



International Center for Tropical Agriculture
Since 1967 / *Science to cultivate change*

Research Data Sharing

Webinar

CGIAR Gender and Agriculture
Research Network

29th September 2016

Leroy Mwanzia

Data and Information Manager

lmwanzia@cgiar.org



What: Research Data

Primary research data:

- Is valuable
- Can be expensive and time consuming to generate
- Has value beyond the use of the original research

Why Share Data



Reduces the cost of duplicating data collection



Increases the impact and visibility of research



Recognition & attribution: Can provide a direct credit to the researcher as a research output in its own right



Promotes the research that created the data and demonstrates continued use and relevance of the research

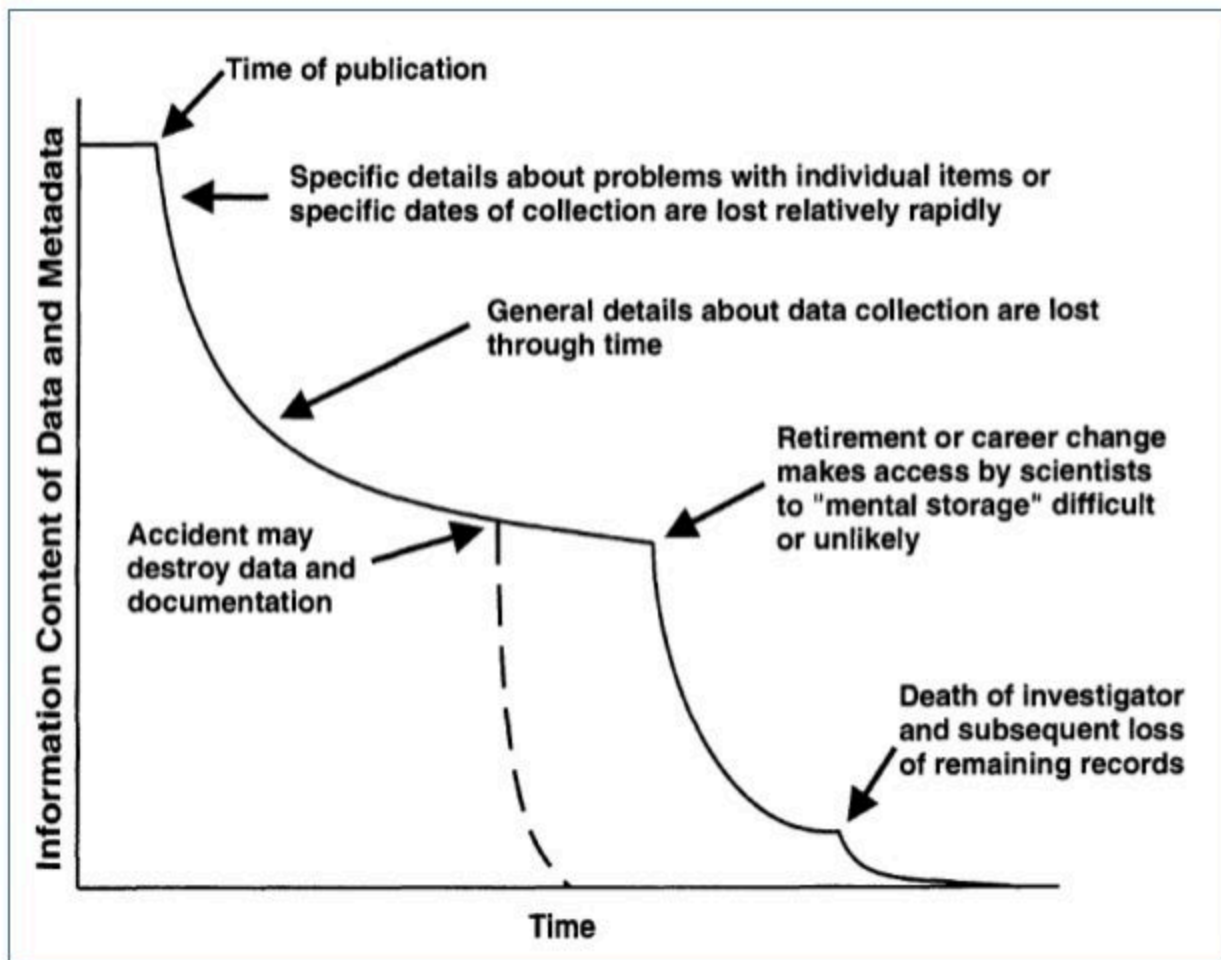


Facilitating science: discovery and access reinforces open scientific enquiry and can lead to new unanticipated discoveries.



Provides important resources for education and training

Why Share? Eradicate Information Entropy



Information Entropy – Demand for 20 year old data

Sent: Tuesday, September 13, 2011 9:17 PM

To: _____ (ICRAF)

Subject: Request for Ethiopia Price Data

Dear _____

My name _____ with USAID's Famine Early Warning Systems Network (FEWS NET). I am writing to inquire about the food price data from Ethiopia that you used in the paper titled "Marketing of Maize and Teff in Western Ethiopia: Implications for Policy Following Market Liberalization" published in 1992.

The FEWS NET Markets and Trade team maintains a data set of monthly staple food prices in Ethiopia for the purpose of monitoring food security in this country. To improve our analysis of medium- and long-term trends, we would like to expand this data set. However, price data from before 2004 are scant. We were wondering whether you could share with us the price data that you used in that paper to help us consolidate our database.

We thank you in advance, and if you have any questions, please let us know.

Sincerely,

Other Reasons to share: Research Donors

- Donors increasingly require that publicly funded research data be made openly accessible as much as possible.
- Policies tend to allow researchers exclusive data use for a reasonable time period to publish the results of the data.

Other Reasons to share: Research Funders

— How We Work —

BILL & MELINDA GATES FOUNDATION OPEN ACCESS POLICY

- 1. Publications Are Discoverable and Accessible Online.** Publications will be deposited in a specified repository(s) with proper tagging of metadata.
- 2. Publication Will Be On “Open Access” Terms.** All publications shall be published under the [Creative Commons Attribution 4.0 Generic License \(CC BY 4.0\)](#) or an equivalent license. This will permit all users of the publication to copy and redistribute the material in any medium or format and transform and build upon the material, including for any purpose (including commercial) without further permission or fees being required.
- 3. Foundation Will Pay Necessary Fees.** The foundation would pay reasonable fees required by a publisher to effect publication on these terms.
- 4. Publications Will Be Accessible and Open Immediately.** All publications shall be available immediately upon their publication, without any embargo period. An embargo period is the period during which the publisher will require a subscription or the payment of a fee to gain access to the publication. We are, however, providing a transition period of up to two years from the effective date of the policy (or until January 1, 2017). During the transition period, the foundation will allow publications in journals that provide up to a 12-month embargo period.
- 5. Data Underlying Published Research Results Will Be Accessible and Open Immediately.** The foundation will require that data underlying the published research results be immediately accessible and open. This too is subject to the transition period and a 12-month embargo may be applied.

Other Reasons to share: Research Funders



ADS 579 – USAID's Policy on Development Data

WHAT

ADS 579 is USAID's policy on sharing Agency-funded data for public benefit, while ensuring proper protections for privacy and national security. The policy notes that all USAID operating units, including its worldwide missions, must ensure that USAID-funded data is centrally cataloged and made available to the public by default, with limited exceptions. This is in keeping with [USAID's Evaluation Policy](#) and the Office of Management and Budget's [Open Data Policy \(M-13-13\)](#). ADS 579 defines USAID's data governance structure; establishes the Development Data Library (DDL) as the Agency's repository of USAID-funded data; creates Data Stewards in each USAID operating unit; outlines the process for USAID staff and implementing partners to submit data to the DDL; defines the data clearance process; and outlines special considerations for research data.

WHY

Making U.S. Government data accessible, discoverable, and usable by our partners and the international community fuels entrepreneurship, innovation, scientific discovery, and enhanced development outcomes. It contributes to improved design and implementation of development programs while reducing expensive and duplicative data collection efforts. Studies estimate that trillions of dollars in economic value can be unleashed globally by opening untapped data sources. Many of the countries where USAID operates also recognize the value of proactively releasing government data to the public, as members of the



A PRESUMPTION IN FAVOR OF OPENNESS

The Open Government Directive (OMB M-10-06) notes that, "With respect to information, the presumption shall be in favor of openness (to the extent permitted by law and subject to valid privacy, confidentiality, security, or other restrictions)."



Department
for International
Development



DFID Research Open and Enhanced Access Policy

V1.1



THE WORLD BANK
Working for a World Free of Poverty



FORD FOUNDATION



National Science Foundation
WHERE DISCOVERIES BEGIN



OECD
BETTER POLICIES FOR BETTER LIVES

wellcome trust



Other Reasons to share: Journals

- Journals increasingly require data that form the basis for publications to be shared or deposited within an accessible database or repository.
- For example 'Nature journals' and 'PLOS journals' have policies that require authors to make data and materials available to readers, as a condition of publication.

nature
research

 **PLOS** | Open for
Discovery

 CIAT

Other Reasons to share: Open Agriculture Data Initiatives



The Research Data Alliance (RDA) builds the social and technical bridges that enable open sharing of data.

The RDA vision is *researchers and innovators openly sharing data across technologies, disciplines, and countries to address the grand challenges of society.*

Agricultural Data IG (IGAD)

IG

Group details

Status: Recognised & Endorsed

Chair(s): Imma Subirats, Devika Madalli, Johannes Keizer

Case Statement: [Download](#)

Wheat Data Interoperability WG

WG

Group details

Status: Recognised & Endorsed

Chair(s): Esther Dzalé Yeumo, Richard Fulss

Secretariat Liaison: Stefanie Kethers

TAB Liaison: Françoise Genova

Case Statement: [Download](#)

Other Reasons to share: Open Agriculture Data Initiatives



GODAN supports the proactive sharing of open data to make information about agriculture and nutrition available, accessible and usable to deal with the urgent challenge of ensuring world food security.

Over 374 partners:



Petition



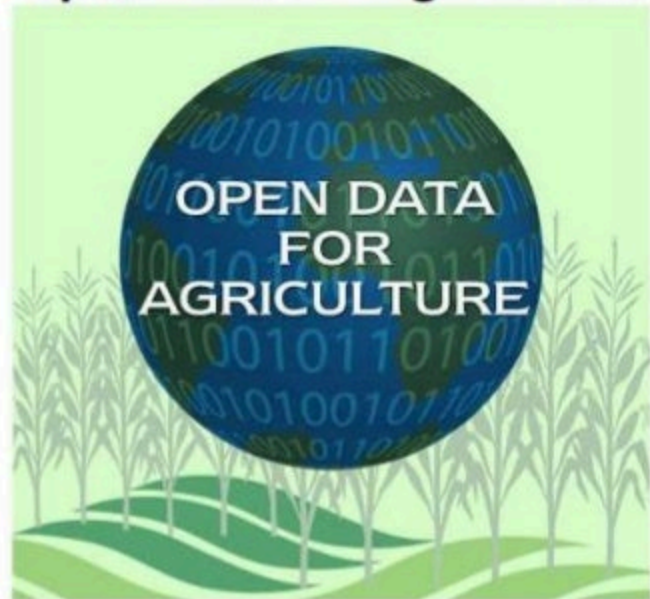
Sign the petition to call on governments, the private sector, and civil society to provide open data on agriculture and nutrition to end world hunger!

Join the Open Data Revolution to end global hunger.



Other Reasons to share: Open Agriculture Data Initiatives

G-8 International Conference on Open Data for Agriculture



“The digital revolution fueled by open data is starting to do for the modern world of agriculture what the industrial revolution did for agricultural productivity over the past century.”

Tom Vilsack - U.S. Agriculture Secretary, G8 International Conference on Open Data for Agriculture, 29th April 2013

Other Reasons to share: Data Citation Benefit?

- Citation benefit from open data, although a smaller one than previously reported. e.g. 2004–2005 deposited data showed a clear benefit of about **30%**
- Authors reuse of their own data for around **2 years**. Third-parties reused the data for up to **6 years**.

Piwovar HA, Vision TJ. (2013) *Data reuse and the open data citation advantage*. PeerJ 1:e175 <https://doi.org/10.7717/peerj.175>



Reasons for the data citation benefit

1. *Data Reuse*. Papers with available datasets can be used in ways that papers without data cannot, and may receive additional citations as a result.
2. *Credibility Signalling*. The credibility of research findings may be higher for research papers with available data. Such papers may be preferentially chosen as background citations or the foundation of additional research.
3. *Increased Visibility*. Third party researchers may be more likely to encounter a paper with available data, either by a direct link from the data or indirectly through cross-promotion. For example, links from a data repository to a paper may increase the search ranking of the research paper.
4. *Early View*. When data is made available before a paper is published, some citations may accrue earlier than they would otherwise because of accelerated awareness of the methods, findings, and so on.
5. *Selection Bias*. Authors may be more likely to publish data for papers they judge to be their best quality work, because they are particularly proud or confident of the results (Wicherts, Bakker & Molenaar, 2011).

Data Citation Principles

“Data should be considered legitimate, citable products of research.”

- Importance
- Credit and Attribution
- Evidence
- Unique Identification
- Access
- Persistence
- Specificity and Verifiability
- Interoperability and Flexibility

CGIAR Big Data Platform



Bringing Big Data to Agriculture,
and Agriculture to Big Data



Where to share: Peer Reviewed Data Journals



ELSEVIER

Home > Journals > Data in Brief

Data in Brief

Editor-in-Chief: Hao-Ran Wang
[View Editorial Board](#)

[Open Access](#)



ISSN: 2352-3409

Wiley Online Library

[Log in / Register](#)

Home > Earth Sciences > General & Introductory Earth Sciences > Geoscience Data Journal

JOURNAL TOOLS

- [Get New Content Alerts](#)
- [Get RSS feed](#)
- [Save to My Profile](#)




Geoscience Data Journal

[Open Access](#)

CIAT

Where to share: Subject Specific Repositories



Data - News - Community - About -

Global Biodiversity Information Facility

Free and Open Access to Biodiversity Data

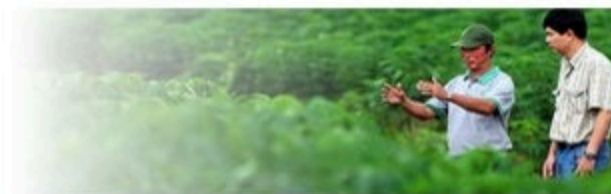
624,609,381	1,643,948	29,170	817
OCCURRENCES	SPECIES	DATASETS	DATA PUBLISHERS



RESEARCH PROGRAM ON
Climate Change,
Agriculture and
Food Security



The Global Agricultural Trial Repository and Database



Home Search Trials About Us Contact Us Site Map Sign

AgTrials—The Global Agricultural Trial Repository


Agtrials.org is an information portal developed by the CGIAR Research

Map Satellite

NCBI Resources How To

SRA SRA Search

Advanced

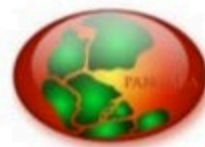


SRA

Sequence Read Archive (SRA) makes biological sequence data available to the research community to sort and allow for new discoveries by comparing data sets. The SRA stores raw sequencing data and alignment high-throughput sequencing platforms, including Roche 454 GS System®, Illumina Genome Analyzer®, Agilent SOLiD System®, Helix Helix® and Complete Genomics®, and Pacific Biosciences SMRT®.

PANGAEA®

Data Publisher for Earth & Environmental Science



All Water Sediment Ice Atmosphere

colombia Search

Help Advanced Search Preferences more...



Where to share: Institutional/CRP Repositories



International Center for Tropical Agriculture
Since 1967 Science to cultivate change

CIAT Research Online

CIAT - International Center for Tropical Agriculture Dataverse (CGIAR)

CIAT - Eco-efficient agriculture for the poor

[Harvard Dataverse](#) > [CIAT - International Center for Tropical Agriculture Dataverse](#)

A screenshot of the ILRI Datasets Portal website. The header is dark red with the ILRI logo and "Datasets Portal" text on the left, and a search bar on the right. Below the header is a navigation menu with links for "ILRI Datasets", "All Datasets", "Organizations", "Groups", "Open Data Policy", and "About". The main content area has a dark red box on the left with the heading "Search Our Data" and a search input field containing "eg. Livestock health". To the right is a white box with the heading "Welcome to ILRI's Datasets Portal." and a paragraph of text: "Here you will find an archive of ILRI datasets. The metadata of all ILRI datasets is publicly available. Access to the data may also be publicly available or only downloadable for registered users. Public available data does not contain sensitive information like names or telephone numbers. Full access to...".

ILRI
Datasets Portal

Search

[ILRI Datasets](#) [All Datasets](#) [Organizations](#) [Groups](#) [Open Data Policy](#) [About](#)

Search Our Data

eg. Livestock health

Welcome to ILRI's Datasets Portal.

Here you will find an archive of ILRI datasets. The metadata of all ILRI datasets is publicly available. Access to the data may also be publicly available or only downloadable for registered users. Public available data does not contain sensitive information like names or telephone numbers. Full access to...

The CIAT logo, consisting of a green circular icon with a white globe and the letters "CIAT" in white text to its right.

Where to share: General Repositories



Guiding principle: FAIR Data

Findable

Accessible

Interoperable

Re-usable

Restrictions to Sharing Data

- Privacy – Information that identifies and individual
- Confidentiality – Information that should not be shared
- Security – Release of data will cause threats to someone or something

CGIAR Intellectual Assets (IA) Principles



CGIAR Principles on the Management of Intellectual Assets ("CGIAR IA Principles")

(effective as part of the Common Operational Framework as of 7 March 2012; approved by the Consortium Board on 1 March 2012 and by the Fund Council on 7 March 2012)

- CGIAR regards its intellectual assets as **International Public Goods (IPG)**
- Maximize the **global accessibility** and ensure broadest possible impact of IPGs
- We work with partners
- Respect of Farmers rights
- Conservation, access and widespread use of Genetic resources
- Private – public partnerships

CGIAR Open Access and Data Management Policy



Consortium

Approved by CGIAR Consortium Board
October 2, 2013

CGIAR Open Access and Data Management Policy (the "Policy")

1. Preamble

CGIAR regards the results of its research and development activities as international public goods and is committed to their widespread dissemination and use to achieve the maximum impact to advantage the poor, especially smallholder farmers in developing countries. CGIAR considers Open Access (defined below) to be an important practical application of this commitment as it enhances the visibility, accessibility and impact of its research and development activities. Open Access improves the speed, efficiency and efficacy of research; it enables interdisciplinary research; assists novel

- Intellectual assets policy as applied to data and other research outputs (**information products**)
 - Open Access – Publications, policies, images, videos, toolkits.....
 - Open Data – Databases, datasets, linked open data
- Effective **October 2013**

When to share: OA-DM Policy Sharing Timelines

Types of Information Products	Policy Deposit Schedule (from October 2, 2018)
Peer-reviewed versions of journal articles	Latest: 6 months from publication
Self-published journals, books, reports etc.	Self-published materials not currently addressed in the Policy
Reports and other papers	Latest: within 3 months of completion
Externally or commercially published books and book chapters	Latest: within 6 months of completion
Data and data sets	Latest: within 12 months of appropriate project milestone within 6 months of publication of products underpinned by that data
Video, audio, scientific images, Photographs	Latest: within 3 months of completion
Computer software/applications/code	Upon completion of software development
Metadata	Latest: before or on publication of the information product
Core/corporate governance documents appropriate for public consumption	As per 'reports' category of Information Product (Core/corporate governance documents not currently addressed separately in the Policy)

When to share: OA-DM Policy Sharing Timelines

Types of Information Products	Policy Deposit Schedule (from October 2, 2018)
Peer-reviewed versions of journal articles	Latest: 6 months from publication
Self-published journals, books, reports etc.	Self-published materials not currently addressed in the Policy
Reports and other papers	Latest: within 3 months of completion
Externally or commercially published books and book chapters	Latest: within 6 months of completion
Data and data sets	Latest: within 12 months of appropriate project milestone within 6 months of publication of products underpinned by that data
Video, audio, scientific images, Photographs	Latest: within 3 months of completion
Computer software/applications/code	Upon completion of software development
Metadata	Latest: before or on publication of the information product
Core/corporate governance documents appropriate for public consumption	As per 'reports' category of Information Product (Core/corporate governance documents not currently addressed separately in the Policy)

Data Sharing: Pre- Publication

- CGIAR Researchers are generally not obligated to share data, during the **pre-publication**, within the CGIAR. (Sometimes even within Centres and CRPs)
- Pre-publication sharing across CRP's and Centres may not be mandatory even when working within the same project.



The screenshot shows the Nature journal website interface. At the top, the 'nature' logo is displayed in white on a dark red background, with the tagline 'International weekly journal of science' to its right. A search bar is visible in the top right corner. Below the header, a navigation menu includes 'Journal home', 'Archive', 'Opinion', and 'Full Text'. The 'Archive' menu item is highlighted in red. The main content area features an 'Opinion' section with the title 'Prepublication data sharing' and a sub-headline 'Toronto International Data Release Workshop Authors¹'. The article text begins with 'Rapid release of prepublication data has served the field of genomics well. Attendees at a workshop in Toronto recommend extending the practice to other biological data sets.' and continues with a paragraph discussing the importance of open discussion and full disclosure of supporting facts in scientific discourse.



Member of the
CGIAR Consortium

www.ciat.cgiar.org
www.cgiar.org



CGIAR

Science for a food secure future