

Africa RISING program data management plan

2019–2021



Produced by

International Food Policy Research Institute, International Institute of Tropical Agriculture and International Livestock Research Institute

Published by

International Institute of Tropical Agriculture

March 2019

www.africa-rising.net

The Africa Research in Sustainable Intensification for the Next Generation (Africa RISING) program comprises three research-in-development projects supported by the United States Agency for International Development as part of the U.S. government's Feed the Future initiative.

Through action research and development partnerships, Africa RISING will create opportunities for smallholder farm households to move out of hunger and poverty through sustainably intensified farming systems that improve food, nutrition, and income security, particularly for women and children, and conserve or enhance the natural resource base.

The three regional projects are led by the International Institute of Tropical Agriculture (in West Africa and East and Southern Africa) and the International Livestock Research Institute (in the Ethiopian Highlands). The International Food Policy Research Institute leads the program's monitoring, evaluation and impact assessment. <http://africa-rising.net/>



This document is licensed for use under the Creative Commons Attribution 4.0 International Licence.

This document was made possible with support from the American people delivered through the United States Agency for International Development (USAID) as part of the US Government's Feed the Future Initiative. The contents are the responsibility of the producing organization and do not necessarily reflect the opinion of USAID or the U.S. Government.

Contents

Executive summary	1
1. Data generation and confidentiality.....	2
1.1 Data generation.....	2
1.2 Data confidentiality	2
2. Data management.....	3
2.1 General background	3
2.2 Main data types.....	3
2.3 Data formats.....	4
2.4 Data management guidelines.....	4
3. Data security and acknowledgement	6
4. Online data management and data entry tools	7
4.1 Dataverse.....	7
4.2 Data entry application of the project mapping and monitoring tool (PMMT).....	14
5. Data sharing.....	16
5.1 Sharing of restricted non-confidential datasets on Dataverse with the public	17
5.2 Sharing of restricted non-confidential datasets within Africa RISING	17
5.3 Sharing of confidential data within Africa RISING	17
6. Appendices	18
Appendix A. Data user agreement for restricted non-confidential Africa RISING datasets (on Dataverse with the public)	18
Appendix B. Data user agreement for confidential data within Africa RISING	20
7. Bibliography.....	23

Executive summary

Africa Research in Sustainable Intensification for the Next Generation (Africa RISING -AR-) is a research for development program designed to pilot agricultural innovations for the sustainable intensification of core farming systems in sub-Saharan Africa. It comprises three linked projects in West Africa (Ghana and Mali), East and Southern Africa (Tanzania, Malawi, and Zambia) and Ethiopian Highlands with a separate Monitoring, Evaluation and Impact Assessment Component. The program is funded by the United States Agency for International Development and is a collaborative effort by several international and local research institutions, including the International Institute for Tropical Agriculture (IITA), International Livestock Research Institute (ILRI), Michigan State University (MSU), Wageningen University (WUR), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), and International Food Policy Research Institute (IFPRI).

During Phase I (2011–2016), the primary hypothesis was that sustainable intensification of mixed crop-tree-livestock systems leads to increased whole farm productivity, which in turn translates into better economic and development outcomes, including nutrition security. It was further hypothesized that a combination of innovations is more likely to increase whole farm productivity than a single intervention. During this phase, selected interested farmers participated in the testing and validation of alternative innovations that better fit their own circumstance. The type of innovations -and their delivery methods- varied across time, space, and local context. For Phase II (2016–2021), the program has adopted an ambitious but achievable vision of success to significantly increase its engagement with research households while at the same time partner with development actors for the scaling of promising innovations from Phase I to more than one million households.

During Phase I, Africa RISING generated a vast amount of agronomic and socioeconomic data that were managed through the Comprehensive Knowledge Archive Network (CKAN) platform operated by ILRI. Since 2017, to improve data management efficiency, program data have been migrated to a new platform – Dataverse – managed by Harvard University. As of March 2018, close to 100 datasets and supporting documentation were being handled through Dataverse.

This document outlines Africa RISING data management guidelines by incorporating relevant updates to the data management plan adopted during Phase I. The document aims to provide guidelines on data management practices, including Dataverse data uploading and sharing. The present data management plan has been discussed, agreed upon, and approved by the three CGIAR leading agencies (IITA, ILRI, and IFPRI), in consultation with the donor (USAID). Therefore, **all Africa RISING staff and collaborators must comply with the provisions and rules herein.**

The rest of the document is organized as follows. Section 1 discusses core issues associated with data production, ownership, and confidentiality in Africa RISING. Section 2 presents elements of data management, together with the different types of program-generated data and their formatting. Section 3 discusses data security and acknowledgement. Section 4 illustrates the online data management and project mapping and monitoring tools. Section 5 sets the guidelines for sharing Africa RISING data within and outside the program.

1. Data generation and confidentiality

1.1 Data generation

Africa RISING partners and organizations with an Internal Review Board (IRB) must secure ethical approval from their respective IRB offices before collecting data from human subjects. Africa RISING partners collecting data from human subjects must secure (written or oral) informed consent from human subjects before requesting data from them.

Furthermore, the program has developed detailed engagements guidelines on how to involve farmers into research, which help to facilitate relationships between researchers, farmers and other stakeholders, as well as foster ethical conduct within the Africa RISING community. They are accessible at: <http://hdl.handle.net/10568/34830>

1.2 Data confidentiality

All Africa RISING data to be uploaded onto the data repository must be anonymized, i.e., stripped of personal identifying information (PII)¹ including names, telephone numbers, and GPS coordinates. Researchers shall keep confidential data in a separate file and store them in a secure location other than the data repository. This is in line with donor requirements and other data protection regulations, such the General Data Protection Regulation (GDPR) of the European Union (<https://www.eugdpr.org/>). Since datasets uploaded onto the data repository must be stripped of PII, researchers should encrypt datasets with PII (e.g. raw or uncleaned, written consent forms) to protect digital data confidentiality while stored on computer systems and transmitted electronically. To protect the confidentiality of hard copy documents, researchers should exercise due diligence in identifying and implementing the best methods for storing and managing their records conditional on their own circumstances. Some of the methods include storing confidential hard-copy documents in a restricted, locked down cabinet, limiting access to such documents, and destroying records securely if they are no longer required.

¹ PII is defined as information that can be used (on its own or combined with other information) to identify, contact, or locate a single person, or to identify an individual in a specific context.

2. Data management

2.1 General background

Africa RISING is generating various types of data as part of the research and monitoring and evaluation activities. These data should be managed properly to help gauge progress towards achieving program goals and increase the evidence base on program impact. Within Africa RISING, the distinction between data collected on a repeated basis with a full data set required to properly test project hypotheses versus one round data collection that constitutes a full data set suitable for testing project hypotheses is specifically important. Other times robust analyses may require multiple data 'sets', which are collected annually for several years (e.g. experimental replication panel data).

Ideally, annual workplans should identify the hypotheses being tested and which data need to be collected over time, when and how they will be collected and analyzed based on the hypotheses being tested. Information on data collection and analyses should be discussed in the "Materials and Methods" section of each proposal. Chief Scientists and Principal Investigators are responsible for ensuring that data collection information is clearly identified in the workplans, confirm that the data have been collected and uploaded annually -or on an appropriately regular basis-, and work with research teams to identify the appropriate timelines for Open Access. The minimum data sets required by each research team to test its research hypotheses shall be guided by the program's Sustainable Intensification Indicator Framework (SIIF). Working with their team members, principal investigators should identify additional data needs (above and beyond what is outlined in the SIIF) that apply to their respective projects. For example, when evaluating the performance of a new variety in a particular environment, information about days to flowering, number of pods, and pest infestation severity shall be collected.

2.2 Main data types

In this document, data are defined as observational, experimental, simulated, or derived representation of information (MIT 2014; UMN 2014) as summarized in **Error! Reference source not found..**

Table 1. Africa RISING data types

Data Type	Description
Observational	Quantitative and qualitative information collected through surveys or captured from sensors to measure the intensity of program interventions, as well as their agroeconomic and environmental impacts
Experimental	Information collected in a laboratory or other controlled settings (e.g., through mother-baby trials)
Simulated	Information that is projected/predicted (e.g., from experimental data to the national scale) using computer assisted modelling tools
Geospatial contextual	Information collected from secondary sources to better contextualize program activities and outcomes (e.g., weather and cropping calendar data, metric conversion units)
Physical	Specimens collected or created (e.g., soil samples)
Project monitoring	Data about project/program activities and outputs (e.g., innovations being tested and scaled up, # of households & villages targeted) for project monitoring.
Feed the Future Indicators	Data on the Feed the Future indicators applicable for Africa RISING (shall be collected by all implementers)

2.3 Data formats

Digital Africa RISING data can be documented in various formats as shown in **Error! Reference source not found.**

Table 2. Africa RISING data types and formats

Storage types	
Display/interface	File types
Text	*.do, *.doc, *.docx, *.doct, *.rtf, *.txt, *.pdf
Numeric	*.xls, *.xlsx, *.xml, *.csv, *.sav, *.dta, *.dat, *.df, *.sas, *.lst, *.mdb, *.accdb
Images/video/audio	*.jpg, *.png, *.bmp, *.gph, *.pdf, *.wav, *.mp3, *.mp4, *.gtff

2.4 Data management guidelines

For experimental data, supporting documentation includes protocols outlining research activities as well as other contextual information. When possible, experimental data should be accompanied by contextual data (e.g., about weather condition and cropping calendars) as well as the Geographic Information Systems (GIS) coordinates of experimental plots (see the Data access section for details on the management of confidential data). GIS data will allow researchers to link field/survey data with other geo-referenced contextual data (e.g., weather and biophysical conditions) to control for variables that determine agricultural potential and mediate the interaction between program interventions and agroeconomic outcomes.

For observational and physical data, supporting documentation includes notes about the sampling design and coverage, survey tools, codebook, and metric conversions if applicable. Africa RISING partners collecting these data should ensure that variables and response codes are appropriately labelled and can be understood by users. Data reported in local measurement units shall be accompanied by information about metric conversions. For example, an experiment involving the use of “one bag of fertilizer” should include information on the metric conversion rate, provided either in trial protocols or as supplemental information.

Data for monitoring research activities shall be generated through the cataloguing of all households engaged in the various research activities. This information should include detailed geographic and household identifiers, household-specific innovations being tested, dates when households joined the program, and organizations involved. These data are especially important for Africa RISING, given the diversity of innovations and farming systems as well as the number of partners involved.

Data for monitoring scaling activities shall include basic information about innovations being scaled up, development partners in charge of the scaling, as well as the total number of households that benefited from scaling-up activities. This information shall be disaggregated by the type of AR innovation, gender of beneficiaries (e.g., gender of the household head or

individual beneficiary as applicable), reporting period, and other dimensions of interest that are relevant to the project.

Data on Feed the Future (FtF) indicators shall be compiled throughout the year for subsequent reporting at the end of the fiscal year. Considering the timeline necessary for uploading FtF indicators data onto the USAID's FtF Monitoring System (FtFMS), each research team should upload complete data on the FtF indicators (along with the 10% discrepancy narratives) onto the Program's Project Mapping and Monitoring Tool (PMMT) by October 31st of each year or a month before the closing of the FtFMS portal, whichever comes first. This will allow enough time to sort out possible issues related with incomplete information, aggregation, and data upload onto the FtFMS in a timely manner. Researchers should compile and upload FtF indicators data onto the PMMT using the login credentials assigned to them. If any AR staff does not have login credentials (user name and/or password), s/he could contact the M&E team for requesting them.

3. Data security and acknowledgement

All original data must be stored by the creator/s for a minimum period of **four years**. During this time, physical specimens such as plants or paper, must be placed in spaces where environmental degradation can be minimized. Any physical specimens associated with personal identifiers should be stored in spaces that can be physically secured by lock and key. Electronic data with PII must have at least one level of security. The data may be protected by a password known only to authorized users. It should also be encrypted, with decryption keys communicated only to creator/s.

CGIAR or USAID provided email is an acceptable medium to transfer digital files. No third-party email addresses should be used unless belonging to collaborators and data creators. In cases where digital files are too large to be shared through email, data creators and users could use file sharing services, such as LeapFile or Dropbox. Files in these services must be shared peer-to-peer or in folders to which only the interested parties have access. When physical specimens are transferred, Africa RISING teams should use only DHL, Fedex, or UPS where available, as the quality, timeliness, and confidentiality of service among public mail carriers is highly variable in Africa RISING countries.

Any scientific report, journal article, draft paper or publication in any form resulting from the use of the data should acknowledge the provider institution and its staff involved as data providers and curators. **Each time a dataset is used, it must be cited by the user** as suggested in Dataverse or other preferred citation combining the creator, data production year, and data title. The citation for the dataset can be found at the top of the dataset page in a blue box as shown below, which will give the option to download citation as EndNote XML, RIS Format, or BibTeX Format.

The screenshot shows the Harvard Dataverse interface. At the top, there's a navigation bar with 'About', 'User Guide', 'Support', 'Sign Up', and 'Log In'. Below this, the 'Africa RISING' logo and 'Dataverse' text are visible. The main content area displays the dataset title: 'Malawi Africa Research in Sustainable Intensification for the Next Generation (Africa RISING) Baseline Evaluation Survey'. Below the title, there's a 'Metrics' section showing '2,440 Downloads'. A 'Cite Dataset' button is circled in red, indicating where users can find citation information. The citation text at the bottom reads: 'International Food Policy Research Institute (IFPRI), 2015, "Malawi Africa Research in Sustainable Intensification for the Next Generation (Africa RISING) Baseline Evaluation Survey", <https://doi.org/10.7910/DVN/28557>, Harvard Dataverse, V1, UNF:5:LtVDFxHKkgEMf1T4ScoQhQ==

4. Online data management and data entry tools

Given the scale and geographic coverage of Africa RISING activities and the large volume of data generated by the program, two online tools have been developed/adopted by Africa RISING. The first tool is a repository for all data generated as part of Africa RISING, **Dataverse**. The second tool is the **project mapping and monitoring tool** (PMMT), with a data entry application that allows researchers to create data reports on FtF indicators as well as other indicators relevant to their respective projects. The PMMT was developed by IFPRI during Phase I of the program and went through several iterations over the lifetime of the programme. Below are specific details about the two tools.

4.1 Dataverse

Dataverse is an open source web application for storing and sharing various types of data to increase the impact of research, owned and managed by Harvard University. Details about Dataverse can be found here: <http://guides.dataverse.org/en/latest/>.

During Phase I, Africa RISING data and supporting documentation have been managed using the CKAN platform managed by ILRI. In February 2017, the Africa RISING PCT agreed that program data shall be managed through another platform, Dataverse. The decision to migrate program-generated data was taken to overcome some of the operational challenges experienced during Phase I with data management, storage, and access. Working with ILRI and IFPRI's Communications and Public Affairs (CPA) Divisions, the M&E team has migrated program data (except ILRI-owned data) from CKAN to Dataverse, now accessible on a dedicated Africa RISING Dataverse webpage here:

<https://dataverse.harvard.edu/dataverse/AfricaRISING>. For ILRI-owned datasets, metadata records have been created in Dataverse with links pointing to actual datasets stored in CKAN. Upload of Africa RISING data on Dataverse is compulsory and provides a back-up of datasets collected through US federal funds to avoid loss of data.

4.1.1 Dataverse management

Africa RISING subpage on IFPRI-Dataverse is managed by IFPRI's Dataverse administrator in the Communications and Public Affairs (CPA) Division with support from the M&E team. Data management responsibility includes reviewing metadata forms received from partners for completeness and consistency; reviewing submitted datasets to ensure they do not contain any PII; uploading metadata and datafiles onto Dataverse; contacting data providers when potential data users request access. As part of its responsibilities, IFPRI manages the data received by AR researchers under the international rules and regulations protecting PII, for the sole purpose of data management, to maximize the benefits of the programme, and in the sole interest to spur synergies across AR researchers, donor, and data requestors.

4.1.2 Creating an account in Dataverse

To maximize the efficiency of Africa RISING data management using Dataverse, Africa RISING project managers, chief scientists, and researchers shall open an account in the Dataverse web application following this link <https://dataverse.harvard.edu/>, by clicking the "Sign Up" button at the top-right corner of the page.



IFPRI's Dataverse administrator grants subscribers the right to access and download their own dataset. Subscribed researchers with an account should share their login credentials (not their access password) with the M&E team so that each researcher can be mapped to the dataset submitted. Unfortunately, researchers who do not share their login credentials with the M&E team will need to submit a request to download their own data. Creating a Dataverse account is also necessary for individuals who would like to request access to restricted Africa RISING datasets.

IFPRI Dataverse administrator can assign a custom role to project managers, chief scientists, and principal investigators who wish to access all datasets uploaded by researchers in their team or region to allow them to review (and download) all datasets submitted by researchers working under their supervision.

When a researcher leaves Africa RISING, the PI/project manager shall communicate the name and email address of the replacement staff to IFPRI's M&E team so that contact information in Dataverse can be updated accordingly. The update will allow the replacement staff to access datasets uploaded by the former staff, who will no longer access Dataverse, as well as receive data access requests.

4.1.3 Uploading data onto Dataverse

By sharing their datasets and supporting documentation with the IFPRI M&E team for upload onto Dataverse, data providers accept their data to be uploaded onto Dataverse. Data providers are also responsible for ensuring that their submitted datafiles do not include any PII in line with donor and other privacy regulations, including the new General Data Protection Regulation (GDPR) of the European Union (See Section 1). The following main steps must be followed when uploading data onto Dataverse.

4.1.3.1 Completion of Dataverse metadata template by data providers

A complete metadata is needed for each dataset uploaded onto Dataverse. The metadata is a description of the actual data, including but not limited to the source, methods, and tools used to generate the raw data. It is accessible to the public and is essential for finding, identifying, citing, and understanding data sources. If any of the datasets have been transformed in any way, the metadata should include information about how the transformation was conducted and what has changed compared to the previous version(s). All data uploaded onto Dataverse will be assigned a unique persistent identifier in the form of digital object identifier (DOI) (DataCite, 2011). See **Error! Reference source not found.** below for Dataverse metadata template.

Table 3. Dataverse metadata template

Data field	Description of data field
Dataset title	Full title by which the dataset is known. Please choose a concise, self-explanatory title. Avoid using abbreviations and long titles.
Related Publication	Publications that use the data from this dataset. If available, please include url to relevant publications and reports based on these data
Description	A summary describing purpose, nature, and scope of the dataset

	(no word limit, although we suggest keeping it to the maximum of two short paragraphs)
Contributor	The organization/s or person/s responsible for either collecting, managing, or otherwise contributing in some form to the development of the resource.
Related Datasets	Any datasets related to this dataset, such as previous research on this subject
Production Date	Date when the data collection or other materials were produced (not distributed, published or archived)
Producer	Person/s or organization/s with the financial or administrative responsibility over the dataset
Collaborative organizations	Organizations involved in the data production
Funding organizations	Grant number and related acknowledgements (if available)
Distributor	The organization designated by the author or producer to generate copies of the particular work (dataset) including and necessary editions or revisions
Time period covered	Time period to which the data refer. This reflects the time period covered by the data, not the dates the data were collected
Date of collection	Contains the date(s) when the data were collected
Geographic coverage	Information on the geographic coverage of the data. Includes the total geographic scope of the data
Geographic Unit	Lowest level of geographic aggregation covered by the dataset, e.g. village, county, region
Universe	Description of the population, the group of people, or other elements that are the object of the study and to which the study results refer. Age, nationality, and residence commonly help to delineate a given universe.
Unit of Analysis	Basic unit of analysis or observation that the dataset describes, such as individuals, families/households, institutions or organizations
Sample and sampling procedure	Type of sample and sample design used to select the survey respondents to represent the population. May include reference to the target sample size and the sampling fraction
Kind of data	Type of data included in the file: survey data, census data, aggregated data, process produced data, time budget diaries, etc. Is this dataset part of a multiyear experiment? Yes/No Expected completion date of the experiment Year _____ Month _____

4.1.3.2 Electronic submission of completed metadata and data files

After completing the metadata form, data providers should submit the form, the corresponding dataset, and other supporting documentation to the IFPRI's M&E team electronically (through email). Supporting documentation includes codebooks, survey questionnaires, sampling methodology, and other documents to assist with proper processing, extraction, and interpretation of the data.

4.1.3.3 Uploading datasets onto Dataverse

Uploading data on the program repository platform –Dataverse– is one important feature that the present data management plan is set to guide. As mentioned, data upload is a mandatory and necessary back up of the data collected to avoid potential losses.

- All de-identified socioeconomic, agronomic, unit-record level AR data collected as part of the program (for which AR funds have been used, even partially) **must** be uploaded onto Dataverse **at least every year, whether they are part of a multiyear experiment or not,** as part of the donor’s policy regarding grantee reporting requirements.
- Consistently with the rules for Open Data set by the CGIAR Policy (2013a), de-identified **AR datasets (in raw, unprocessed form at the unit-record level) that are not part of a multiyear experiment must be made public access within 12 months** of completion of the data collection (embargo period). As such, data uploaded on Dataverse will be restricted (to the public) until the 12 months expiration of the embargo period, after which they will be openly accessible by the public. According to CGIAR Policy (2013a:5) “Open Access means the immediate, irrevocable, unrestricted and free online access by any user worldwide to information products, and unrestricted re-use of content (which could be restricted to non-commercial use and/or granted subject to appropriate licenses in line with the CGIAR IA Principles), subject to proper attribution.”
- For AR datasets that are part of a multiyear experiment, the embargo period extends until the experiment comes to an end and “complete” (i.e. multiyear) datasets are available. This is consistent with the CGIAR policy regarding "completion of the data collection or appropriate project milestone" as well as USAID Open Data Policy (<https://www.usaid.gov/sites/default/files/documents/1868/579.pdf>)

Monitoring timely data uploads and embargo period is the primary responsibility of the regional data managers and the Chief Scientists based on the commitments made in the workplans, with the support of the M&E team managing the Dataverse data repository.

After checking for completeness of documentations (such as metadata, codebook, questionnaire, sampling methodology) and conformity of submitted datasets with general privacy and data protection regulations (discussed in Section 1), the Dataverse administrator will upload data and supporting documentation submitted by researchers. The review of received documentation and datasets includes the following: verifying that all the applicable fields in the metadata template have been populated; checking completeness of supporting documentation for reuse (e.g., codebook, questionnaire, protocol); checking appropriate variables labelling; checking consistency between codebook and values (including value labels) in the datasets; checking the absence of PII (e.g., name, phone number, address, national identification number). Researchers must ensure that metadata template are fully populated, and datasets are accompanied by supporting documentation necessary for properly tagging the datasets, as well as for interoperability and discoverability of the data.

All Africa RISING datasets will be uploaded onto Dataverse as restricted files accessible only by authorized users, unless the data provider authorizes their access to the public (i.e., makes them open data). If researchers later update datasets, datafiles (or part thereof) already uploaded on Dataverse, they should inform the M&E team of the changes, and the associated updated dataset should be uploaded onto Dataverse with accompanying updated

metadata. When updated files are uploaded, or older files replaced, within a dataset, Dataverse will automatically create updated versions of the published data. The outdated versions can be accessed through the “Versions” tab as shown below (and then clicking on the version chosen).

Files	Metadata	Terms	Versions
Dataset	Summary	Contributors	Published
1.1	View Details	IFPRI KM	February 21, 2018
1.0	This is the first published version.	IFPRI KM	November 21, 2017

4.1.3.4 Registering Dataverse AR data in the USAID Development Data Library

Once the embargo period is over, Africa RISING datasets will be made open data and linked to/registered in the USAID’s Development Data Library (DDL). Details about the DDL can be found here: <https://www.usaid.gov/data>.

4.1.4 Accessing Dataverse datafiles

Without signing into Dataverse, users can search for datasets and files and view descriptions of restricted datasets. To download datasets, follow the steps below.

1. Search for the relevant dataset on Africa RISING page in Dataverse by entering key words in the search field. Data user will then be re-directed to a pop-up screen, requested to provide minimum data about the request, and asked to accept the Terms of Use to download the data.

Dataset Terms

Please confirm and/or complete the information needed below in order to continue.

Terms of Use



This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

Name *

Guest

Email *

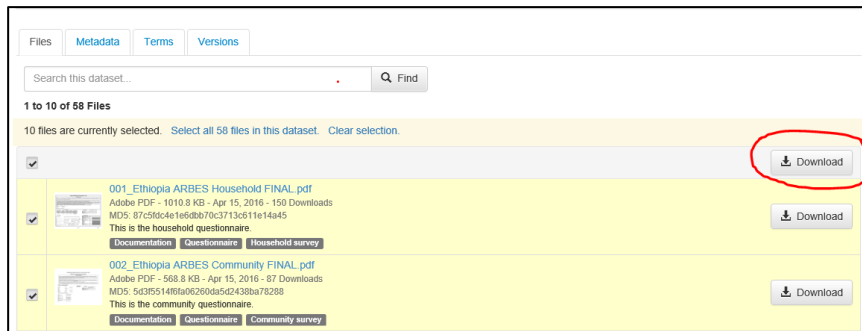
Institution *

Position

Accept

Cancel

2. For Africa RISING datasets that are currently open data (i.e. IFPRI's baseline survey data), prospective data users can directly download the datasets by clicking the download button to the right of the dataset as shown below.



3. For Africa RISING datasets that are restricted (all other datasets except IFPRI baseline survey data), prospective users must request access by filling out the Data User Agreement for non-confidential datasets using the Google form at the top of the Africa RISING Dataverse page.

- Prospective data users must first create a Dataverse user account (see Section 4.1.2), necessary for mapping the data requestor with the specific dataset requested. It is imperative that data requestors enter the correct Dataverse user account when completing the DUA, as highlighted above.
- By clicking on the Google Form link, users will be directed to a form where they are required to fill in personal details on their Dataverse user account, select datasets of interest (up to three datasets per request), provide the reason for the data request, and agree with the terms of the DUA. A screenshot of the online DUA is shown below.

The screenshot shows a Google Form titled 'Agreement for Using Africa RISING Data on Dataverse'. The form includes the following sections:

- Header:** 'Agreement for Using Africa RISING Data on Dataverse'
- Instructions:** 'Please ensure that you have already created a Dataverse account; you will need to provide your ID to successfully submit this form. If you have any questions or encounter any problems filling out this form, kindly email: IFPRI-AfricaRising@cgiar.org'
- Required Fields:**
 - Email address ***: A text input field with the placeholder 'Your email'.
 - Name ***: A text input field with the placeholder 'Your answer'.
 - Dataverse Account ***: A text input field with the placeholder 'Your answer'.
- Footer:** 'Please ensure you enter your correct Dataverse account. Permission, if granted will be linked to the account that you provide here. If you do not have a Dataverse account, you can create one by following this link: <https://dataverse.harvard.edu/dataverseuser.xhtml?>

Agreement for Using Africa RISING Data on Dataverse

* Required

Data Access

You agree that if permission is granted, you will use the data for the following purposes(s) and for no other purpose (check all that apply): *

☐ Secondary analysis of the data
 ☐ Comparative study - comparing the data with information from another dataset
 ☐ Followup study of the original sample
 ☐ Other: _____

Please provide a brief description (1000 characters max) of your proposed research and how the data you are requesting will be used in your research. *

Your answer

- Once the DUA is completed, the data request will be submitted to the IFPRI M&E team, who will in turn contact the data provider to seek permission to access the requested dataset.
 - With the goal of automating the process as much as possible, the final permission granted will be recorded on a Google sheet. All the data providers will be given access to this Google sheet.
 - Once a new request is received, the relevant data provider will receive an email notification informing on the request and ask for approval or denial on the Google sheet. The email will re-direct the data provider to the Google sheet where s/he can see details about the requestor, including the expected use of the requested dataset.
 - Additionally, dataset providers will receive periodic reminders of pending data requests.
- Upon approval from the data provider, the IFPRI M&E team will give permission to the data requestor to download the requested data.

Africa RISING project managers and chief scientists will also be able to review the Google sheet containing details about data requests approved and denied.

4.1.5 Monitoring Dataverse data requests

The integration of the DUA into Dataverse using the Google sheet would allow the Africa RISING Dataverse administrator to keep track of requests and permissions in a systematic and seemingly automatic manner. Once a form is filled out by the requestor, a record of the specific request is collected as an observation in the Google sheet to be reviewed by the M&E team, data providers, and program managers.

Once the data provider accepts (or denies) a request, IFPRI Dataverse administrator will send out an email to the requestor informing her/him of the decision of the request. In case permission is granted by the data provider, the administrator will give access to the requestor for that specific dataset.

Given strict privacy regulations, the requestor will be given prior notice of the fact that her/his personal information will be shared with the data provider, IFPRI Dataverse administrator, and Africa RISING project managers. Additionally, the requestor will be alerted that filling out the Google form DUA implies that s/he agrees and abides to the present Africa RISING Data Management Plan and privacy policy.

4.2 Data entry application of the project mapping and monitoring tool (PMMT)

The Africa RISING data entry application allows users with appropriate credentials to add/modify data on FtF Indicators and other “custom” indicators through an intuitive, step-by-step interface. Behind the PMMT is a database that stores detailed project information that describes project activities, their location, and other secondary data layers.

SIGN IN







Sign In

Upon authentication, a data-entry application will be accessible to Publishers, Editors, and Viewers. The data entry and reporting module supports the collection of project indicators at the site level. For each site, any number of indicator values can be entered, including both pre-defined Feed the Future indicators as well as user-defined (custom) indicators specific to the activities occurring at a site.

The data entry application has three types of users, each with different use cases and use privileges. A single user can fill these roles for different action sites:

- **Viewer** – A viewer has the opportunity to browse information about AR sites. Where data is available, they can generate reports based on existent statistics in the PMMT.
- **Editor** – An editor has all the rights that a viewer also does. In addition, an editor can make changes to values in existing reports in the PMMT.
- **Publishers/Administrators**– In addition to all the rights that an editor and viewer also have, a publisher can create new reports and add or change other users' access rights.

The IFPRI's M&E team has created a PMMT account for all Africa RISING researchers and sends an email blast every summer (well ahead of the FtfMS reporting period) to remind users of their login credentials.²

² A User Guide on the PMMT Data Entry Application can be found here:
<https://www.dropbox.com/s/bo93efl0b9xvt0z/Africa%20RISING%20PMMT%20Data%20Entry%20App%20User%20Guide%20Sep%202018.doc?dl=0>

5. Data sharing

Africa RISING has made a significant commitment to deliver, maintain (and enhance, where appropriate) an open-access data management platform, in line with the USAID and CGIAR respective Policies. Public sharing of research outputs has the potential to speed research, spur innovation, and generally facilitate the learning process. In the context of developing countries, open data access has the potential not only to increase innovation, but to increase the visibility of and opportunities for emerging country researchers (DFID 2013). Closed and paid access favors researchers who can afford covering costly subscriptions, increasing the costs of doing research, and decreasing the efficiency of research utilization.

Unrestricted open data allows developing country researchers to pursue research topics that are relevant to their countries and interesting to local stakeholders. It also broadens the potential impact of research to developing country residents. It is only logical and fair that data collected in a developing country are available to its citizens, and data collectors should not be viewed as data owners, but data custodians/providers, the owners being the subjects from whom the information was collected.

The CGIAR Open Access and Data Management Policy (“the Policy”) was approved and adopted by the CGIAR Consortium and is effective as of October 2, 2013 with a 5-year transition implementation period with demonstrable implementation by the end of 2018. The Policy is not part of the Common Operational Framework of the CGIAR, but, as a CGIAR Consortium Policy, **implementation of and compliance with the Policy is mandatory for the CGIAR Consortium and its members**, and partners within the scope of the Strategy and Results Framework and the CGIAR Research Programs. In Section 4.2.4 of the Policy it is stated that:

Different information products are subjects to different rules for ownership and diffusion, under the CGIAR (2013a) Open Access and Data Management Policy. Below are some guidelines for sharing the different types of data generated as part of Africa RISING.

The USAID DRAFT Research Policy (February 2014) states that “Improving the accessibility of USAID funded data can bring the Agency, its partners, and other stakeholders a deeper and more up-to-date understanding of development challenges, which in turn will help USAID and others design, manage, and evaluate development programs more effectively. By making USAID data available through user friendly platforms in machine readable formats, host countries, scientists, and communities can propel research forward in solving complex development problems.” Open access is the barrier-free and permanent right of any global citizen to view full versions of research outputs. While the goal of open access is to share data to maximize its impact, it must be pursued with the protection of third parties in mind (CGIAR, 2013a).

AR researchers are expected to share data among them and, under certain conditions outlined below) with the general public as a requirement by the donor. If a dataset is part of

a multiyear experiment, the corresponding datafile on Dataverse shall be made open data after the experiment comes to an end. A multiyear experiment is one where data must be collected on a repeated basis with the full data set required to properly test project hypotheses outlined in the project proposal.

5.1 Sharing of restricted non-confidential datasets on Dataverse with the public

All Africa RISING datasets on Dataverse are currently restricted³, i.e., they are not accessible to the public. These datasets can only be accessed upon approval from the data provider. Prospective data users first need to create a Dataverse account (see Section 4.1.2) and then complete the program DUA for non-confidential data. The IFPRI M&E team will subsequently submit the completed DUA forms with the data provider who will then decide whether to grant the requestor access. Over the course of the multiyear experiment, data access and sharing with institutions and individuals not directly involved in the Africa RISING program will be discretionally determined by each data institution/research team provider on a case-by-case basis for legitimate research purposes and non-commercial uses as outlined in the DUA in Appendix A. After the experiment comes to an end, datasets will be made open data.

5.2 Sharing of restricted non-confidential datasets within Africa RISING

Africa RISING researchers are expected to share restricted non-confidential data with other Africa RISING researchers to promote cross-regional and cross-disciplinary collaborative research, irrespective of whether data are part of a multiyear experiment or not. Given that non-confidential datasets from multiyear experiments remain restricted to the public until the experiment comes to an end, Africa RISING researchers interested in a restricted dataset shall submit their request for access through Dataverse. Any controversy about access to datasets by Africa RISING researchers shall be brought to the attention of the Africa RISING project managers and the Program Coordination Team (PCT).

5.3 Sharing of confidential data within Africa RISING

For the sake of programmatic purposes (e.g., project monitoring, impact assessment, project evaluation, technology assessment, and so forth), Africa RISING researchers are also expected to share confidential data with other Africa RISING researchers. For example, data on innovations tested by program beneficiaries will inevitably contain identifying information about names and location of beneficiary farmers and households and experimental plots. Africa RISING researchers interested in accessing confidential data shall sign the confidential DUA in Appendix B with the data provider. Any controversy about access to confidential data by Africa RISING researchers shall be brought to the attention of the Africa RISING project managers and the Program Coordination Team (PCT).

³ Except datasets from Africa Research Baseline Evaluation Surveys (ARBES).

6. Appendices

Appendix A. Data user agreement for restricted non-confidential Africa RISING datasets (on Dataverse with the public)

1. Name of data requestor
2. Organization
3. Address (City and Country)
4. E-Mail Address:
5. **You agree that if permission is granted, you will use the Data for the following purpose(s) and for no other purpose (check all that apply):**
 - a. Secondary analysis of the Data
 - b. Comparative study - comparing the Data with information from another data-set
 - c. Follow-up study of the original sample
 - d. Other Specify_____
6. Please provide a brief description (about 250 words) of your proposed research and how the data you are requesting will be used in your research

7. Expected date of completion of your research: (mm/dd/yyyy)
8. Conditions for Data Access.
 - a. I will not use the data for any other purpose than the above specified request.
 - b. I will not knowingly divulge any information that could be used to identify individual participants in the study and I agree to use such precautions as are reasonably necessary to prevent such identification.
 - c. I will notify the data provider should the scope of the research change.

- d. I will make appropriate acknowledgment of the contributor and will include the citations as suggested under "How to Cite" tab of Dataverse when reports and/or publications are generated.
- e. I will not copy or release the data to any other person or organisation, directly or indirectly, without the written consent from the data provider. [Please send the request to IFPRI-Data@CGIAR.ORG].
- f. I abide by Africa RISING Data Management Plan.
- g. I am aware that the information I submitted as part of this data request will be shared with the data provider and Africa RISING project management to help them decide on whether to grant me access to the requested data.
- h. I agree with the Africa RISING Privacy Notice

BY TYPING MY NAME IN THE SPACE LABELED "SIGNATURE" BELOW (WHICH SHALL CONSTITUTE MY SIGNATURE) AND BY CLICKING ON THE "I AGREE" BUTTON BELOW, I CONFIRM:

- A. THAT I HAVE READ AND UNDERSTOOD EACH AND EVERY TERM SET FORTH ABOVE
- B. THE ABOVE REPRESENTATIONS AND THE INFORMATION I PROVIDE BELOW ARE ACCURATE
- C. THAT I HAVE THE AUTHORITY TO SIGN THIS FORM
- D. I AGREE TO BE BOUND BY THE ABOVE TERMS AND CONDITIONS
- E. AGREE NOT TO REPRODUCE ANY OF THE DATA, IN ANY FORM, WITHOUT PRIOR EXPRESS WRITTEN CONSENT FROM THE DATA PROVIDER.

Signature

Date

Appendix B. Data user agreement for confidential data within Africa RISING

[Insert Agency Name of Data Provider, Approving Official/Title, Contact Information (address, phone, email]

and

[Insert Name/Title Principal Researcher, Agency Name, Contact Information (address, phone, email, Country]

This Data Use Agreement is mutually made and voluntarily entered on [Insert Date] by and between [Insert Provider Name], hereafter “Provider” and [Insert Researcher Name], hereafter “Researcher.” Researcher will be responsible for the observance of all conditions for use and for establishment and maintenance of security agreements as specified in this agreement to prevent unauthorized use. As part of this agreement the Researcher has included an attachment containing justification for making the data request including study duration and personnel, a detailed plan as to where and how the data will be stored, who will have access, how data will be maintained, transported as well as how the data will be disposed of at the end of the study’s activities. Except for summary results any other request for the dissemination of the data is not covered by this agreement. That determination is a separate request. Summary results are those items which cannot be used to identify any individual participating in the study. It should be noted that the stripping of an individual’s name or individual identification number does not preclude the identification of that individual and therefore is not sufficient to protect the confidentiality of individual data.

1. This agreement sets forth the terms and conditions pursuant to which Provider will disclose certain data in the form of a Limited Data Set to the Researcher.
2. Permitted Uses and Disclosures
 - 2.1. Except as otherwise specified herein, the Researcher may make all uses and disclosures of the Limited Data Set necessary to conduct the research described herein: [include a brief description of the research and/or agency protocol number] (“Research Project”).
 - 2.2. In addition to the Researcher, the individuals, and or classes of persons, who are permitted to use or receive the Limited Data Set for purposes of the Research Project include: **[insert names or classes of persons who may use or receive the limited data set, e.g. the researcher’s staff, any collaborators, other clinical sites involved in the research, sponsors if applicable, outside laboratories]**. Access to the data covered by this agreement shall be limited to the minimum number of research individuals and to those individuals on a need-to-know basis to achieve the purpose stated in the protocol/study design. To the extent that the classes of

persons are not part of the Researcher's workforce who are directly involved in the Research Project, the Researcher shall notify and consult with the **[ORGANIZATION]** IRB Chair before adding such individuals or classes of persons. The approval from the Chair must be given in writing. The Researcher will then enter a data agreement with the other classes of persons before such release of the Limited Data Sets. All such notations will be sent via email to the **[ORGANIZATION]** IRB Chair within 7 days of the approval to include such individuals or classes of persons.

3. Researcher Responsibilities

- 3.1. Researcher will not use or disclose the Limited Data Set for any purpose other than permitted by this Agreement other than to the Research Project only or as required by law;
- 3.2. Researcher will use appropriate administrative, physical and technical safeguards to prevent use or disclosure of the Limited Data Set other than as provided for by this Agreement;
- 3.3. Researcher will report to the Holder any use or disclosure of the Limited Data Set not provided for by this Agreement of which the Researcher becomes aware within seven (7) days of becoming aware of such use or disclosure;
- 3.4. Researcher will ensure that any agent, (i.e. including donor, subcontractor), to whom it provides the Limited Data Set, agrees to the same restrictions and conditions that apply through this Agreement to the Recipient with respect to the Limited Data Set;
- 3.5. Researcher will not identify the information contained in the Limited Data Set. There are no intended third-party beneficiaries to this Agreement.
- 3.6. Researcher will not contact the individuals who are the subject of the Research Project contained in the Limited Data Set.

4. Term and Termination

- 4.1. The terms of this Agreement shall be effective as of **[insert effective date]**, and shall remain in effect until all data in the Limited Data Set provided to the Researcher is destroyed. An affidavit or statement certifying that the Limited Data set was destroyed outlining the means and method of the destruction should be provided to the Provider within seven (7) days of data destruction. The **[ORGANIZATION]** IRB Chair should be the contact to send the information. Such proof may be offered by email address to the **[ORGANIZATION]** IRB Chair and to the IRB to keep for their administrative purposes.
- 4.2. Upon the Holder's knowledge of a material breach of this Agreement by the Researcher, the Provider shall provide an opportunity for Researcher to cure the breach or end the violation. If efforts to cure the breach or end the violation are not successful within the reasonable time period ten (10) days specified by the Provider, the Provider shall immediately discontinue disclosure of the Limited Data Set to the Researcher and notify the IRB and agency officials of the breach. If the Provider agency officials determine that cure of the breach is not possible then the agreement is immediately terminated, and no other data will be provided to the Researcher. A report will be sent to the IRB who will in turn contact the

appropriate regulatory agency.

- 4.3. Upon Provider's knowledge of a material breach by the Requestor, the Provider shall have the right to immediately terminate this Agreement

5. General Provisions

- 5.1. This Agreement shall not be assigned by Researcher without the prior written consent of the Provider. The respective rights and obligations of Researcher under this Agreement shall survive termination of this Agreement. No provision of this Agreement may be waived except by an agreement in writing signed by the waiving party. A waiver of any term or provision shall not be construed as a waiver of any other term or provision.
- 5.2. Each party agrees that it will be responsible for its own acts and the results thereof to the extent authorized by law and shall not be responsible for the acts of the other party or the results thereof.

The persons signing below have the right and authority to execute this Agreement and no further approvals are necessary to create a binding agreement. In the event of any conflict between the terms and conditions stated within this Agreement and those contained within any other agreement or understanding between the parties, written, oral or implied, the terms of this Agreement shall govern.

IN WITNESS WHEREOF, the parties hereto execute this agreement as follows:

Date: _____	[INSERT NAME OF DATA PROVIDER] By: _____ (Title person with authority to sign agreement for the holder of the data)
Date: _____	RESEARCHER By: _____ (Title of researcher or person with authority to sign agreement for the recipient)

If your research requires the use of a data use agreement, a copy of the signed agreement (by both parties) should be submitted to the IRB office with other research materials for review. A copy of the agreement should also be retained by both parties involved in the research as an assurance of this agreement.

7. Bibliography

- AEA (American Economic Association). 2014. "The American Economic Review: Data Availability Policy." Accessed April 11, 2014. <http://www.aeaweb.org/aer/data.php>.
- CGIAR. 2013. CGIAR Open Access and Data Management Policy, approved October 2, 2013, and available at: <http://www.cgiar.org/consortium-news/cgiar-consortium-now-officially-open-access/>
- CGIAR. 2013. *DRAFT Implementation Guidelines for the CGIAR Open Access and Data Management Policy*, available at: <http://www.cgiar.org/consortium-news/cgiar-consortium-now-officially-open-access/>
- DataCite. 2011. DataCite Metadata Schema for the Publication and Citation of Research Data. Version 2.2. doi:10.5438/0005.
- DCC (Data Curation Centre). 2014. "Data Management Plans." Accessed April 9, 2014. <http://www.dcc.ac.uk/resources/data-management-plans>.
- DFID (Department for International Development). *DFID Research Open and Enhanced Access Policy v1.1*. London, United Kingdom: Department for International Development.
- Holdren, J.P. 2013. *Memorandum for the Heads of Executive Departments and Agencies*. Executive Office of the President. Accessed April 1, 2014. http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.
- MIT (Massachusetts Institute of Technology). 2014. "Data Management and Publishing." Accessed April 11, 2014. <http://libraries.mit.edu/guides/subjects/data-management/plans.html>.
- NSF (National Science Foundation). 2010. "Data Management for NSF SBE Directorate Proposals and Awards." Accessed April 11, 2014. http://www.nsf.gov/sbe/SBE_DataMgmtPlanPolicy.pdf.
- The White House. 2012. "We Petition the Obama Administration to Require free access over the Internet to scientific journal articles arising from taxpayer-funded research." Accessed April 1, 2014. <https://petitions.whitehouse.gov/petition/require-free-access-over-internet-scientific-journal-articles-arising-taxpayer-funded-research/wDX82FLQ>.

- UMN (University of Minnesota). 2014. "Creating a Data Management Plan." Accessed April 11, 2014. <https://www.lib.umn.edu/datamanagement/DMP>.
- USAID. 2017. Open Data: Rules and Responsibilities for Implementing Partners.
- USAID (United States Agency for International Development). 2014. USAID DRAFT Research Policy: Washington, DC.