****

**CGIAR Research Program on   
Climate Change, Agriculture and Food Security (CCAFS)**

CCAFS[[1]](#footnote-1) - Data Management Strategy

**August 2012**

**DRAFT**

Contents

[Introduction 2](#_Toc333231251)

[Goal 3](#_Toc333231252)

[Objectives 3](#_Toc333231253)

[Scope 3](#_Toc333231254)

[Supporting mechanisms 3](#_Toc333231255)

[Strategic Elements 3](#_Toc333231256)

[Programme Level 3](#_Toc333231257)

[Centre Level 4](#_Toc333231258)

[Implementation of Strategy 4](#_Toc333231259)

# Introduction

CCAFS is mandated to producing international public goods and has developed this Data Management Strategy (DMS) to enable the programme to fulfil its obligations with respect to making data and the relevant supporting documentation from its research activities available to the global community.

The Program Participant Agreements established with CGIAR centres and other partners stipulate that data is to be made freely available and sets up the time scales for data publishing by scientists involved in CCAFS research activities:

“The Contracted Party agrees to publicly share any data and/or models generated as a result of activities under this Agreement through CCAFS’s data portals as soon as practically possible, but no later than twelve (12) months of generation for meta data and twenty four (24) months for other data and/or models. Such data portals include, but are not limited to, the CCAFS agricultural trial data repository (www.agtrials.org), the Adaptation and Mitigation Knowledge Network (www.amkn.org) and the CCAFS climate data portal (www.ccafs‐climate.org). Access to the data should be fully granted to the CCAFS data manager at CIAT, who will guide the process of adding each dataset to the proper data portal and will provide the necessary help and support.”

The aim of the Data Management Strategy (DMS) is to guide the creation of an enabling environment where scientists and partners are able to produce and share high quality data outputs throughout CCAFS, while at the same time enabling a variety of data management procedures and good practices at project level. This is achieved through creating “data portals” specifically designed for common types of data where scientists can publish their data and by the provision of guidance and support to scientists and CGIAR Centres to facilitate producing well-managed and documented datasets that are easy to use both now and in the future.

Guiding principles for this strategy are: accessibility, ease of use, ethical use and sharing of data about people, provision of support to data generators, ensuring that credit and visibility go to data generators, adherence to international standards for data documentation, curation and storage.

CCAFS aims to providing a “one-stop shop” for data generated by its research activities and expects to attract data contributions from scientists working in related areas even if not directly managed or funded by CCAFS. It will increase accessibility and visibility of scientific outputs to a global community for adding even more value to the products of CCAFS research with development outcomes in mind.

In this strategy we use the term “*Data+*” to indicate the actual data generated by the research process once it has been cleaned and is considered of good quality, as well as the documentation that will enable the use of these datasets in the future. This includes but is not restricted to documents about the methodology for data collection/generation, computer programs used for data manipulation and data processing, data quality assessment, and any metadata that helps in building a description of the context in which the data have originated.

In defining this strategy, we have adopted the following principles:

*It has to be easy to implement and any burden to researchers that is generated from its implementation must be balanced by the benefits that the researcher will get from making his/her data available, and by the support that CCAFS will provide.*

*It should not affect the autonomy of scientists to carry out their research; the strategy ensures the independence and creativity of scientists in the collection of data that is relevant to the CCAFS research objectives.*

# Goal

The goal of this DMS is for CCAFS *Data+* products to be archived and made available for the long-term for use by partners and the scientific community.

# Objectives

The objectives of this strategy are as follows:

1. To make available quality-assured *Data+* to potential users now and well into the future;
2. To encourage appropriate levels of standardization, adoption of international standards and harmonization so that data from separate research activities can be brought together to enrich our understanding of processes, outcomes and impacts in the areas of the world where CCAFS works;
3. To set the path for building a useful, complete and accessible depository of data for future research;
4. To guide CCAFS in designing and implementing support mechanisms to reach the goal.

# Scope

This DMS looks at making *Data+* available in public archives. It does not include research outputs such as papers and publications resulting from analysis of primary data. CCAFS is utilizing alternative portals to share this type of information.

# Supporting mechanisms

Supporting mechanisms will be necessary for the implementation of this strategy. These include:

1. Providing guidelines for making data available in such a way as to respect the trust that information providers have deposited on CCAFS scientists;
2. Creating, maintaining and supporting portals to the data repositories. These portals should enhance visibility of *Data+* provided by researchers with credit clearly attributed to the data generators; support the work of CCAFS PMC; and ensure accessibility to the data through human generated searches and automatic searches by web search engines.

# Strategic Elements

## Programme Level

In order to achieve the objectives set out above, the CCAFS programme needs to:

* Based on Consortium level policies, discuss, define and adopt a data sharing and data ownership policy and Intellectual Property policy;
* Negotiate and coordinate actions with the Consortium Office of CGIAR (CO), as well as CGIAR centres and partners that are part of CCAFS;
* Include the required elements of these policies into the contracts established with centres – e.g. CCAFS Program Participant Agreements;
* Develop and provide a package of supporting materials to centres:
  + Provide guidelines on how data can be accepted into the system;
  + Provide a clear definition of data so that the data management strategy contributes but does not encroach into the management of research outputs;
  + Support projects in achieving good data management, quality assurance and the timely delivery of data;
* Set up and resource mechanisms to receive and archive data in the following data portals;
  + Socio-economic and other data: Dataverse;
  + Trial data: AgTrials;
  + Climate data: CCAFS climate portal;
  + Spatial data: We are currently seeking guidance from the CO;
* Establish a mechanism to promote the existence of the data archive; facilitate accessibility, visibility, and links to web search engines.

## Centre Level

In order to fulfil centre contractual obligations under the PPA agreements, CCAFS expects that centres will pro-actively do the following:

* Allocate sufficient resources to allow for the implementation of the DMS;
* Utilise the provided support package for the implementation of the DMS;
* Submit their *Data+* to the CCAFS appropriate data repository in a timely manner.

# Implementation of Strategy

Three key elements are essential to the implementation of the strategy:

1. Establishing a ***process***

A clear process for data sharing and management must be established, from legal agreements through to operating and reporting principles.  This will require all CCAFS activities to identify data products, and the data information specialist to liaise with the researchers to ensure that the data is generated, managed and made available through appropriate systems.  All data produced in CCAFS should be reported, and the data information specialist should establish a data pipeline, and prioritise the most important products for sharing based on extent, scope, and relevance.

1. Facilitating the ***systems***

For each of the data types identified, the most appropriate system (for instance, the CCAFS portals) established to ensure that data is archived, and shared following the principle of open-access.  The system for each data type should be identified through a robust analysis of different technical options, and implemented by the data information specialist of CCAFS.  Support, clear guidelines and documentation of the system should be established, and shared with all researchers in CCAFS.

1. Enabling a data ***culture***

The hope is that all researchers follow the principles and processes established by CCAFS, but this requires significant cultural shifts with program participants.  The appropriate incentives should be established to promote data sharing across CCAFS.  Metrics on data sharing from each program participant should be used as a criteria for measuring performance, and appropriate incentives put in place to follow up and reward those most effective in managing and sharing the data generated under the CCAFS program.  Researchers and their data technicians participating in CCAFS should also be trained on the use of the systems for making data available with clear user guides to be established by the CCAFS data information specialist.

* Among the conditions to facilitate the establishment of the data culture CCAFS must make it easy to submit data;
* Highlight benefits to researchers derived from data sharing such as increased visibility and reputation;
* Availability of statistics about data downloading and use so as to be able to use this information as a planning tool for the programme to promote CCAFS’s research agenda and that of our scientific partners among the global audience.

*Temporary end note: The support package for the implementation of the DMS is currently under development, and will include planning and reporting templates, a database to facilitate CCAFS’s management and evaluation of submitted Data+, and practical guidance on meta-data, data management, data quality assurance and ownership/authorship. A listing of the implementation guidelines will be available by 5 September 2012, and the full package will be available to Centres in early 2013 so that the DMS can be implemented for at least part of the 2013 planning and reporting cycle and will be fully operational for the 2014 cycle.*

1. For more information on CCAFS, go to: www.ccafs.cgiar.org [↑](#footnote-ref-1)