data
2. Molecular marker data
3. Crop phenotypic data from on-station field & greenhouse trials
4. Crop phenotypic data from on-farm field trials
5. Crop-specific phenotyping protocols
6. Farmer & household survey data
7. Anthropometric/health data from subjects
8. Seed value chain/local market data
9. Pest/pathogen phenotypic data
10. Other research data

Institution and Contact Person Responsible for Data – Indicate which institution and collaborator has prime responsibility for the data and can respond to needs for clarification

Description – Provide relevant details to describe the dataset: What information was collected, where, and when?

Data Privacy & Use Restrictions – Describe any privacy concerns or restrictions on making the data public, e.g. dataset subject to IRB compliance constraints; contains Personally Identifiable Information, medical histories, or other sensitive information; intellectual property/trade secrets; etc.

Pre-submission Data Processing – Describe any data processing steps (cleaning, anonymization, dithering to higher geospatial scale, etc.) that the project will perform on the raw dataset to ensure that data quality & data privacy concerns are addressed before making it publicly available.

Final Data Deliverable – What will the final submitted dataset consist of? Provide relevant details as appropriate, e.g. the file format, file size, number of entries, accompanying metadata, etc.

Timeline – Describe the life cycle of the data: e.g. collection, cleaning, analysis, embargo period, publication, and submission of final dataset to USAID (or other public data repository)

Data Repository & Post-Award Curation – In what public repository will the dataset be placed? (e.g. GenBank, USAID's Development Data Library (DDL), etc.) Who is responsible for maintaining the data after the award period ends?

Responsible Party – Who is responsible for ensuring data are submitted by the project to the final data repository?

Target Submission Date – When will data be submitted to the final location?

Associated Costs – Beyond the data collection, analysis, and publication costs included in your research budget, are there additional costs associated with depositing the data to a public repository (e.g. data processing, curation)?