



Africa RISING

Africa Research in Sustainable Intensification for the Next Generation

Data Management Plan

Carlo Azzarri, Cleo Roberts, Melanie Bacou, Maria Comanescu, Beliyou Haile

Produced by International Food Policy Research Institute, Monitoring and Evaluation Team

Published by International Food Policy Research Institute, International Livestock Research Institute, International Institute of Tropical Agriculture

November 21, 2014

www.africa-rising.net



The Africa Research In Sustainable Intensification for the Next Generation (Africa RISING) program comprises three research-for-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 a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License

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.



Table of Contents

1.	Africa RISING Program	1
2.	Data Management	1
	Data Sources	2
	Types and Formats.....	3
	Metadata Management and Requirements	4
3.	Data Access Tools.....	6
	Open Access.....	6
	M&E Open-Access Web-Site.....	7
	Africa RISING Baseline Evaluation household and community Surveys (ARBES)	8
	Project Mapping and Monitoring Tool.....	9
	Africa RISING Program Data.....	10
	Reference Data	11
4.	Ownership and Diffusion	11
4.1	CKAN platform	12
	Feature Evaluation.....	13
	Useful Links	14
	Cost	15
	Comparison with other platforms	15
4.2	Data Diffusion	15
4.3	Discretionary External Diffusion	17
4.4	Acknowledgment	17
5.	Security	17
	Storage.....	17
	Transmission	17
	Appendix A. Compendium of Africa RISING Tools	18
	Appendix B. Data Sharing Agreement.....	19
	References	21

1. Africa RISING Program

Africa Research in Sustainable Intensification for the Next Generation (Africa RISING) is a research for development program designed to pilot potential interventions for the sustainable intensification of mixed crop-tree-livestock systems and provide data and information that will lead to the better design of development projects. 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. Africa RISING is funded by the United States Agency for International Development and involves several 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).

The primary hypothesis of the Africa RISING Program is that sustainable intensification of mixed crop-tree-livestock systems leads to increased whole farm productivity, which in turn leads to development outcomes and better food and nutrition security for those who depend on these systems. It is further hypothesized that a combination of relevant interventions is more likely to increase whole farm productivity than single interventions. The hypothesis will be tested by implementing baskets of interventions in selected communities.¹ Within a community, interventions will be 'offered' to volunteers. The type of interventions -and delivery methods- is expected to vary across time, space, and local context, even across farms depending on the choice of the individual household. They will also vary according to the farm/household typology that will classify farm households 'sufficiently similar' in relation to the expected effects of Africa RISING. Farming systems analysis and modelling will be used to help identify and target appropriate interventions across different farm types and to perform ex-ante impact analysis. Crop modelling analysis can also be applied.

This data management plan has been developed to provide guidance on data management practices and standards for research institutions and teams working on the Africa RISING program. Upon discussion and agreement among partner institutions in the program, the plan can be considered binding for all participating research teams. The rest of the document is organized as follows. Section 2 discusses open data access, Africa RISING Program data sources and types, metadata management, and data standardization. Section 3 discusses Program data management and access tools. Section 4 discusses internal and external diffusion of Program data. Section 5 discusses data storage and transmission.

2. Data Management

Managing information created using US federally funded research is important for several reasons. First, it provides verification of which activities have been undertaken and which outputs they have produced. Second, it can shed light on any information deficits that the US government faces.

¹ The definition of 'community' varies between countries depending on the local administrative and geographical divisions.

Data Sources

Within Africa RISING program, data are being produced at various points in project implementation and evaluation. During the testing of interventions for sustainable intensification, various types of experimental data will be collected, such as improved seed varieties, fertilizers, management practices, biomass, soil coverage, water retention, water use efficiency and their combinations thereof. In addition, observational data will be collected about local farming systems and farmers who rely on them. Additionally, information on plant specimens and demonstration plots are likely to be available. Prior to the dissemination of technologies and eventual distribution of inputs, socio-economic baseline data have been collected from farmers. These observational data will be used to better tailor the design of research activities. As interventions are taking place, research teams will collect additional observational data on various forms of inputs distributed (such as combination of technologies and packages, training), and on the recipient farmers.

Data from Testing of Technologies and Management Practices: Africa RISING research teams are conducting a series of tests to determine which technologies and sustainable management practices perform well in controlled environments. They will observe the effects of the interventions (such as improved seed varieties, fertilizers, pesticides, inoculation, but also new agricultural and livestock practices and training, and their combinations) on a series of outcome variables (e.g., plant growth, agricultural yields). When possible, the research teams should keep digital records of test results, including where and when they were taken, which plant varieties and technologies were involved, and the outcomes of the tests. Geographic Information Systems (GIS) coordinates at different level (demonstration plots, watersheds, specific landscapes) and, when possible, of participating farmers should also be taken.

Monitoring Data: Throughout the life of Africa RISING it will be necessary to capture how the project is carried out. The research and M&E teams will need to coordinate to collect information about which farmers participate in field days, trainings, demonstration plots, and other trials. In addition, they will need to keep track of which households are beneficiaries of the project, which type of technology or technologies each household receives, and where they are located. The households' location should include GIS coordinates if possible.

Monitoring information may include the following:

- Type of project activity
- Targeted crops and/or livestock
- Africa RISING team(s) involved
- Number of Africa RISING participating beneficiaries and how they were selected
- Geographic distribution of beneficiaries
- Demographic and socio-economic characteristics of beneficiaries

Socio-economic and Agricultural Household Surveys: The M&E team administered a baseline household survey to Program beneficiary households, non-beneficiary households in the same communities, and non-beneficiary households in different, but similar, communities. These data have been collected to primarily assess the impact of the research activities on agricultural

performance and various socio-economic outcomes of interest, as well as to inform Africa RISING project management. This effort encompassed gathering community-level data in beneficiary and non-beneficiary villages to monitor research activities at a broader scale. GIS coordinates of all sampled households have also been collected. In addition to the M&E team, some of the research teams (e.g., in Zambia) have also conducted socio-economic and agricultural household surveys.

Types and Formats

Africa RISING data can be presented in various formats. See Table 1 below for some examples.

Table 1: Data Formats

Medium	Storage Types	
	Display/Interface	File Types
Digital	Text	*.do, *.doc, *.docx, *.doct, *.rtf, *.txt, *.pdf
	Numeric/spreadsheet	*.xls, *.xlsx, *.xml, *.csv, *.sav, *.dta, *.dat, *.df, *.sas, *.lst, *.mdb, *.acddb
	Images/video/audio	*.jpg, *.png, *.bmp, *.gph, *.pdf, *.wav, *.mp3, *.mp4, *.gtff
	Models	*.gms
	Web/software	
Physical	Plant specimens	
	Minerals	
	Paper records	
	Physical repositories of multimedia (i.e., tapes, CDs)	
	Other physical artifacts	

In this document, data are defined as the observational, experimental, simulated, or derived representation of information (MIT 2014; UMN 2014). Table 2 shows the types of data available in Africa RISING program.

Table 2: Data Types

Data Type	Description
Observational	Survey data (quantitative, qualitative) or information captured from sensors
Experimental	Information collected in a laboratory or other controlled settings (such as trial, control, or farmer-managed plots)
Simulated	Information projected using computers or other modelling mechanism
Derived/compiled/analyzed	Secondary information collected and/or contextualized; maps, graphs, and other data visualizations; software or web tools

Physical	Specimens collected or created in the field
Metadata	Descriptions of data, including but not limited to the source, methods, and tools used to collect data

Different information products are subjects to different rules for ownership and diffusion, under the CGIAR (2013a) Open Access and Data Management Policy (see below in Chapter 3 “Data Access Tools”).

Metadata Management and Requirements

For each of the data types listed above, proper documentation (metadata) is essential. Metadata is important for identifying and understanding data sources, and it is also necessary for the proper citation and integrity of the data. Moreover, data documentation facilitates the process of finding existing data, aiding in research visibility and efficiency, and in open access thus making data more discoverable (MIT 2014). Metadata should include the data type and its source. If any of the data have been transformed in any way, the metadata should include information about how the data were created. In addition, the documentation should include any preservation requirements and sharing guidelines or restrictions (DCC 2014). A unique persistent identifier should be created in the form of digital object identifiers (DOIs) (DataCite, 2011).

All filenames for Africa RISING must include the metadata information, used to describe the contents and context of data files. For all data, digital or not, the metadata file must be submitted in electronic format with the following general information:

Table 3: Metadata

<i>Element</i>	<i>Description</i>
Identifier	Unique and persistent identifier assigned to the dataset in its current form
Title	Name of the dataset
Creator	Person, team, or institution that produced the dataset
Format	Physical or digital medium in which the data are stored
Subject	General description of the contents of the dataset
Publication Year	Year that the dataset is made public
Version	Necessary if multiple versions of the data are released
Language/s	Language or languages in which the data are available
Funder/s	The person, team, or institution that provided the funds to collect and/or transform the data
Usage and Sharing Rights/Restrictions	List of usage and sharing rights, along with acknowledgment requirements
Keywords	Search terms that could lead a user to the dataset
Filename	For electronic data, a label associated with each file, which contains a brief description of its contents

Table adapted from MIT (2014) and DataCite (2011)

Different elements should be reported in the metadata depending on the data type (dataset or project). Each **dataset** metadata should include:

- Contact person
- Contact email
- Main researcher
- Other staff involved
- Partners
- Groups
- Commodities
- Technologies
- Subjects
- Other tags
- Region
- Country
- Sub-National Level
- Spatial coverage
- Date of collection
- Availability date
- Data owner
- Sharing agreement
- Usage conditions
- Citation
- Citation for acknowledgements

Each **project** metadata should include:

- Project title
- Project abstract
- Grant code
- Donor
- Partners
- Start date
- End date
- Spatial coverage
- Principal investigator
- Other staff involved

In addition to data, researchers in Africa RISING produce data collection tools (see Annex A for some examples of the tools developed and used). These tools may include questionnaires, focus group guidelines, or other templates and technical manuals used to organize and collect the data. All Africa RISING researchers must share data collection tools and supporting documentation to facilitate understanding of the data, and further enhance collaboration among partner institutions. To this end, and to keep track of all tools and data collection efforts, it is necessary to keep a repository and compendium of tools in Africa RISING.

All tools and data will be collected in the ILRI-CKAN web-based repository, accessible by researchers through a prominent link in the Project Mapping and Monitoring Tool (PMMT) website (see discussion in section 3). This compendium of tools used by each research team should be updated regularly and uploaded onto the ILRI-CKAN system.

3. Data Access Tools

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 (see below).

Open Access

Open access is the barrier-free and permanent right of any global citizen to view full versions of research outputs. It requires none of the onerous permissions processes or institutional affiliations often associated with proprietary data (DFID 2013; MIT 2014). 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).

Open access to US federally funded data is a pressing priority of researchers and government officials alike. Free internet access to scientific journal articles based on federally funded research was the subject of a petition to the White House signed by 65,704 individuals (WH 2012). Dr. John Holdren of the White House's Office of Science and Technology responded to this petition, agreeing that "Americans should have easy access to the results of research they help support." He issued a memorandum on February 22, 2013 directing the heads of executive departments and agencies with over \$100 million in research and development expenditures to develop public access plans for their data (Holdren 2013).

The simplest logic for open access for American citizens is that those who paid for the research should be able to use it (WH 2012). However, the value of data extends far beyond that of a commodity. The 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 just to increase innovation, but to increase the visibility of and opportunities for emerging country researchers (DFID 2013). Closed and paid access favors researchers who have the resources to pay for costly subscriptions. Unrestricted access to 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 the research to developing country residents. Furthermore, it reduces the costs of doing research, improving the efficiency of research utilization. Finally, it is only logical and fair that data collected in a developing country be available to its citizens. In this regard, data researchers should not be viewed as data owners, but data custodians, the owners being the subjects from whom the information was collected. This the reason why, for any data collection involving human subjects, the Internal Review Board of each Institution needs to enforce the submission of an informed

consent to data collection and treatment that each respondent needs to be aware of, agree upon, and sign.²

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:

“Data and databases. Data (and any relevant data collection and analysis tools) shall, subject to any additional donor requirements, be deposited in a suitable repository and made Open Access as soon as possible and in any event within 12 months of completion of the data collection or appropriate project milestone, or within 6 months of publication of the information products underpinned by that data, whichever is sooner. Data deposited shall be prepared in a manner consistent with the aims of this Policy. Existing and future databases shall be made Open Access.”

In early 2014, <http://cgiar.org/open> was set up to serve as an open online repository of tools and best practices to assist Centers in integrating the Policy into their operations, as well as the future home of aggregated and/or harvested open information products.

M&E Open-Access Web-Site

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.”

Following the USAID and CGIAR Policies, the Africa RISING M&E team has developed a data and knowledge management platform to leverage data assets and tools produced and maintained by project partners (e.g., ILRI ImpactLite and SLATE, HarvestChoice’s 600+ SSA data layers www.harvestchoice.org).

The platform will serve four key purposes:

1. Provide implementation partners with a secure, web-based data storage and documentation repository that over time constitutes a major Africa RISING knowledge pool supporting further discovery, integration and analysis;

² For example, this was a strict rule under which the IFPRI Africa RISING Baseline Survey (ARBES) data have been collected and distributed. For the exact consent form read and submitted to the respondents, see the ARBES survey instruments for each country available from the IFPRI M&E team upon request.

2. Provide a set of procedures to capture, validate, and integrate indicators, which can generate periodic monitoring reports on indicators agreed with Africa RISING partners (USAID, CGIAR CRPs and Centers, and other national and transnational partners)³;
3. Provide a live repository for non-indicator variables that are used to provide baselines, context and input variables to inform systems modeling and evaluations of interventions intended to support farming systems, post-harvest activities, and market-related activities;
4. Serve as a one-stop, structured and searchable inventory of Africa RISING project and partner organizations, activities, and outputs catalogued in a consistent manner across the entire Africa RISING portfolio, thereby enabling investment and institutional data to be linked to a range of data layers. The platform will include both tabular (e.g., plot, household and community) as well as spatial data and will support management of indicators and other variables as time series (in regular or irregular time series formats).

Africa RISING Baseline Evaluation household and community Surveys (ARBES)

Between the baseline and end-line household interviews, there will be an opportunity to extract and analyze data collected during the baseline survey. In this period, it is expected that descriptive statistics and analyses about individuals, households and communities in the intervention and counterfactual areas will be produced. Both the research and M&E teams may also produce articles, scholarly papers, and reports based on data collected through multiple channels. Additionally, the M&E team will produce interactive maps using spatial data from household and community surveys.

At the end of the Africa RISING project, end-line household and community surveys will be conducted. Data from these surveys will be used to provide evidence about Africa RISING's performance. These data will be used to compare outcomes between intervention and non-intervention households over time. Following the collection, data will be made available for use for legitimate research purposes, including scholarly papers, provided the identification and the confidentiality of information from human subjects is respected.

Conducting standardized farm/household surveys can provide essential data for characterizing the agricultural research process and informing the research agenda. At the same time, survey data can serve as baselines for monitoring research performance at the farm/household level. Baseline, mid-line and end-line surveys on identified indicators (both the common set of standards and – where necessary – custom indicators) are imperative for assessing and monitoring progress during the different project cycles. In project countries, there has been an explosion of baseline farm/household surveys, many of which are not effectively used; most only meet the needs of specific projects. Building strategic partnerships and alliances with multiple stakeholders would be an important step toward tapping into existing data and spreading information possibly helpful to other institutions. In countries where baseline surveys exist and are available, the M&E team in collaboration with the implementers conducted a gap scrutiny to identify the need for complementary baseline surveys. The team developed effective data

³ Wherever possible Africa RISING data will be gathered dynamically from partner-curated data holdings accessed through metadata query and harvesting tools and APIs. Data coding/metadata standards will be adopted/developed and supported by the M&E data and knowledge management support team over the life of the Africa RISING initiative.

structures and systems to cover all the identified indicators. These data structures and systems must be endorsed by the research teams at the project level before their use for data collection.

All ARBES data sets must be accompanied by all data tools including survey instruments and instruction manuals. There must be a description of the sample, including how it was selected, its size and representativeness. Furthermore, the documentation should include a description of how the data collection was implemented and sample weights calculated, if the latter exist.

Project Mapping and Monitoring Tool

The M&E team will also provide data sharing policies, protocols and vehicles for aggregate biophysical and technological performance data. The team has developed an open-access web-site for storing and managing project data, and it will maintain a transparent data analysis platform to serve the needs of stakeholders. The open-access web-site will be a platform of knowledge sharing system for sustainable intensification innovations and information on promising technologies in Sub-Saharan Africa among a large network of stakeholders.

The Africa RISING Project Mapping and Monitoring Tool (PMMT) is ultimately intended to help users understand where and how Africa RISING activities are taking place, and improve project strategies and partnerships for greater impact in their work. Its features and functions have been designed to provide the following benefits:

- **Inform strategic and project management decisions.** The PMMT can help inform decisions by allowing users to take geographic information about AR sites into account, whether it is the location of markets, related projects and partners, travel time, annual precipitation, or maize crop yields.
- **Communicate programmatic projects to key stakeholders.** A primary benefit to users of the PMMT is to see the spatial layout of AR activities relative to geographic context. Users have the ability to add their projects to the PMMT database and then visualize those projects in a variety of ways.
- **Understand how programmatic efforts relate to other projects as well as to useful agricultural information.** Users have the ability to browse and map other people's projects alone and alongside their own projects. This functionality provides the framework for multiple organizations to communicate vital strategic information together in a coordinated fashion.

The Africa Rising PMMT is composed of two functional modules which perform specific and complementary functions:

- **Mapping Application** – This allows users to contextualize where AR activities are taking place and view data related to them
- **Data Entry Application** – Users with the appropriate credentials can add additional data to the PMMT through an intuitive, step-by-step interface

The African Rising PMMT 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 statistics that exist in the PMMT.
- **Editor** – An editor has all the rights that a viewer does. In addition, an editor can make changes to values in reports that exist in the PMMT.

Administrator – In addition to all the rights that an editor has, an administrator can create new reports and add or change other users’ access rights.

The PMMT contains two primary data types: program data and contextual reference data. These two types of data can be used together to create informative, interactive maps. A third type, external data, are suitable for advanced uses where existing data sources from outside the PMMT database can be “mashed up” onto a map, although the treatment of this type of data into the PMMT is still a matter of discussion among Africa RISING project managers due to its implementation difficulty and expected use. The intermediate solution adopted so far is to use the ILRI-CKAN web service (accessible through a prominent link on the PMMT website), although this decision still needs to be agreed upon and ratified by Africa RISING stakeholders.

Initially, the PMMT will be housed at HarvestChoice, given the complexity of the functionalities involved, and in the short-term the ILRI-CKAN web repository has been identified to house researchers’ data for reference and use (see Chapter 4 on “Ownership and diffusion”).

The PMMT has undergone through two waves of enhancements, based on feedback received at the 2013 annual expert M&E meeting. The tool is constantly being updated and refined, and the M&E team, through its local M&E coordinators, periodically organizes in-country PMMT training for all research teams in each mega-site. In addition, the M&E team developed a detailed PMMT training manual and video tutorial to be used at these trainings and beyond, for ease of reference. The HarvestChoice/Africa RISING M&E team is also expected to assist and manage operation, representation, data flow, and research outputs of Africa RISING. Its tasks are to:

- Oversee the quality and relevance of Africa RISING M&E project knowledge products, including data and tools
- Organize and summarize large amounts of technical documents
- Perform a range of program coordination and data management activities
- Build and implement data collection plans
- Develop and manage project-related presentation materials, and respond to internal as well as external project information requests.

Africa RISING Program Data

Behind the PMMT is a database that stores detailed project information. This project information describes what activities are taking place and where. The PMMT database contains information about the PMMT user’s projects, other projects and external layers, so connections can be made on the ground.

Key Concept: Action

In the PMMT, projects are comprised of actions. While the Project is described by basic and largely unchanging attributes, actions are described in greater detail in the dimensions of location and time. Put another way, Project details are those that were most likely provided as part of a grant application. Actions are aspects of a project that may have been described when the project was conceived but should represent the work in its current state after final implementation plans and adjustments made in response to ongoing challenges and opportunities.

Project data include specific project attributes like name, description, grant amount, duration, their actions and the locations where these actions occur. Each action represents a programmatic activity of the given project. It has basic attributes similar to those of the project but specific to the action. What actions have, and projects do not, are on-the-ground field locations.

Locations illustrate where actions are taking place and may represent a variety of things, ranging from training locations to chilling plants to development sites.

Locations are represented as points or areas (polygons). These are drawn by the designated user or users, uploaded in batches (points only) or taken from the PMMT database of known places and administrative district boundaries. Contact and partner information are also stored as part of the project data.

Reference Data

The PMMT contains reference data to help users put project data into context. Contextual reference data consist of reference data from international institutions involved in global development (CGIAR, FAO). The content spans a wide variety of topics, including demographics, crop suitability and market characteristics, mostly drawn from HarvestChoice. These layers are provided with descriptive metadata so that users can understand how to use the data appropriately. Additional layers and categories of data are planned based on availability and user feedback. Specific requests for reference data can be made either through the feedback tool or by a support request to the IFPRI-HarvestChoice team.

4. Ownership and Diffusion

Again, following the USAID DRAFT Research Policy (February 2014) “Results of research (including published and unpublished manuscripts, and their underlying data) for which 50 percent or more of its funding comes from USAID, once secured of any concerns related to privacy, security, or other principled exceptions, must be made publicly accessible online for search, retrieval, analysis, and application.[...]”

After collection, cleaning, transformation of data, and before production of relevant publication, it is acceptable that data remain in the sole custody of the data collector for a period of one year. In fact, the USAID Research Policy states that “As noted previously, federally funded scientific research is subject to a more rigorous peer review process than other types of data before it can be made available to the public. To facilitate this review process and in recognition of time lags associated with publication in peer reviewed journals, research data may be temporarily withheld from public release (embargoed) pending completion of the review process. Specifically, research results may be embargoed for up to twelve months following the end of the award.”

As such, teams or institutions working in concert to collect data must share them in a timely manner to be set-up at the start of their collaboration. Data sharing needs to occur through electronic form and following certain standard and requirements. Minimum feature requirements for Africa RISING open-data platform were identified as follows:

- **File storing:** capability to host physical data files, but also links to external sources.
- **Metadata management:** ability to specify different types of content and to use different metadata templates according to data types (survey data, modeled data, spatial data, etc.).
- **Versioning:** keeping track of history of both data files and metadata with option to view/restore older versions.
- **Permissioning:** ability to control access to individual items based on user roles/groups; supporting OAuth for user authentication/authorization; ability for end-users to set group access to individual or collections of items.
- **Openness:** indexing feeds from external systems (in RDF or DDI format) and publishing resources as RDF feeds (open API and embeddable widgets desirable).
- **Visualization:** ability to view the data file structure (fields and field types, row count for the most common file formats -CSV, TSV, DTA, XLS, DBF-); capability to visualize the actual data as table and/or charts desirable; filtering and sorting data as well as visualization of spatial data (vector and raster) desired.
- **Social features:** users' ability to rate and comment on items, possibly through third-party discussion applications, such as Disqus, Facebook, Twitter.
- **Performance:** adequate performance over low-bandwidth connections, and on smart phones and tablet devices.
- **Analytics:** capability to track usage per item and file over time (views and downloads from non-authenticated users).

Given the above requirements, for Africa RISING it was decided on a trial basis that, once the data have been collected and curated, they will be uploaded onto the web-based ILRI-CKAN at <http://data.ilri.org/portal/> for internal diffusion and use. Data and metadata uploaded onto CKAN will be compliant to the Open Access Policy (this feature is already embedded into CKAN so there is no extra work for the user-side). Thus, data and metadata can be displayed into external platforms (like global data management platforms) or into Africa RISING Project Monitoring and Mapping Tool (PMMT).

4.1 CKAN platform

The Comprehensive Knowledge Archive Network (CKAN) is a web-based open source data management system for the storage and distribution of data, such as spreadsheets and content of databases. Its code base is maintained by the Open Knowledge Foundation (OKF), and the

system is available as a publicly hosted platform at <http://thedatahub.org/> or may be installed locally.

The screenshot shows the Data Hub interface for a dataset titled "Kenya Budgetary Data". At the top, there is a navigation bar with the Data Hub logo, the tagline "the Data Hub — The easy way to get, use and share data", and links for "Search", "Groups", "About", "Login", and "Register". A search box labeled "Find datasets" is also present. Below the navigation bar, the dataset title "Kenya Budgetary Data" is displayed. A secondary navigation bar includes "View", "Resources (0)", "Related (0)", "History", and "Followers (0)". A descriptive text box states: "The Kenyan Treasury site with limited information on expenditure and grant subsidies (currently in non-machine readable format)". The "Resources" section is empty. The "Additional Information" section contains a table with the following data:

Field	Value
Source	http://www.treasury.go.ke/index.php?option=com_docman&task=cat_view&gid=110&Itemid=54
Author	Author not given
Maintainer	Maintainer not given

Below the table, there is a "Cite this" section with the text: "Kenya Budgetary Data . No author. Retrieved 09:39, Jun 17, 2013 (UTC). the Data Hub". On the right side, there is a sidebar with "Views: 7", "License: License Not Specified", and "Tags" including "kenya" and "opending".

Feature Evaluation

- (+) The public platform at <http://thedatahub.org/> is already configured and may be used to upload, distribute, and visualize data.
- (+) CKAN is highly customizable. Currently there are several features implemented by the CKAN community under the form of CKAN extensions; additional extensions can be built in order to implement specific features.
- (+) Customizable look and feel. CKAN is easy to customize using own CSS and images.
- (+) Additional fields or custom vocabularies may be added either by using already built extensions or by developing custom additions depending on requirements.

For example, ckanext-spatial is an extension adding geospatial capabilities to CKAN such as:

- o Spatial model for CKAN datasets and automatic geo-indexing
- o Spatial search integration and API call
- o Spatial search widget integrated on the search form
- o Dataset extent map widget showing dataset extent
- o Web Map Service (WMS) previewer
- o Basic CSW Server to server metadata from the CKAN instance
- o GEMINI Harvesters for importing INSPIRE-style metadata
- o Harvest Metadata API to view the harvested metadata XML either as a raw file or styled to view in a web browser.

(+) Visualization. CKAN has previewing tools that can easily and nicely display data as tables, graphs, maps or images. Own data previewing tools can be plugged into the API to create custom visuals on the fly (see examples at <http://tinyurl.com/c72xhpw>). Data can also be sorted and filtered.

(+) Social features. Users can add comments and discuss datasets, share and promote content using social media platforms (Google+, Twitter, Facebook), create RSS/Atom feeds, “follow” a dataset, flag as “to do” incomplete or faulty datasets.

(+) Performance. The platform may be mirrored on servers located in regions with low average bandwidth for optimization. However, performance is not guaranteed if the CKAN public platform is used.

(+) Analytics. CKAN Google Analytics Extension is built as a user-friendly reporting interface to display Google Analytics statistics, including file downloads.

(+) All the uploaded datasets and resources are available via a data API with powerful query support.

(0) The main content type is a dataset consisting of resources (data files) and cataloging information describing the data. Within the default CKAN instance the resources are defined by: Resource (Link to a file/Link to an API/Uploaded file), Name, Description, Format; the cataloging information consists of: Title, Description, Tags, License, Visibility, Author, Maintainer, Group, tree sets of Custom Fields (key-value pair), Resource (Link to a file/Link to an API/Uploaded file, Name, Description, Format)

(0) Versioning. CKAN uses the Open Knowledge Foundation’s Versioned Domain Model (VDM) to keep a complete history of all edits and versions of dataset metadata. Using this feature, it is possible to look at old versions, undo changes, look at a complete history of changes to a dataset, and compare different revisions. However, it does not store history of changes of the resources (data files).

(0) Permissioning. Datasets may be defined as public, private or associated to preset groups.

(-) If customization is needed, CKAN must be installed locally using managed storage resources (server, DNS hosting). The installation process is platform-restrictive (package install requires Ubuntu 10.4, but source install works with any other operating system). The process of installing, configuring and upgrading the CKAN platform in order to meet requirements needs substantial developer’s time.

Useful Links

- o <http://thedatahub.org/> - CKAN open data platform
- o <http://data.gov.uk/> - Open Data portal of the UK Government
- o <http://publicdata.eu/> - Europe Public’s Data
- o <http://iatiregistry.org/> - IATI Registry

Cost

There is no cost claimed by CKAN if the instance is to be hosted on an internally managed server. However, developer time has to be factored in order to install, configure and maintain a local CKAN instance. A customized data portal meeting requirements is available from the CKAN team starting from \$2,000/month. More details at <http://ckan.org/solutions/pricing/>

Comparison with other platforms

Table 4 reports the scoring summary across platforms.

Table 4. Scoring

Features	CKAN 1 - 5	Dataverse 1 - 5	Socrata 1 - 5	Knoema 1 - 5	DKAN 1 - 5	DevInfo 1 - 5	GeoPortal 1 - 5	DataPublic 1 - 5	Weight 1 - 5
File storing	5	5	5	5	5	5	5	5	3
Metadata mgmt	4	2	5	2	5	2	5	5	5
Versioning	4	3	5	5	5	4		5	4
Permissioning	4	5	5	5	5	4		5	3
Openness									3
Visualization	5	2	5	4	5	5		4	5
Social features	5	4	5	5	5	4		4	3
Bandwidth	4	4	4	3	3	4		4	3
Analytics	5	4	5	4	4	4		4	4
Pricing	4	5	1	5	5	2	5	5	5
Extensibility	5	2	3	2	5	1	4	5	5
Total (out of 190)	180	137	167	155	190	133	85	185	

Note: The grading above reflects the platform's level suitability to the requirements, 5 being the most suitable. It should not be considered as a qualitative evaluation of the platform itself.

4.2 Data Diffusion

Raw and processed data from each Africa RISING team will be available on the ILRI-CKAN website. Data will be freely available right after the embargo period under the open data access policy. Before this date, data will be available to the requestor provided the data provider grants electronic authorization, and separate requests are submitted with name and role of the requestor/user, as well as the intended use of the data. Anonimization needs to be ensured prior to data publication on the ILRI-CKAN platform. Following data publication or one year from data collection, all data transfers (both among Africa RISING teams and with external partners) must be accompanied by metadata. In all cases, the team (hence, not the individual scientist only) that collected the data retains first publishing rights for scholarly articles (peer-reviewed or not) for a period of **12 months after data collection and cleaning**.

The data collected will remain in the custody of the provider institution(s)/scientist(s) for cleaning, initial analysis, and processing. The institution(s)/scientist(s) will submit and upload the data (in raw form, not aggregated) and the associated metadata onto the ILRI-CKAN platform. The present document has been agreed upon and approved by the two AR CGIAR leading agencies for the three mega-sites (IITA and ILRI), IFPRI, and after consultation with the donor

(USAID). Therefore, all staff (and researchers more specifically) are expected to comply with the provisions and rules of this plan. Data can be downloaded by all Africa RISING researchers upon submission of an electronic request form to the data provider. Once received the request, the data provider can grant or deny access to the requested data, which would have been pre-uploaded onto CKAN. The reason(s) of an eventual denial can be appealed by the requestor, bringing the case to the Africa RISING Program Coordination Team (PCT) in the event of unresolved disputes. Data might also be accessed by external users during the embargo period, and subject to the usual permission of the data provider, although in this case no instance can be brought to the Africa RISING PCT should unresolved dispute arise in case of access denial. In general, public users will be able to access only the metadata during the “embargo” period. The latter will cease automatically upon reaching the time elapsed, since each dataset will be associated a publication date from which the embargo time will be calculated.

Consistently with the CGIAR Policy (2013a), the following embargo period rules for Open Access are reported below:

- Peer-reviewed journal articles: **as soon as possible**, ideally at the time of publication, and **no later than 6 months** from the date of publication;
- Reports and other papers (conference papers, policy briefs and working papers): **as soon as possible** and in any event **within 3 months** of their completion;
- Books and book chapters: **as soon as possible** after publication and in any event **within 6 months**;
- Data and databases: **as soon as possible** and in any event **within 12 months** of completion of the data collection or appropriate project milestone, or **within 6 months** of publication of the information products underpinned by that data, whichever is sooner;
- Video, audio and images: **within 3 months** of their completion;
- Computer software: if developed internally, the associated source code must be deposited in a free/open software archive **upon completion of the software development**, and access may be granted subject to appropriate licenses (e.g. Copyleft);
- Metadata: it must be deposited in a suitable repository **before or on publication** of the information product.

Requests for data containing personal or geographic identifiers suitable to identify or contact any individual from whom the dataset was created must not be approved, except for AR programmatic purposes and subject to approval by the data provider, even after the embargo period. All data requests must be addressed in writing or electronically to the data provider in Africa RISING (see Appendix B for the data sharing agreement). All shared data must be accompanied with appropriate metadata and must be labeled properly, to enable the requestor to work on the data independently.

The provisions above discipline data the Africa RISING program will generate in the course of its lifetime. In case of past data already collected at the date of **November 30, 2014**, they should be uploaded onto the CKAN web-based platform no later than **April 30, 2015**. A list of researchers and Institutions not complying with this provision will be kept by the PCT.

4.3 Discretionary External Diffusion

Within the embargo period, 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, subject to the eight conditions reported in Appendix B. After the embargo period, all data are public access and must be shared under the open data access policy rules.

4.4 Acknowledgment

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. The preferred citation method is Creator. Publication Year. Title. Identifier.

5. Security

Storage

All original data must be stored by the creator/s for a minimum period of four years. 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. The identities of individuals with access to the storage space must be provided by the partner institutions (ILRI, IITA, IFPRI) to USAID once a year. Electronic data with personal identifiers must have at least one level of security. The data may be protected by a password known only to authorized users. It may also be encrypted, with decryption keys communicated only to creator/s. A list of parties with access to data including personal identifiers must be provided to USAID.

Transmission

CGIAR or USAID provided email is an acceptable medium through which to transfer digital files. No third party email addresses should be used unless belonging to collaborators in data creators. In cases where digital files are too large to be shared through email, data creators and users could use file sharing serviced, 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 DHS, Fedex, or UPS where available, as the quality, timeliness, and confidentiality of service among public mail carriers is highly variable in the base countries of Africa RISING partners.

Appendix A. Compendium of Africa RISING Tools

	Tool	Description	Where applied?				Experiences/Notes
			WA	ESA	EH	IFPRI	
Participatory Methods	Innovation platforms (at various scales)	community, district, regional, national, etc.)			x		
	Mother-baby trials		x	x			
	Participatory community analysis (PRAE, PRA, RRA, village mapping)		x		x		EH: prioritized livelihoods; different challenges/opportunities for different farmers; identify actors)
	Participatory technology D??		x				
	Preference ranking/pairwise ranking			x			
	participatory action research			x			
	Matrix ranking (feed)	assesses availability, prioritization of feed resources		x			
	Reflection workshops	get farmer feedback on different technologies/iterative feedback into project planning			x		
	Focus Group Discussion	identify priorities as precursor to formation of R4D platforms		x			
Participatory resource mapping	builds on SLATE tool, maps natural assets			x			
Data Collection Tools	FEAST (Feed Assessment Tool)	Documents constraints & opportunities in livestock feeding	x	x	x		(ESA: found livestock underfex, get 60% maximum)
	Impact LITE	Baseline survey or information generation tool	x		x		
	Biomass assessment tool		x				
	Land Degradation Surveillance Framework (LDSF)	measures biophysical condition of land		x			
	Techfit (Technology Filter Tool)	Technology prioritization and selection tool		x	x		e.g. EH - use to match feed recommendations to particular environment
	Nutrition Behavior Change Communication		x				
	Community survey tool (Malawi)	generates community-level data		x			
	AKTS (Agroecological KGE Tool kits)	based on farmers , characterization of different			x		conducted by ICRAF with participatory community analysis
	SLATE (Sustainable Livelihood Asset Tool Evaluation)	Identify indicators for various asset classes			x		
ARBES	household level socioeconomic survey tool		x				
Farm system modeling /FarmDESIGN	(Wageningen)	x	x				
Africa RISING baseline survey	Implemented by IFPRI across AR sites	x	x	x	x		
GIS tools (disaggregate)	Geospatial analytical approach applied to a		x	x	x	e.g. EH's defunct quick-win on technology targeting	
Approaches	value chain mapping	identify major actors in the value chain			x		
	value chain analysis	identify the role of various actors in the value	x		x		
	Community seed production		x				
	Integrated Soil Fertility Management			x			
	Transect mapping	used to differentiate different areas during site				x	
	Integrated Pest Management			x			
Randomized Controlled Trials (RCT)			x				
Communication Tools	Communications tools (Wiki, blog, slideshare, flickr)	Internal project management/knowledge sharing tools		x			
	Participatory video	documenting farmer learning process			x		
	Media: TV news & features		x	x			
	Media: Radio			x			
	Media: Local newspaper			x			
	Media: farmer training videos		x				ESA: maize lethal necrosis disease awareness, micro-dosing benefits when using
	hands-on demonstrations/demonstration plots		x		x		
	Farmer-to-farmer extension						
	Farmer field days		x	x	x		
engage politicians during field days			x				

Appendix B. Data Sharing Agreement

DATA SHARING AGREEMENT BETWEEN [INSTITUTION]

AND

[DATA REQUESTOR]

Person(s) Requesting Data/:

Institutional Affiliation:

Business Address:

E-Mail Address:

Purpose of Data Request:

(Attach copy of Study Outline):

Dataset(s) and Module(s) Requested

Conditions for Data Release:

1. The data shall not be used for any other purpose than the above specified request.
2. The data shall not be copied or released to any other person or organisation, directly or indirectly, without the written consent of [INSTITUTION]. [In the case of IFPRI data, send request to IFPRI-Data@CGIAR.ORG]
3. Draft copy of any major publication (journal article, research monograph, report) arising from the use of the data shall be submitted to [INSTITUTION] for comments and clearance before final publication.
4. Any publication from the data shall acknowledge the contribution of [INSTITUTION]. The data should be properly cited and referenced as mentioned in the documentation manual.
5. The data shall be used for the sole purpose of research and data analysis to further understanding about the economy, society, people and in the advancement of science. Any commercial use is strictly prohibited.
6. The collaboration with professional staff in [INSTITUTION] on data analysis and specialized output, especially in the form of co-authorship of joint research publication, is strongly recommended and advocated.

7. All costs involved in making the data available shall be borne by the Requestor.
8. The dataset shall not be used to identify or contact any individual from whom the dataset was created.

I/We agree that all data released under this agreement are confidential and remain the property of [INSTITUTION] and USAID. I/We undertake to observe the terms and conditions specified above, knowing that if I/We default, necessary legal action may be taken against me/us. This agreement shall be effective as of [DATE] and shall continue until the agreement is terminated in accordance with the provisions made below, but in any event no later than [DATE].

The [INSTITUTION] may terminate this agreement:

- Immediately if a finding or stipulation that the data requestor has violated any standard or requirement of any security or privacy laws is made in any administrative or civil proceeding in which the requestor has been joined;
- Immediately if [INSTITUTION] determine that the requestor has breached or violated a material term of this agreement.

Signature:

Name:

Date:

PERSON(S) REQUESTING DATA

Signature:

Name: Point of contact at [INSTITUTION]

Date:

References

- AEA (American Economic Association). 2014. "The *American Economic Review*: Data Availability Policy." Accessed April 11, 2014. <http://www.aeaweb.org/aer/data.php>.
- CGIAR (Consultative Group on International Agricultural Research). 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 (Consultative Group on International Agricultural Research). 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 (United States Agency for International Development). 2014. USAID DRAFT Research Policy –Washington, DC