

Data Citation Implementation @ Dataverse

Mercè Crosas
Chief Data Science and Technology Officer,
IQSS, Harvard University
@mercecrosas

Workshop: Data Citation Pilot Project Kick-off
bioCADDIE supplemental project, NIH Big Data to Knowledge
Feb 3, 2016

Data Citation in Dataverse complies with the **Data Citation Principles**

Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014

Altman, Crosas, The Evolution of Data Citation: From Principles to Implementation, IASSIST Quarterly; 2013

Data Citation Generated by Dataverse

Authors

Published Year

Export Formats
for users

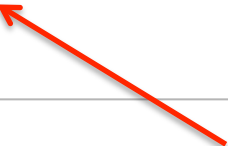
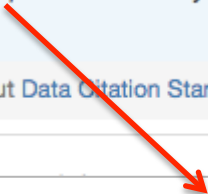
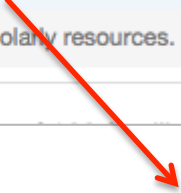
Correlates of Father Participation in Family Work, 1979-1981

Rosalind C. Barnett; Grace K. Baruch, 2007, "Correlates of Father Participation in Family Work, 1979-1981", <http://hdl.handle.net/1902.1/00620>, Harvard Dataverse, V3

If you use these data, please add this citation to your scholarly resources. Learn about [Data Citation Standards](#).

Download Citation ▾

- EndNote XML
- RIS Format

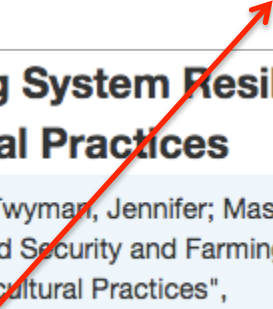
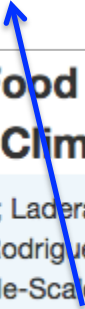


Persistent Identifier:
Handle or DOI

Repository
Name

Dataset Title

Version



Increasing Food Security and Farming System Resilience in East Africa through Wide-Scale Adoption of Climate-Smart Agricultural Practices

Winowiecki, Leigh; Laderach, Peter; Mwongera, Caroline; Twyman, Jennifer; Mashisia, Kelvin; Okolo, Wendy; Eitzinger, Anton; Rodriguez, Beatriz, 2015, "Increasing Food Security and Farming System Resilience in East Africa through Wide-Scale Adoption of Climate-Smart Agricultural Practices", <http://dx.doi.org/10.7910/DVN/28703>, Harvard Dataverse, V7

Download Citation ▾

Persistent Identifier Resolves to Dataset Landing Page



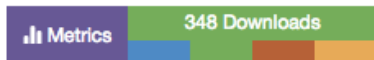
RESEARCH PROGRAM ON
**Climate Change,
Agriculture and
Food Security**



[CCAFS - Climate Change, Agriculture and Food Security Dataverse \(CCAFS\)](#) <http://ccafs.cgiar.org/>

[Harvard Dataverse](#) > [CCAFS - Climate Change, Agriculture and Food Security Dataverse](#) >

Increasing Food Security and Farming System Resilience in East Africa through Wide-Scale Adoption of Climate-Smart Agricultural Practices



Increasing Food Security and Farming System Resilience in East Africa through Wide-Scale Adoption of Climate-Smart Agricultural Practices

Winowiecki, Leigh; Laderach, Peter; Mwongera, Caroline; Twyman, Jennifer; Mashisia, Kelvin; Okolo, Wendy; Eitzinger, Anton; Rodriguez, Beatriz, 2015, "Increasing Food Security and Farming System Resilience in East Africa through Wide-Scale Adoption of Climate-Smart Agricultural Practices", <http://dx.doi.org/10.7910/DVN/28703>, Harvard Dataverse, V7

 Download Citation ▾

If you use these data, please add this citation to your scholarly resources. Learn about [Data Citation Standards](#).

Description	The overall project goal is to improve food security and farming system resilience of smallholder mixed crop-livestock farmers in East Africa while mitigating climate change through wide-scale adoption of climate-smart agriculture (CSA). The project integrates interdisciplinary approaches, including participatory research, integrating a meta-analysis of CSA practices, real-time land and soil health assessments, crop suitability modelling, socio-economic appraisals and multi-dimensional trade-off analyses, as well as on-farm participatory evaluations of CSA to identify, test, implement, and outscale locally appropriate CSA practices.
Subject	Earth and Environmental Sciences; Medicine, Health and Life Sciences; Social Sciences; Other
Keyword	Climate Smart Agriculture, Food Security, Land Health, Soil, Socio-Economic, Adaptation, Mitigation

The Persistent Identifier applies to the entire Dataset, not to individual Files

The screenshot shows a web interface for a dataset. At the top, there are four navigation tabs: "Files", "Metadata", "Terms", and "Versions". The "Files" tab is highlighted with a red circle. Below the tabs is a search bar with the placeholder text "Search this dataset..." and a "Find" button. On the right side, there is a "Download" button. Below the search bar, it says "7 Files". The main content area displays a list of files:

- 0000 Increasing FS and farming system resilience in East Africa through wide-scale adoption of CSA.pdf**
Adobe PDF - 438.8 KB - Apr 23, 2015 - 52 Downloads
MD5: 47a940551eed82c5f5e2e6ed9c698aab;
This document gives a brief description of the project, including the type data collected.
Please download and read this before downloading any other.
00 ReadMe
Download
- 0200 CSA-RA manual V2.pdf**
Adobe PDF - 4.6 MB - Feb 2, 2016 - 0 Downloads
MD5: 9f022e49f891ec6b4a80c7dad9107f1d;
CSA-RA Manual updated including the CSA Prioritization Workshops
02 Manuals
Download
- 0300 CIAT SAGCOT CSA-RA Report.pdf**
Adobe PDF - 1.8 MB - Apr 23, 2015 - 70 Downloads
MD5: 7c8a695de400ea0c40ff3ed6fac34aac;
A Climate Smart Agriculture Rapid Appraisal (CSA-RA) was carried out by CIAT in collaboration with Sokoine University of Agriculture (SUA) for the Southern Agricultural Growth Corridor of Tanzania (SAGCOT) in September 2014. The CSA-RA aimed to assess within and between district variations in farming systems, agricultural management practices, challenges for current agricultural practices, and climate vulnerability, in order to inform targeting of climate smart agriculture (CSA). The CSA-RA used key-informant interviews, participatory workshops, transect walks, farmer interviews, as well as oender-disaacoregated methods to oather information on important

The same Persistent Identifier applies to All Versions of the Dataset

Only major versions (not minor) appear in the generated data citation

Files Metadata Terms Versions

View Differences

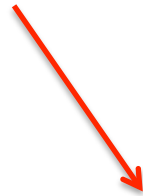
<input type="checkbox"/>	7.0	Files (Added: 2; Removed: 1; Changed File Metadata: 5); View Details	Paola Camargo	February 2, 2016
<input type="checkbox"/>	6.1	Citation Metadata: Contact (1 Changed); Additional Citation Metadata: (); Terms of Use/Access Changed View Details	Cathy Garlick	April 23, 2015
<input type="checkbox"/>	6.0	Files (Added: 2; Removed: 1); View Details	Juliana Muriel Osorio	March 3, 2015
<input type="checkbox"/>	5.0	Files (Changed File Metadata: 3); View Details	Juliana Muriel Osorio	February 18, 2015
<input type="checkbox"/>	4.0	Files (Added: 1); View Details	Juliana Muriel Osorio	January 30, 2015
<input type="checkbox"/>	3.0	Citation Metadata: Author (2 Added); View Details	Juliana Muriel Osorio	January 23, 2015
<input type="checkbox"/>	2.0	Files (Added: 1); View Details	Juliana Muriel Osorio	January 22, 2015
<input type="checkbox"/>	1.0	This is the first published version.	Hector F. Tobon R.	January 16, 2015

Citation for Quantitative (tabular) Data

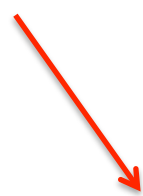
Authors, Published Year, Dataset Title, **Persistent Identifier**,
Repository Name, Version, **Universal Numerical Fingerprint**
(UNF), **[File name]**, **[var 1]**, **[var 2]**, **[var...]**



Checksum
independent
of file format



Specify File in
Dataset



Specify a subset of
variables in Tabular
Data File

Following: Altman, King, A Proposed Standard for the Scholarly Citation of
Quantitative Data, D-Lib, 2007

Dataverse – DataCite Workflow

EZID API

1. Dataset Created in Dataverse
2. Mint DOI with status “reserved” in EZID, send citation metadata
3. Dataset published in Dataverse
4. Change status to “public” in EZID
5. New version of Dataset
6. Send updated citation metadata

DataCite API

1. Dataset Created in Dataverse
2. Reserve local DOI in Dataverse
3. Dataset published in Dataverse
4. Mint DOI in DataCite, send citation metadata
5. New version of Dataset
6. Send updated citation metadata

Additional Metadata in Dataverse

Citation Metadata

- Authors
- Title
- Description
- Dates
- Contact
- Subject
- ...

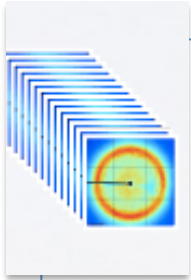
Domain Metadata

- Life Sciences: based on ISA-Tab (and OBI and NCBI taxonomy)
- Other domains (social science, astronomy)

File Metadata

- File header metadata
- File description, type
- Variable metadata

What's Coming Next

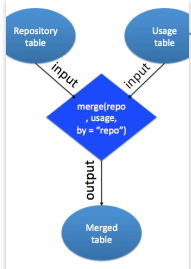


SBGrid Data Repository,
Biomedical Dataverse (Sliz
HMS, Crosas IQSS)

THE LEONA M. AND HARRY B.
HELMSLEY
CHARITABLE TRUST



Social Science Big Data (King,
Crosas at IQSS)



Data Provenance (Seltzer
SEAS, Crosas, King IQSS)



Privacy Tools to share sensitive
data (SEAS, Berkman, Privacy
Lab, IQSS, MIT)



Future Data Citation Extensions

- Provenance Metadata to be used in citation services
- Extended Domain Metadata (e.g., Life Sciences) to be used in citation services
- Support for Privacy, Sensitive Datasets:
 - A DataTag (blue, green, yellow, orange, red, crimson) assigned to each dataset that defines its sensitive level, with security and access requirements
- Support for Large (Streaming) Datasets:
 - Many files per Dataset. E.g., Primary Structure Dataset with thousands of images
 - Large Streaming Dataset. E.g., Geospatial Tweets

Citation for Big Data: Large, Streaming, or Sensitive Datasets

Authors, Published Year, Title, **Persistent Identifier**, Repository Name, Version, [Subset: Query or Variable], [DataTag]

- Be able to cite entire Big Data dataset (with one Persistent Identifier), as well as specify granularity when needed
- Should the query be a RESTful url?
- Should the subset be defined by variable/attributes metadata?
- Should the DataTag be part of the citation for sensitive data?