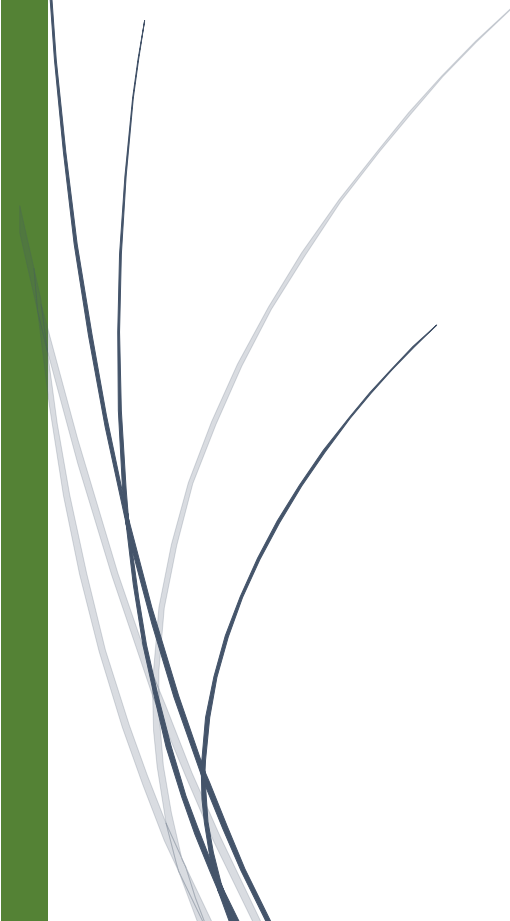3/31/2021

Data Management Plan

Accelerating Impacts of CGIAR Climate Research for Africa (AICCRA)

Technology Integration / Innovation and Business Development Unit
ALLIANCE BETWEEN BIOVERSITY AND CIAT

Contents

Further clarifications to be taken out when the document is finalized:

1. We are not collecting data. We are just providing the mechanism for accomplishing the FAIR. The implementation of the data management plans should happen at cluster level.
2. The cluster level protocol is actually the data management plan because this is the level where the data collection happens.

Introduction

This global data management plan's primary purpose is to help AICCRA support evidence-based decision-making by ensuring that the research data, which is essential for the integrity, quality, and value of research, is appropriately managed at the project level. It also strengthens the culture of results-based project monitoring and evaluation (M&E) by ensuring the reporting of high-quality information products. AICCRA will make these information products available for the long-term to partners and the scientific community following an Open Access Policy. The CGIAR and the Alliance of Bioversity International and CIAT also recommend the implementation of FAIR (Findable, Accessible, Interoperable, and Reusable) principles for each output.

Additionally, this plan also includes establishing data sharing and management processes from legal agreements through research operations and reporting. This "conveyor belt" process will be applied to AICCRA through the MIS reporting system (MARLO). The system identifies the generated data and information products and verifies that products are made publicly available within timeframes agreed upon contractually with partners. Simultaneously, these processes will guide and support scientists and CGIAR Centers to facilitate the production of well-managed and documented datasets that are easy to use both now and in the future.

AICCRA aims to provide a centralized place to collect all information products generated by its research activities. It expects to attract contributions into the platform from scientists working in related thematic areas even if they are not directly managed or funded by AICCRA. It is expected that this will increase accessibility and visibility of scientific outputs to a global community for adding even more value to the products of research with development outcomes in mind.

Main goals:
- Support evidence
- Ensure proper dissemination of products
- Provide long-term storage for information products
- Collect standardized data

Overall Plan

This document will consider two data management planning levels, the project level, led by the AICCRA Management Unit, and the cluster level, led by CGIAR Centers and Thematic/Regional/Country leaders).

At the project level, AICCRA will follow the "CGIAR Open Access and Data Management Policy (OADM)" as the basis of establishing principles and agreements concerning data generated as well as considering the resources that will be needed. Clear planning and reporting schedules will support the principles and agreements and will facilitate collecting, aggregating, and synthesizing research outputs produced during the year.

At the cluster level, the "Data Management Plan" details the specific procedures that leaders will carry out to put the OADM policy into operation. Each research cluster leader should have a data management plan. The Research Leader has responsibility for drawing up and ensuring implementation of the overall policy, while the Data Managers are responsible for cluster-level plans and procedures.

Each cluster-level plan should then detail the steps that will be taken to prepare the data for archiving, including anonymization, producing the data dictionary, etc. The plan will also say who is responsible for each task.

Data coming from the clusters will be collected and entered into the M&E system regularly as per the agreed planning and reporting processes. The quality of the metadata reported by each information product will be checked and assured for quality assurance by the AICCRA Regional Coordinators, Thematic Leaders and the Management committee before being consolidated, synthesized and released by the Cluster Leaders twice in the first year and annually thereafter.

Principles & Agreements

As mentioned before, one of the main goals of this AICCRA Data Management Plan is that the information and data products can be available for long-term use by partners and the scientific community. Therefore, the following principles will be adopted:

- Easy to implement
- Research Outputs should attain highest possible diffusion
- Adherence to international and CGIAR standards for data interoperability
- Archiving, preservation and maintainability
- Ethical use and sharing of personal and private data
- Information products to be **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable (FAIR Principles)

A clear process for data sharing and management must be established, from **legal agreements** through to **operating and reporting processes**, considering the points below:

a. **Annual Project Plans:** The Annual Plan of Work and Budget (APWB) process will take place in the adopted M&E System called **MARLO-AICCRA,** which stands for a customized version of **M**anaging **A**gricultural **R**esearch for **L**earning and **O**utcomes.
  - **Formalization:** Once the planning process has finalized, the details of the annual activities/deliverables will be accessed through the online system and its Business Intelligence Dashboard. In addition, will be then formalized through **Partnership Performance Agreements (PPA)** established with CGIAR Centers and other strategic partners to stipulate that data is to be made freely available and set up the timeframes for data publishing by scientists involved in AICCRA research activities.
b. **Intellectual Asset:** The Project Participants must follow the CGIAR Principles on the Management of Intellectual Assets. The CGIAR IAs include knowledge, databases, publications and other information products; they do not include improved germplasm, plant variety rights and patents. All products produced by AICCRA and Project Participants must be, wherever possible, disseminated following open access principles (see next point).
c. **Open Access:** keeping with the CGIAR Open Access and Data Management Policy (OADM), research data, tools, and associated information generated under AICCRA will continue to be

made available for indexing and interlinking, such that research outputs adhere to FAIR principles (Findable, Accessible, Interoperable, Reusable) to enhance innovation, impact, and uptake.

* Numbers in parenthesis represent months

Data Sharing, Access and Archiving

There are different levels and ways of sharing data and providing access to it. Moreover, this will depend on the applications, tools and platforms that are available within the project and that are developed as a result of the cluster portfolio of information products in the regions and countries. It is expected that each of those platforms should have the proper mechanism to share and archive the data by taking into consideration confidentiality issues and data ownership as per the Open and Intellectual Assets policies described above. The operationalization of best practices for sharing data must be carried out by the country lead organization.

At a project level, and in order to increase the accessibility and visibility of scientific outputs to the global community as well to maintain a standardized metadata information for each of the information products reported by the program, we have available two main repositories which are also broadly adopted by the CGIAR:

1. **CGSpace:** which is a joint repository of several CGIAR centers, research programs and partners. It is a tool to archive, curate, disseminate and permanently preserve research outputs and information products (e.g.: Handle). Currently, the portal is implementing the new CG Core Metadata Schema to ensure greater interoperability across other CGIAR repositories. It is expected that CGSpace can be used to storage agricultural research products and knowledge products.

2. **Dataverse**: which is an open-source web application to share, preserve, cite, explore and analyze research data. Researchers, data authors, publishers, data distributors, and affiliated institutions all receive appropriate credit via a data citation with a persistent identifier (e.g., DOI, or Handle).

At the same time, the monitoring of the information products will be done through each of the deliverables that are part of the AICCRA clusters indicated in the M&E System, as well as a mechanism to measure the FAIR principles. It is expected that this tool also allows mapping the information products, previously disseminated, to the indicators and other components of AICCRA.

3. **MARLO** is an online platform expected to assist AICCRA in the strategic results-based program planning and reporting of research projects. It covers the project cycle from planning to reporting and learning. Reports generated by the system support outcome-focused programmatic reporting with additional synthesis at the Flagship and Regional levels.

Information collected in the M&E System will be used in external websites and any other knowledge sharing platforms. IT is expected that the inputs filled in the system should therefore be complete and suitable for an external audience.

Ethics, Privacy & Licensing

AICCRA respects research participant's and subject's rights and will protect personally identifiable information, sensitive and confidential data. AICCRA will only publish such data in an aggregated, anonymous, or coded form, protecting privacy. Researchers working with such data are encouraged to follow the CGIAR responsible data guidelines and follow the ethical clearing processes of there institutions and countries where the research is being done. AICCRA shall seek and obtain the most appropriate licensing arrangement possible for all information products that reflect this plan's principles and agreements. For data and information deposited into the AICCRA repositories, the default license applied will be the Creative Commons—Attribution License (CC BY 4.0) (https://creativecommons.org/licenses/by/4.0/). Alternative and more restrictive licensing may be sought as an exception if it furthers the research objectives of AICCRA.

Processes & Quality Checks

Using MARLO for AICCRA (https://aiccra.marlo.cgiar.org), the Planning and Reporting processes will take place a number of times in the year depending on the needs to collect progress reports for the World Bank. These processes are built bottom-up and comprise different submission points depending on the level of disaggregation of data and information collected in the system. MARLO is also monitoring the open access properties of deliverables using the FAIR approach.

The development of the M&E System will follow a number of key principles, including a focus on simplicity with the ability to add management components as needed. To eliminate redundancies and confusion, clusters of activities are only added once, by those closest to the source of information, with the goal to add only the information that people will use. The system will allow AICCRA to agree on common standards, and collaborate and share knowledge.

On the granular project components and deliverable details, there will be a couple of quality assurance processes to make sure the information fulfills the program needs. They usually are:
1. Cluster of activity leaders
   Submission of detailed plan/reports (inc. deliverables, contribution to IPIs, etc.)
2. Aggregation and Synthesis at Thematic and AICCRA Regional level
3. Aggregation and Synthesis at Project level
4. Final report
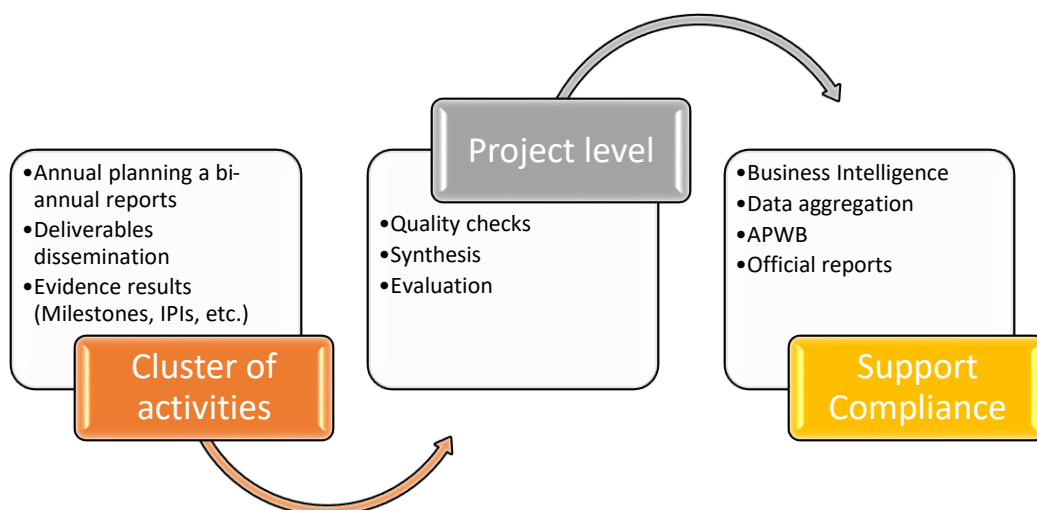
At planning, the cluster of activities set their expected deliverables[1] with a complementary narrative description. And at reporting they are asked to report against what they have planned. It is also expected that the information products are finalized and disseminated when the reporting is done.

ID	Deliverable title	Sub Category	Delivery year	FAIR Compliance	Status
D20205	Combining the effects of increased atmospheric carbon dioxide on protein, iron, and zinc availability and projected...	Journal Article (peer Reviewed)	2019	F A I R	●
D9715	Supplementary data on livestock feed biomass availability developed for the IMPACT Model	Database/dataset/data Documentation	2019	F A I R	●
D19275	Global Foresight for Food and Agriculture Tool	Data Portal/tool/model Code/computer Software	2019	Not applicable	●
D19341	Agricultural investments and hunger in Africa modeling potential contributions to SDG2 – Zero Hunger	Journal Article (peer Reviewed)	2019	F A I R	●
D19342	Gaps between fruit and vegetable production, demand, and recommended consumption at global and national levels	Journal Article (peer Reviewed)	2019	F A I R	●
D19344	The palm oil dilemma: Policy tensions among higher productivity, rising demand, and deforestation	Policy Brief/policy Note/briefing Paper	2019	F A I R	●
D19347	Key determinants of global land-use projections	Journal Article (peer Reviewed)	2019	F A I R	●
D11946	Options for keeping the food system within environmental limits	Journal Article (peer Reviewed)	2018	F A I R	●
D11947	A modelling study on optimal tax levels and associated health impacts	Journal Article (peer Reviewed)	2018	F A I R	●
D12604	Risk of increased food insecurity under stringent global climate change mitigation policy	Journal Article (peer Reviewed)	2018	F A I R	●
D12668	Can Ethiopia feed itself by 2050? Estimating cereal self-sufficiency to 2050	Policy Brief/policy Note/briefing Paper	2018	F A I R	●
D5760	Long-term simulation of the livestock sector through the IMPACT model	Discussion Paper/working Paper/white Paper	2018	F A I R	●
D5320	Development of livestock module for the IMPACT model	Data Portal/tool/model Code/computer Software	2018	Not applicable	●
D12248	Improvement of fish module for IMPACT model	Database/dataset/data Documentation	2018	F A I R	●

Metadata information requested at deliverable level would be:

General information:
- Title
- Description
- Category and Sub-category
- Status (i.e. on-going, complete, extended, cancelled)
- Expected/complete year
- PDO and IPI Indicator
- Funding source
- Geographic scope
- Cross-cutting dimensions on Gender, Youth, Cap-Dev
- Partners contributing

Dissemination and Metadata:
- Disseminated URL (E.g. CGSpace, Dataverse, etc.)
- Open Access (Yes, No)
- Is the journal an ISI publication? (Yes, No)
- License adoption (Yes, No)
- Trainings / CapDev
- Does the publication acknowledge AICCRA? (yes, No).

---

[1] A specific, time-bound, tangible information and knowledge product that is linked to an output. It is proof, in digital, electronic, physical or other kind of soft or hard copy of the completion of a set of activities. Examples of deliverables are: workshop reports, journal articles, datasets, training materials.

Quality Check

For those deliverables that are Databases, Datasets, Data documentation or Maps/Geospatial data, please respond to the 3 questions related to its quality (see screenshot).

Compliance check section guarantees that a data deliverable is Gold Data. If you select 2 out of the 3 questions with yes and documented, and the ranking is over 3.5, it qualifies to be a Gold Data deliverable.

1. Have you had a process of data quality assurance in place?
   (Yes, but not documented; Yes, and documented; No)
2. Di you have a data dictionary?
   (Yes, but not documented; Yes, and documented; No)
3. Are the tools for used for data collection available (e.g. surveys, training materials, etc.)?
   (Yes, but not documented; Yes, and documented; No)

* In order to simplify the reporting processes, the MARLO is also connected to external platforms such Web of Science and Scopus.

Resources & Responsibilities

The CGIAR CCAFS Program, in collaboration with the University of Reading and recently by the Alliance of Bioversity and CIAT, have available a Data Management Support pack designed to help to produce high quality, reusable and open data from your research activities. It consists of documents, templates and videos covering the different aspects of data management and ranging from the overarching concepts and strategies through to the day-to-day activities. For each of the videos in the pack we have included a transcript of the narrative.

Who are you?

- Principal Investigator
- Researcher
- Data Manager
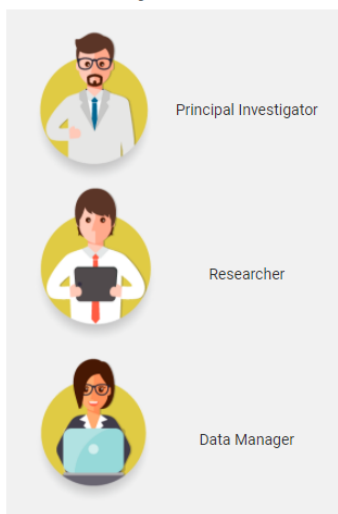
Where are you?

1. **Proposal Stage**
   You are writing a project concept note or proposal for funding consideration
2. **Grant Opening**
   Your project has been funded and you are starting research activities
3. **Project Research**
   You are carrying out activities that will achieve project objectives
4. **Publishing**
   You are sharing the results of your project
5. **Grant Close out**
   The project is coming to an end and you are housekeeping

What

0. Research Protocols
1. Policy Documents
2. Data Ownership and Authorship
3. Planning
4. Data & Document Store
5. Fieldwork
6. Managing Data
7. Metadata
8. Archiving & Sharing

https://ccafs.cgiar.org/data-management-support-pack

In order to ensure a proper implementation of this data management activities, the following list of

actors should be involved within the program:

1. ***Cluster Leaders** (incl. Country, Regional and Thematic leaders)* are expected to:
   - Provide overall leadership to the cluster, make sure there is a data management plan in place, and following the guidelines as per the Cluster Protocol (see section below).
   - Make sure the planning and reporting processes are completed in the system in the schedule proposed.
   - Make sure information products (deliverables) are proper disseminated prior to the reporting cycles and following the FAIR principles.
2. MARLO ***Management Liaisons***:
   - This can be the Project Management Committee (PMC), Thematic leaders or AICCRA Regional Coordinators.
   - A cluster must have always a ML associated
   - Ultimate responsible for the cluster
3. M&E Officer
4. Comms officer
5. Finance Manager
6. *AICCRA Knowledge and Data Sharing team:*
   - Support AICCRA compliance with the CGIAR OA/OD & IA Policies;
   - Ensure organizational awareness about the information products we provide;
   - lead the AICCRA Planning & Reporting Processes;
   - Lead the development of the M&E tool as per AICCRA project needs.
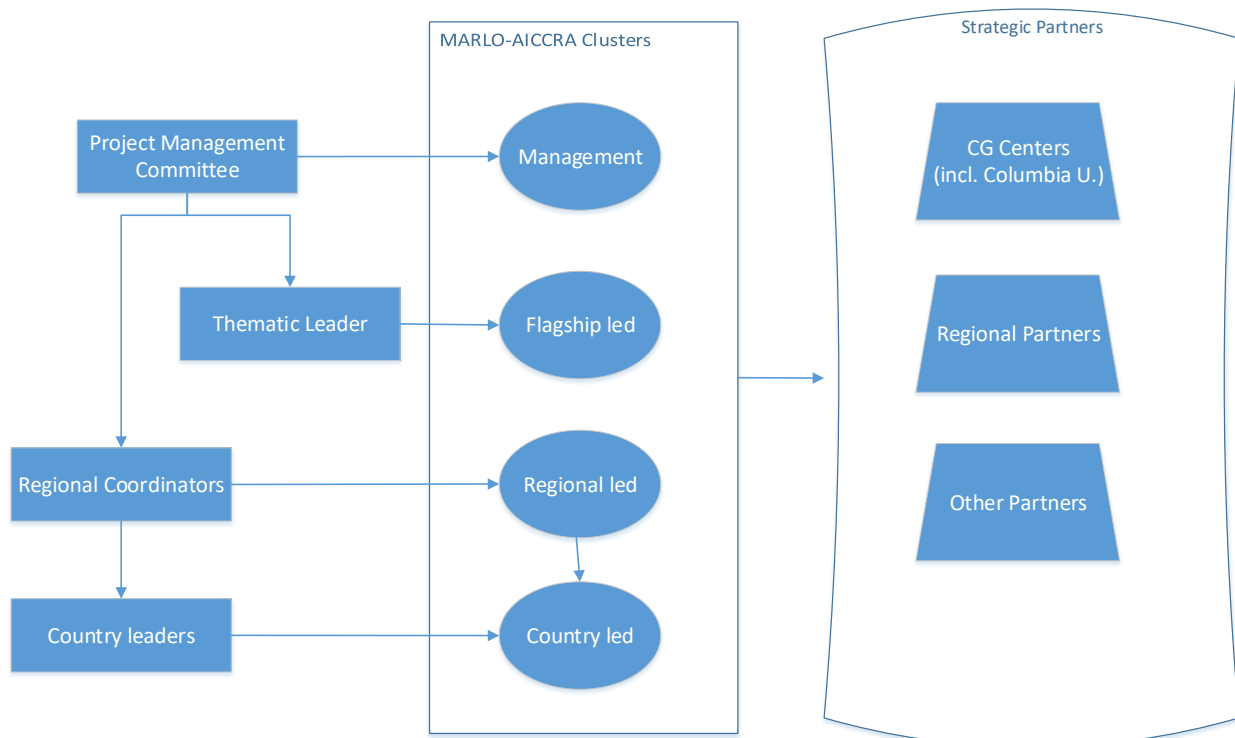
Associate Videos

Videos accompanying the original release of the CCAFS Data Management Support Pack in 2013 are available as a playlist on the Statistical Services Centre YouTube Channel at

https://www.youtube.com/channel/UCs7EU95YMjhvNozJKCD92xQ/playlists. These videos have not been updated since the original release but are mostly still relevant.

In particular the playlist includes a video on Data Management Plans available from the following link: https://www.youtube.com/watch?v=Q8jX_cH0C60&index=3&list=PLK5PktXR1tmNRaUPsFiYlyhg2lui 0xgpj

Data Management protocol at Cluster Level (For Country, Regional and Thematic teams)

The Cluster level data management plan would be drawn up by the data manager or the person with data management responsibilities within the cluster of activity teams. This is a much more detailed document explaining how the principles are going to be achieved. For example, how they intend to set up a data and document storage facility for data sharing among the team.

The plan will naturally vary according to the type of activity but would generally include the following elements:

Do you need a Data Manager?

One of the first decisions to be made is whether a data manager is needed for a cluster. The choice is basically between:

1. Having a specialist data manager within the cluster to whom all data management responsibilities are allocated; or
2. Your Institutions data management team assumes the responsibilities for data management in the cluster or;
3. Allocating data management responsibilities to scientists and/or existing cluster staff

The decision depends on the size and complexity of the cluster and the skills of the scientists, as well as what the CGIAR Centers have the capacity to offer.

> *Most clusters that involve a team of scientists (as opposed to a single researcher) are likely to need a data manager.*

We would recommend the inclusion of a data manager in most of the country and regional teams. This may correspond to the allocation of data management responsibilities to an existing member of the team who has the relevant skills, time and inclination to do the job well, or may involve recruiting a new member of staff with the relevant skills.

The following section contains a list of considerations for a data management plan.

DATA MANAGEMENT PLAN TEMPLATE

(Revised in 2020)

Data Management and Sharing Plan

1. Basic Project Information

Please complete the form below with the relevant fields applicable for your project. For multiple items e.g. multiple funders please separate the entries with a comma (,).

Name of Project:	
Project Code:	
Principal Investigator	
Data Manager (if any):	
Duration of project:	
Aims & Purpose of the project:	
Producer (Abbreviation):	
Funder(s):	
Partner Institutions:	

2. Data production activities

Briefly describe (what, when, where, how, how much) data will be produced by this project. Include information such as; the kind of data (new observational data, experimental data, survey etc.), how much (volume) will be produced each year? What existing datasets could you use or build upon? If available list procedures for quality control of the data
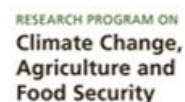
3. Documentation and metadata

Briefly describe how documentation will be handled e.g: How will metadata (contextual details needed to make the data meaningful) be supplied and standards to which will it adhere? Will metadata creation be automated? Include other forms of data documentation such as experimental protocols, methodology reports, laboratory notebooks, questionnaires, codebooks, data dictionaries, database schema and software syntax.

4. Roles and responsibilities

Briefly describe the roles and responsibilities of project staff (PIs, researcher, data managers, partners etc.) for implementing this plan. E.g. who is responsible for obtaining 3rd party data? For capturing data in the field? Producing metadata? Storing and sharing the data?

Alliance

Bioversity International

CIAT
International Center for Tropical Agriculture
Since 1967 Science to cultivate change

CGIAR

RESEARCH PROGRAM ON
Climate Change,
Agriculture and
Food Security

CCAFS

5. Storage, backup and security

State information such us; where the data will be physically stored, how it will be transmitted, how will the data be backed up, how regular will backups be, who is responsible for backups, how will data security be managed, how will permissions, restrictions and embargoes be enforced.

6. Data protection, rights, ownership and access

Kindly state; are there ethical, confidentiality and privacy issues? How will these be resolved? Will the project need IRB approval? Does the project have a data ownership agreement? Is the dataset covered by copyright or database right? Who owns the copyright and other intellectual property? Who will be given access to the data pre-publication or pre-sharing? Kindly contact Alliance legal office.
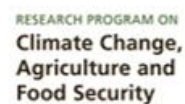
7. Data preservation, sharing and licensing

Briefly describe the long term strategy for maintaining and sharing the data: e.g. on what basis will data be selected for preservation? How long will the data be kept beyond life of the project? How will sensitive data be disposed of or transferred? Is anonymization or de-identification required prior to sharing? Which repository/central database will the data be deposited for sharing and how long will it be retained? What license will be applied to shared data? What related information will be shared?

Alliance