COALITION OF THE WILLING FOR SOIL AND AGRONOMY DATA ACCESS, MANAGEMENT AND SHARING

Data Sharing Guidelines

EIAR
Ethiopian Institute of Agricultural Research

ግብርና ሚኒስቴር
MINISTRY OF AGRICULTURE

Ethiopian Institute of Agricultural Research (EIAR)

EIAR's mission is to conduct research that will provide market competitive agricultural technologies that will contribute to increased agricultural productivity and nutrition quality, sustainable food security, economic development, and conservation of the integrity of natural resources and the environment.

www.eiar.gov.et

COALITION OF THE WILLING FOR SOIL AND AGRONOMY DATA ACCESS, MANAGEMENT AND SHARING

Data Sharing Guidelines

Corresponding author
  Lulseged Tamene, Soil Scientist, Alliance of Bioversity International and the International Center for Tropical Agriculture
  lt.desta@cgiar.org

March 2020

ACKNOWLEDGEMENTS

ACRONYMS AND ABBREVIATIONS

AAUDE	Association of American Universities Data Exchange
BMGF	Bill and Melinda Gates Foundation
CCAFS	CGIAR Research Program on Climate Change, Agriculture and Food Security
CDC	Centers for Disease Control and Prevention
CIAT	International Center for Tropical Agriculture (now part of the Alliance of Bioversity International and CIAT)
CoW	Coalition of the Willing
CSA	Central Statistical Agency
DFID	United Kingdom's Department for International Development
EMA	Ethiopian Mapping Agency
FAO	Food and Agricultural Organization of the United Nations
GEOSS	Global Earth Observation System of Systems
GIZ	Deutsche Gesellschaft für Internationale Zusammenarbeit
GODAN	Global Open Data for Agriculture and Nutrition
GSP	Global Soil Partnership
ILRI	International Livestock Research Institute
INSA	Institute of National Security Agency
INSII	International Network of Soil Information Institution
NACO	National AIDS Control Organization
NCHS	National Center for Health Statistics
NMA	National Meteorological Agency
ODI	Open Data Institute
SDG	Sustainable Development Goal
UK	United Kingdom
USAID	United States Agency for International Development

CONTENTS

Photo: Tsigereda Tafesse

INTRODUCTION

Background

Data can inform decision making, enable innovation and promote transparency in agriculture and nutrition. Better data access, use and sharing can ultimately result in improved yields and food systems efficiency.[14] When more data is made available, it not only provides innovation opportunities, but also helps to ensure validity, accountability and accuracy of the data while avoiding the duplication of efforts.[15]

The global agriculture sector is driving shared and open approaches to data. This drive is led by organizations such as the Global Open Data for Agriculture and Nutrition (GODAN) network who advocate for stronger open data ecosystems,[16] and the GO FAIR initiative, who aim to implement the FAIR data principles - making data Findable, Accessible, Interoperable and Reusable.[17] Data that conforms to the FAIR data principles will be easier to find, access and reuse. This reduces sources of friction that can lead to data not being used to its full potential.

Currently, functional and consistent data management and sharing systems in many African countries

---

14 Open Data Institute, "Enabling data access to support innovation in agriculture", https://theodi.org/project/enabling-data-access-to-support-innovation-in-agriculture/

15 Gelagaye, H. S. 2018. Geospatial Data Sharing Barriers Across Organizations and the Possible Solution for Ethiopia. Spatial Data Infrastructure Program (SDIP), Information Network Security Agency of Ethiopia, Addis Ababa, Ethiopia.

16 Open Up Guide for Agriculture, "The Solution of Open Data", https://openupguideforag.info/why-open-data/the-solution-of-open-data/

17 GO FAIR, "FAIR Principles", https://www.go-fair.org/fair-principles.

are lacking. Studies reveal that data is either not made available or remains scattered spatially and temporally due to a lack of data sharing mechanisms.

The Ethiopian Agriculture Research and Development Community has identified a need to alter the dynamic of disclosure and transparency of Soil and Agronomy Data within Ethiopia. With support from the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), the CGIAR Research Program on Water, Land and Ecosystems (WLE), the International Center for Tropical Agriculture (CIAT) (now part of the Alliance of Bioversity International and CIAT) and other partners, a 'Coalition of the Willing' (CoW) for data collaboration has been created. The CoW is a group of individuals willing to share soil and agronomy data among themselves to facilitate wider scale soil and agronomic data access and sharing.

The CoW has drafted a set of Soil and Agronomy Data Sharing Guidelines aligned with the FAIR data principles. These Guidelines for the Coalition of the Willing outline the requirements that members must adhere to. These Guidelines require an open and transparent approach to sharing Soil and Agronomy Data. Delivering the outcomes will require continued investment of time, resource, collaboration and perseverance by the members of Coalition of the Willing.

Data sharing landscape

Globally, funding organizations, national and international organizations are increasingly encouraging data management and sharing policies and guidelines, for example:

- CGIAR has a set of policy and operational guidelines for implementing open data across the research programs.[18]

- GODAN supports global efforts to make agricultural and nutritionally relevant data available, accessible and usable for unrestricted use worldwide.[19]

- The Food and Agriculture Organisation (FAO) promotes open data policies in its soil data sharing guidelines while calling for data access with

minimal additional restrictions on use for those countries lacking open data policy.[20]

- A number of donors such as USAID, DFID, and BMGF have introduced data access policies encouraging partners to openly share research data generated by funded projects.[21]

Data access and use, data licensing and reuse rights, prioritizing data for release, privacy considerations, data publishing standards, engaging with data users, monitoring commitments, and overall policy transparency are elements considered in most of these guidelines.[22]

Within these trends for greater access to agricultural data, soil data in particular has become an area of growing demand worldwide. The Global Soil Partnership (GSP) has established an international network of soil information institutions (INSII), consisting of 60 countries/institutions to allow the exchange of new and harmonized national data as part of new global datasets. Soil data sharing is one of the emerging areas for enhancing conservation and promoting national and international development.

Despite this growth in demand, barriers to accessing soil and agronomy data persist. Soil and agronomy research data can involve human and other national security elements which are considered confidential or sensitive. Approved protocols, confidentiality agreements and specific publishing processes may restrict access.[23]

Some institutions have guidelines to help alleviate issues related to sensitive data and data privacy. The Association of American Universities Data Exchange (AAUDE) categorizes the data so that standards are applied for sensitive data into four exchange categories:[24]

- Publicly reported data: publicly available and are not subject to data sharing guidelines and confidentiality rules.

- Ad hoc/special requests: group-based access to member institutions.

18 ibid.

19 Global Open Data for Agriculture and Nutrition (2013), "About", https://www.godan.info/about.

20 FAO 2017. GSP Guidelines for sharing national data/information to compile a Global Soil Organic Carbon (GSOC) map http://www.fao.org/3/a-bs975e.pdf

21 ibid.

22 Smith, F., Fawcett, J. and Musker, R. 2017. Donor Open Data Policy and practice: An Analysis of Five Agriculture Programs, GODAN.

23 ibid.

24 ibid

- Confidential exchange items: sensitive and confidential information that require specific data sharing agreements.

- Confidential exchange items with additional rules: highly sensitive and confidential information that require more restrictive data sharing agreements

In order to ensure that data can be accessed more widely despite the prevalence of restrictions on confidential data, organizations such as the UK Data Archives provide different options for data sharing that are relevant to soil and agronomic data:[25]

- depositing the data with a specialist data center, data archive or data bank;

- submitting them to a journal to support a publication;

- depositing data in an institutional repository;

- making them available online via a project or institutional website;

- making them available informally between researchers on a peer-to-peer basis.

Ultimately this is due to the belief that wider access to soil and agronomic data will provide widespread benefits. In the example of Ethiopia, greater access to this data could boost farmer yields, providing more food to the population and revenues to farmers which in turn can grow the economy and raise people out of poverty. The UK Data Archives states that tangible benefits which could lead to wider societal improvement are that data sharing:[26]

- encourages scientific enquiry and debate;

- promotes innovation and potential new data uses;

- leads to new collaborations between data users and data creators;

- maximizes transparency and accountability;

- enables scrutiny of research findings;

- encourages the improvement and validation of research methods;

- reduces the cost of duplicating data collection;

- increases the impact and visibility of research;

- promotes the research that created the data and its outcomes;

- can provide a direct credit to the researcher as a research output in its own right;

- provides important resources for education and training.

Data Sharing Guidelines Articles

These data sharing guidelines are prepared based on consultation with CoW members and a review of guidelines and policies of various national agencies/ organizations, such as CSA, NMA, INSA, and international organizations such as CGIAR, GODAN, FAO, ODI and UK Data Archives. In addition, agencies/ organizations that have data sharing guidelines or experiences have been consulted regarding the way they share the data they generate and how they access data collected by others. These guidelines are divided into 12 articles, with each article having sub-articles.

---

25 UK Data Archive, 2011, "Managing and Sharing Data: Best Practice for Researchers", https://ukdataservice.ac.uk/media/622417/ managingsharing.pdf

26 ibid.

Article 1 - Definition of terms

Confidential Data	Data that has been classified according to a transparent classification scheme where sharing can be restricted.
Coalition of the Willing	Association of data generators and users who are willing to share their data and agreed to be governed by these guidelines.
Data	A representation of information, numerical compilations and observations, documents, facts, maps, images, charts, tables and figures, concepts in digital and/or analog form.
Data Rights Owner	The party who holds Intellectual Property Rights as a result of collecting or creating Soil and Agronomy Data.
Data Standard	Documented, reusable agreements that help people and organizations to publish, access, share and use better quality data.
Data Steward	The party who controls and administers, with permission, any database containing Soil and Agronomy Data. They may or may not also be the Data Rights Owner.
Database	A collection of data that is organized so that it can be easily accessed, manipulated, managed and updated.
Derived Data	Data produced by a process acting on a set of Input Data, this includes any translation, adaptation, arrangement, modification, or any other alteration of the Input Data or of a substantial part of it.
FAIR principles	Internationally agreed principles that data should be Findable, Accessible, Interoperable and Reusable.
Geographic Coverage	Describes the spatial area covered by any data.
Georeference	To associate something with locations in physical space.
Geospatial Data	Detailed information that depict geographic locations and properties of both artificial and natural features that are found on and under the surface of the earth.
Grace Period	Where data can be embargoed while research is carried out.
Input Data	Data that is used as an input to any process that creates Derived Data.
Intellectual Property Rights	Refers to the general term for the assignment of property rights through patents, copyrights and trademarks. These property rights allow the holder to exercise a monopoly on the use of the item for a specified period.

Licence	An official document which gives you permission to do, use, or own something
Machine Readable	Data in a format that can be automatically read and processed by a computer such as CSV, JSON and XML.
Metadata	General information that describes the origin, structure, author, quality and other characteristics of the data.
Metadata Platform	A database designed to store Metadata.
Open Data	Data that is available for anyone to access, use and share. It is published under an open licence that allows it to be used for any purpose.
Open Standard	A Data Standard that is available for anyone to access, use or share under an open licence.
Pre-existing Data	Soil and Agronomy Data that was collected or created before this policy came into force.
Provenance	Information that includes the data's origins and what happens to it as it goes through processes.
Published Data	Data that has been released in a published form for use or reuse by others. It is released data for public use thus to make them available to everyone to use as they wish.
Raw Data	Data that has not been processed and/or analyzed for use.
Research Data	Data that has been collected for the primary purpose of research.
Soil and Agronomy Data	Data that relates to the soil, crops and land of Ethiopia.
Soil and Agronomy Data Sharing Platform	Database to deliver data collection, storage and sharing requirements.
Soil and Agronomy Data Sharing Infrastructure	Includes data assets such as datasets, identifiers and registers, the standards and technologies used to curate and provide access to those data assets, the guidance and policies that inform the use and management of data assets.
Third Party Rights	Rights in Derived Data that may be owned by another party as a result of using Input Data. This might lead to restrictions on whether the Derived Data can be shared.

Article 2 - Objectives of the guidelines

These Data Sharing Guidelines have the following objectives:

- Guide the CoW members and their institutions in sharing data among them and with the scientific community and increase their collaboration based on FAIR principles with consistent, defined rules and regulations, including protection of moral and actual rights of Data Stewards and Data Rights Owners.

- Reduce duplication of efforts, increase resource use efficiency to reduce costs, avoid redundancy and facilitate information-based decision making.

- Demonstrate how national data sharing and access guidelines and policy could work.

- Facilitate the creation and population of a Soil and Agronomy Data Sharing Platform for the members of the Coalition of the Willing.

- Make Research Data more easily transformed into actionable knowledge by researchers.

- Broaden the mass base of the Coalition of the Willing (Interest Group) and scalability of the benefits ushering therein by encouraging and allowing new members to join.

Article 3 - Applicability

These Data Sharing Guidelines are applicable to:

- All registered members of the Coalition of the Willing.

- Any authorized user, researcher, and institution that has either access to or shares their data through the CoW Soil and Agronomy Data Sharing Platform.

- Custodianship of the Cow Soil and Agronomy Data Sharing Platform by EIAR under an assigned data administrator.

- Scope of access. At the time of publication, access to data within the CoW Soil and Agronomy Data Sharing Platform will be limited to the registered members of the Coalition of the Willing.

- New members who join the Coalition anytime provided that they sign into the agreement and abide by internal rules.

Article 4 - Data sharing principles

Data should be Findable

- In order for all Soil and Agronomy Data that is collected or created to be findable it must have a Metadata record created and stored in a public Metadata Platform to enable potential data users to search the records.

- Metadata should conform to a known and agreed Open Standard.

- The Data Rights Owner will be the original creator of the data or their funding organisation.

- Metadata should clearly identify the Data Rights Owner.

Data Should be Accessible

- Access to data should not be denied other than on grounds of privacy, national security, or confidentiality and in those instances a defined process should be followed to access the data.

- Soil and Agronomy Data and Metadata should be shared with the data steward/manager of the CoW data portal as soon as possible after collection/creation.

- Research data should be shared immediately with other users after the publication of any associated research papers. However, those willing to share the data before publication can do so.

- Members of the Coalition of the Willing should upload or store the data on the Soil and Agronomy Data Sharing Platform within one year after the data is generated.

- When data access is requested, the request should include specifics of the geographical location and time the data is generated, and should state the purpose for which data is going to be used.

- Users can access the data that has been submitted by members of the CoW or partners (online, through sending email, or application).

Data should be interoperable

- Data should be submitted using data formats and standards that falicilitate interoperability, sustainability and future use of data as defined by the Coalition of the Willing.

Photo: Tsigereda Tafesse

- Agreed upon guides for data standardization should be developed and used to ensure all data created satisfy a minimum standard.

- Wherever possible data should be shared using agreed, defined Data Standards.

- The type of data requested should be stated, e.g., raw data, published article, maps, etc.

Data should be reusable

- All data coming to the Soil and Agronomy Data Sharing Platform from different organizations/ individuals must be designated as shareable with members unless specified.

- Data owners should not be held responsible for errors that may emanate from analytical procedures/tools applied by the concerned body acquiring the data.

Article 5 - Required Data Infrastructure

In order for these guidelines to deliver the intended objectives, the existence of facilities and other supportive environments is required.

Soil and Agronomy Data Sharing Platform

This should consist of but not be limited to:

- portals for making data available, sharing and maintenance;

- conditions for future easy adjustment as technology changes;

- mechanisms for members to upload and download data from the repository system.

The Soil and Agronomy Data Sharing Platform will be part of the data infrastructure of EIAR and the Coalition of the Willing will use this for data storage, management and sharing.

The Soil and Agronomy Data Sharing Platform should fulfill nationally and internationally set data standards and requirements including metadata, while considering the standards and requirements of current and potential partner organizations.

It should incorporate standards and specifications in the system design to promote the use of best practices for information sharing.

Supportive Infrastructure

A Board elected by the Coalition of the Willing:

- Empowered to set the work plan for the members of the Coalition of the Willing.

- Accountable to oversee the Soil and Agronomy Data Sharing Platform.

A central database administrator who is responsible for:

- Building the prevention of data misuse or unauthorized use/modification of the platform;

- Regularly updating the database to meet the contemporary data management and sharing forms and standards;

- Maintaining database results by setting and enforcing standards and controls;

- Handling issues with regard to data sharing, uploading or any other data management issues assigned by the Board.

Photo: CIAT/Lulseged Tamene

A clear and transparent data sharing process, as part of this process it is required that:

- Coalition of the Willing members shall have explicit data validation procedures.
- Coalition of the Willing members shall not tolerate data manipulation or falsification attempts.
- Coalition of the Willing members will ensure data quality shall apply to ensure quality and standards required.

Article 6 - Incentives for sharing data

Some of the possible benefits for Coalition of the Willing members to share their data include:

- Proper acknowledgement of data sources that in turn encourages quality data;
- Reduction of costs from having one's own platform to share data;
- Automatic compliance with guidelines and policies by using the Soil and Agronomy Data Sharing Platform;
- Inclusion of owners of data in scientific publications;
- A strong network of trust between Data Rights Owners and Data users.

ANNEX: RESEARCH METHODOLOGY AND REVIEW

1. ADDITIONAL BACKGROUND INFORMATION

In recent years, data is increasingly shared not only for data users, but also for enhancing development of nations through transparency, reduced costs, easy access to information and efficient decision making. The assumption is that data is useful when delivered to the right hands (data users) within a context where it can be most valuable (timeliness). Both data availability and its availability to the public are becoming vital to ensure validity, accountability and accuracy of the data.[14] Data sharing also helps to avoid duplication of efforts and enhances the possibility of innovation. However, there is currently a lack of functioning, consistent data management and sharing systems in many African countries. Studies reveal that data is either not made available or remains scattered spatially and temporally due to a lack of data sharing mechanisms.

Institutions in Africa face various challenges in data management and sharing due to the protective nature of organizations and individuals[15] and a lack of consistent policy and guidelines, both at national, regional and institutional levels. Ethiopia has not enacted national policy or guidelines to share soil and agronomy data; as a result, data generated by different research, academic and development institutions is not easily accessible to users. While a draft policy has been created, the process of enacting it takes time. Once enacted, this policy will require implementation guidelines to support bringing it to fruition.

In the interim period, with support from GIZ, CIAT and other partners, a 'Coalition of the Willing' (CoW) has been created. The CoW is a group of individuals willing to share soil and agronomy data among themselves in order to facilitate wider scale soil and agronomic data access and sharing.

This work was initiated to review the existing data sharing guidelines of different research institutes and organizations and develop guidelines for sharing soil and agronomy data among the CoW members, this will form the basis of the development of further guidelines under the policy once it is enacted.

---

14 Gelagay, H. S. 2018. Geospatial Data Sharing Barriers Across Organizations and the Possible Solution for Ethiopia. Spatial Data Infrastructure Program (SDIP), Information Network Security Agency of Ethiopia, Addis Ababa, Ethiopia

15 Pierre , H.J. nd. Analysis of data sharing environment and major challenges currently being faced in data sharing in Rwanda, Geographic Information Systems & Remote Sensing Research and Training Centre (CGIS-RS) at National University of Rwanda (NUR)

The review focused on the challenges and opportunities of data sharing, policies and guidelines, principles, data storage, management and infrastructure etc. prevailing in the country and globally.

In addition, institutions such as CSA, EMA, ILRI, have been contacted in order to get insight about their experiences in sharing the data they generate or access data generated by others. These guidelines are based on insights from a literature review, discussions with individuals and institutions and the aspirations of the members.

These guidelines are a vital step in enabling data sharing which will in turn promote the possibility of partnerships and collaborative works among institutions/individuals. It outlines an efficient way of gathering, administering, distributing and communicating data to the relevant agencies/bodies.

2. RESEARCH REVIEW

2.1. Global trends

Globally, there is a general trend that funding organizations develop data sharing policies and guidelines for the projects they support.[16] Data access and use, data licensing and reuse rights, prioritizing data for release, privacy considerations, data publishing standards, engaging with data users, monitoring commitments, and overall policy transparency are the important elements considered in most of the guidelines.[17]

---

16 Eynden, V., V., Corti,L. Woollard, M., Bishop, L. and Horton, L. 2011. Managing and sharing data: Best practice for researchers, UK data archive,

17 Smith, F., Fawcett, J. and Musker, R.2017. Donor Open Data Policy and practice: An Analysis of Five Agriculture Programs, GODAN.

Photo: CIAT/Georgina Smith

Coalition of the Willing for Soil and Agronomy Data Access, Management and Sharing

National and international organizations are increasingly encouraging open data policy. For instance, CGIAR, the consortium of 15 international research organizations, has a policy and operational guidelines for implementing open data across the research programs (publications, datasets, and audiovisual materials. Accordingly, they have agreed on an open data policy and developed guidelines to implement it. They have agreed to share information resulting from CGIAR research – including publications, datasets and audiovisual materials.[18] The focus is to bring together research communities with different cultures and practices of data management and sharing and to draw on the best practices from constituent Research Councils.[19]

The Global Open Data for Agriculture and Nutrition initiative (GODAN) advocates for an open data policy,[20] while the CGIAR guidelines of CCAFS agree with an open data policy but within a stipulated time frame to make the data open to the public.[21] The guidelines state the responsibilities of agencies entering into data sharing agreement,'...Participant Centre agrees to publicly share any data and/or models generated and compiled as a result of activities under this agreement through CCAFS's data portals as soon as it is practically possible but no later than twelve months of generation for metadata and twenty-four months for other data and/or models'.[22]

This agreement further details how to share the data and who with, and the means to share, such as which of a number of different websites is to be used for specific data. It also requires the data users to fully acknowledge the organization that generated the data, and data generating organization to provide with details of possible drawbacks of the data which may occur during the analysis or further use of data.

A number of donors such as USAID, DFID, and BMGF have introduced open access data policies encouraging partners benefiting from their funds to openly share data generated by funding from these donors. However, under data licensing and reuse rights, USAID requires implementing research partners to have certain restrictions on data usage.[23] In the scientific community, many journals and academic institutions now advise the authors to upload their raw data, in case claims occur, or if further analysis is needed.

FAO, in its soil data sharing guidelines, promotes an open data policy. Cognizant of national policy variations, the guidelines call for data access with minimal additional restrictions on use for those countries lacking open data policies. In addition to accessibility and availability requirements of soil data to users, FAO makes it clear that data should be available for free or at the cost recovery level. The data sharing principles of FAO also allow the re-use and redistribution of soil data. However, they require data users to conform with all applicable laws and regulations, such as the World Data System data sharing principles, and GEOSS Data Sharing Principles, which may also label data as 'sensitive' or 'restricted'.[24]

The World Data System[25] has four fundamental principles. These are summarized as:

a. data, metadata, products and information should be fully and openly shared, subject to national and international jurisdictional laws and policies;

b. data, metadata, products and information produced for research, education and public domain use have to be made available for sharing without delay and free of charge or no more than cost of dissemination;

c. all who produce, share and use data and metadata are the stewards of those data and have responsibilities of ensuring authenticity, quality and integrity of the data are preserved

---

18 CGIAR. 2014. CGIAR Open Access and Data Management Implementation Guidelines.

19 Eynden, V., V., Corti,L. Woollard, M., Bishop, L. and Horton, L. 2011. Managing and sharing data: Best practice for researchers, UK data archive.

20 Smith, F., Fawcett, J. and Musker, R. 2017. Donor Open Data Policy and practice: An Analysis of Five Agriculture Programs, GODAN.

21 CGIAR. 2011. Research Program on Climate Change, Agriculture and Food Security (CCAFS): Intellectual Property and Data Sharing.

22 ibid.

23 Smith, F., Fawcett, J. and Musker, R.2017. Donor Open Data Policy and practice: An Analysis of Five Agriculture Programs, GODAN.

24 FAO.2017. GSP Guidelines for sharing national data/information to compile a Global Soil Organic Carbon (GSOC) map.

25 World Data System (WDS). 2015. Data sharing principle, https://www.icsu-wds.org/services/data-sharing-principles.

d. data should be labeled 'sensitive' or 'restricted' only with appropriate justification and following clearly defined protocols

FAO clearly states that copyrights and ownership of data need to be defined clearly and the rights of the owners remain intact based on the governing policies and regulations. As stated under principle 3 above, the world data system guidelines remind us that those who produce, use and share data are also custodians of the data and have to ensure that the authenticity, quality, and integrity of the data are preserved. The FAO Global Soil Partnership has developed a data sharing policy that aims to ensure:

• ownership rights to shared soil data are respected;

• the specific level of access and the conditions for data sharing are clearly specified;

• the ownership of each dataset and web service is properly acknowledged and well-referenced;

• the data owners are protected from any liability arising from the use of their original and/or derived data.

To build the Global Soil Information System (achieve GSP Pillar 4), the Global Soil Partnership (GSP) has established an international network of soil information institutions (INSII), consisting of 60 countries/institutions. The aim is to allow the exchange of new and harmonized national data as part of new global datasets including soil profile data, soil polygon maps, soil property grids and the monitoring of indicators (SoilSTAT). Both (GSP and INSII) are facilities through which the task of a Global Soil Organic Carbon (GSOC) mapping can be developed.[26] Soil data sharing is one of the emerging areas for enhancing conservation and promoting national and international development. Globally, there is a high demand for soil information in relation to carbon sequestration (climate change), the prevention of soil degradation, and improvement of agricultural production.[27]

26 FAO.2017. GSP Guidelines for sharing national data/information to compile a Global Soil Organic Carbon (GSOC) map.

27 INSPIRE. 2013. Infrastructure for Spatial Information in Europe: Data Specification on Soil – Technical Guidelines, European Commission Joint Research Centre.

It is acknowledged that many researchers or organizations might not initially have obvious motivations to participate in data sharing and use. According to Allemang and Teegarden, a private enterprise, scientist or farmer might not see a reason to share the data generated by their efforts. However, the motivations for sharing data can be profound once stakeholders understand the benefits, which vary widely depending on who the stakeholders are.[28]

The Guidelines of the Association of American Universities for Data Exchange[29] advocate for free exchange of timely and accurate information while fully appreciating privacy and confidentiality requirements for the data to be shared. The association further states that shared data should be stored in a safe and secure manner, and used in accordance with established data-sharing rules and guidelines when presented and disseminated. This means there is a need to encourage partners to develop data sharing and management policy and guidelines to develop common understanding.

ODI guidelines suggest that governments have three ways to make sure that open data benefits everybody.[30] Firstly, governments may join international initiatives, such as the open data charter that provides a framework of principles for how governments should share and publish information. There are countries listed as exemplars in this regard, such as Ukraine for joining international initiatives and sharing data. Secondly, strong engagement with user groups helps them to find ways of solving problems with data sharing. Mexico is cited as an example where the government is experimenting with releasing data that people need through the Labora startup hub. Thirdly, through widespread data literacy: the knowledge, skills and confidence to innovate with data. Innovation hubs such as dLab in Tanzania are cited as an example that provides essential training to help government initiatives with local communities.

28 Allemang, D., and Teegarden, B. 2016. A Global Data Ecosystem for agriculture and food.

29 ibid.

30 Dodds, L. (2015). How to write a good open data policy – The ODI. [online] Theodi.org. Available at: https://theodi.org/article/how-to-write-a-good-open-data-policy/

2.1.1. FAIR Data

While these policies help researchers to think about access to data, there are other aspects to consider in order to maximise value from using that data.

There are a number of issues that can lead to data, however licensed, not being used effectively. Users need to know that relevant data exists and, once found, must be easily accessible to them. Publishing data in standard, well-structured formats will make it simpler to use in a range of tools and programming languages.[31]

Problems and costs that come with finding, accessing and using data quickly add up, particularly when a user has to draw on multiple datasets. By using the FAIR principles set out by the GO FAIR Initiative, it becomes easier to reduce these issues as an important part of unlocking value from data.

The FAIR data principles identify four important characteristics of datasets that will make them easier to use:[32]

- **F**indable – datasets should have a unique identifier metadata which describes its contents, sources and structure, and should be published so they can easily be found with a search engine or in a data portal.

- **A**ccessible – datasets and their metadata are easily accessible – eg over the web, with appropriate access controls for shared data.

- **I**nteroperable – datasets should be published and organised using open standards for data, so they can be easily accessed using a range of tools, and combined with other sources.

- **R**eusable – datasets should be published with a clear licence and/or terms of use, and have appropriate documentation and metadata that describes how the dataset has been collected and processed, allowing users to understand its potential and its limitations.

The four principles have underpinned 15 recommendations to help data publishers implement the principles consistently. Supporting guidance has been developed to provide more detailed recommendations to help publishers assess whether individual datasets are published in a FAIR and open way.

2.2. Country experience: Ethiopia

Increasing demand for data sharing is partly attributed to internationally led development frameworks, such as the SDGs, and other donor supported programs; similarly national socio-economic conditions are also changing fast and need to be monitored. National policy makers are required to base their decisions on concrete evidence. This increased demand for data sharing exists in Ethiopia, but here organizations lack clear policy and guidelines for data sharing. During the past few years, there have been attempts by different organizations and ministries to develop their own data sharing policies and guidelines (e.g. INSA, CSA, EMA), and this work continues.

2.2.1. Ethiopian Spatial Data Information

Ethiopia established a Spatial Data Infrastructure (Ethio-GIS) in 1999. The aim was to reduce the duplication of efforts in data production and dissemination. This effort encouraged follow-on national initiatives to establish national spatial data infrastructure such as the Ethiopian mapping agency's spatial data infrastructure in 2002.[33]
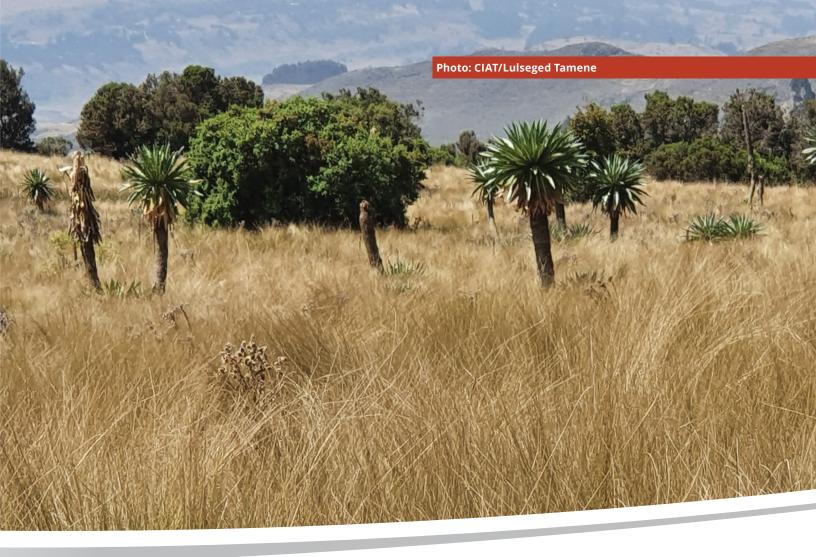
2.2.2. Information Network Security Agency (INSA)

The Information Network Security Agency (INSA) officially launched the Ethiopian National Spatial Data Infrastructure in 2014, albeit with little modification of the earlier infrastructure developed by EMA.[34] INSA has a policy of sharing geospatial data under strict rule of ensuring national security.[35] The components of this policy include broader definitions of concepts related to data and data sharing. INSA (2016) in its policy document states the importance of reducing reliance on foreign technology through building national spatial technology capabilities; spatial

---

31 Smith F, Dodds L, Day C et al. Creating FAIR and open data ecosystems for agricultural programmes [version 1; not peer reviewed]. Gates Open Res 2018, 2:42 (document) (https://doi.org/10.21955/gatesopenres.1114883.1).

32 ibid.

33 Gelagaye, H.S. 2017. Geospatial Data Sharing Barriers across Organizations and the Possible Solution for Ethiopia, International Journal of Spatial Data Infrastructures Research, Vol.12, 62-84.

34 ibid.

35 ibid.

information collection, storage, processing and distribution processes.

This policy document requires data gathering and storage to conform to national security policies. It also states the need for raising citizens' awareness on what, how and for what purposes spatial information and technology shall be utilized and the possible consequences of disclosing sensitive spatial information on national security. As this agency is mainly concerned with national security issues, data gathering, storage and sharing is based on its national mandate.

In the effort to develop self-reliant data systems, the agency also calls for capacity development. Enacting legal frameworks that encourage legitimate users and making illegal users accountable are part of the policy document of the agency. Cooperation with

stakeholders and coordination of different efforts are important components of this policy.

The scope of the policy covers spatial information maintained in a digital and printed form including spatial information and technologies used to collect, store, process and distribute spatial information. The policy is applicable to organizations and individuals engaged in the spatial information and technology sector in the country. This means, all organizations and individuals involved in spatial information generation, storage and use have to abide by this policy and governed by subsequent legal documents.

In addition, the INSA policy document underscores the need for legal framework on collection; storage; dissemination and utilization of spatial information technology for the realization of the country's development and good governance agenda. According

to the policy document, the legal framework helps to address problems related to quality, accessibility, ownership liability, security, integration, and cooperation.

2.2.3. National Meteorological Agency

The National Meteorological Agency (NMA) is another important data creator and provider in the country. Its products, as stated on the agency website, include rainfall, temperature, humidity, wind, sunshine hours, hail, thunderstorm, air pressure, cloud coverage, cloud type, forecasts and climatological maps, presented as data and a map. The agency classifies its product/data into: those that can be provided through telephone requests and those that require top management decisions for sharing. The data management and dissemination department receives the application and process based on the data sharing guidelines of the agency. Those data are accessed after the necessary payments are made.

2.2.4. Ministry of Environment, Forest and Climate Change

The ministry of environment, forest, and climate change[36] is another organization that has developed its own data sharing policy. The policy requires the establishment of a repository system with the aim of making quality and quantity information open to users. The policy calls for easy access and free to use data. Interoperability, sustainability, security, and conformity to legal framework are other guiding principles of the ministry's data sharing policy. Similar to AAUDE , the ministry has four categories of data: Open access, Access through agreement, Restricted Access, and those that are never disclosed.

2.2.5. Ministry of Information, Communication and Technology

The Ministry of Information Communication and Technology [37] has drafted a policy on open data. The draft policy is based on the existing national legal framework and is working to introduce an Open Data Initiative across ministries and agencies. According to this document, the national Open Data Portal

has been launched and several key ministries have established procedures to collect and manage data. In terms of developing the required infrastructure, Ethiopia is working to establish a national Public Key Infrastructure (PKI) framework. The government has also adopted an e-government enterprise architecture and interoperability framework. However, many institutions lack data management and sharing guidelines, except for some of the organizations mentioned above. Others have varied legal and cultural frameworks under which they operate.

Although the draft policy document emphasizes the open data policy, it also recognizes the existence of restrictions, probably imposed by organizations/ agencies. Data categories recognized by this draft policy include:

- Restricted data: Data accessible only through a prescribed process of registration and authorization by respective departments / organizations.

- Sensitive data: as defined in various acts and rules of the government.

- Shareable data: Those data not covered under the scope of restricted and sensitive with regards to privacy, national security, public order and safety or other confidentialities. This category of data can be accessed freely and openly.

2.2.6. Central Statistical Agency

Central Statistical Agency (CSA)[38] data sharing is based on directive no.1/2004 of the agency. According to this directive, data access is possible through the Data Processing Department of the CSA that makes the necessary preparatory arrangements provide access to raw data supplemented with the data dictionary as well as other essential information.

The directive strictly states that raw data can only be accessed through the CSA's Public Relation Services after filling-in the request form prepared for this purpose. Data provided in hard copy is accessed only through payment, subject to a support letter stating the purpose for which the data is needed.

Unless specified in the directives for organizations and individuals who can access raw data free of charge,

---

36 MEFCC. nd. Data Sharing Policy of Ministry of Environment, Forest, and Climate Change, Ethiopia

37 The Federal Democratic Republic of Ethiopia Ministry of Communication & Information Technology. 2018. National Open Data Policy of The Government of Ethiopia

38 CSA. Directive No. 1 /2004 Directive Issued to Establish Procedures for Accessing Raw Data to Users.

other users are required to pay for the data that are not freely available on the website. However, the agency stores some of its products on its website that can be accessed free of charge.

2.3. Challenges of data sharing in Ethiopia

International organizations and national governments have agreed to work towards the achievement of sustainable development goals. One of the goals is ensuring food security through sustained agricultural development, which depends on access to quality, adequate and timely information, including soil and agronomy data.[39] However, organizations and countries are facing challenges including:

- Lack of coordination between organizations, poor data quality and technological compatibility.

- Institutional and technological issues, which are major barriers to geospatial data sharing, particularly in Ethiopia.[40]

  - geospatial data are scattered across several ministries, local agencies, research institutes and universities.

  - There is a lack of central repository system to access and share such data among organizations and individual researchers.

  - Gelagaye (2017) identified inconsistency in data standards or a complete absence of standards, and poor data quality that hampers inter-organizational geospatial data sharing in Ethiopia.[41]

  - CGIAR data sharing guidelines describe the challenges associated with digital storage and preservation of data. These are associated with future accessibility and usability of the data when needed. As technology is growing fast, data storage and management has to consider formats and standards that allow future accessibility. This requires updating the storage

and repository system with technological advancement.[42]

- Absence of legal frameworks to govern data sharing, protect ownership and copyright and cost recovery issues hampers institutions from effectively sharing data in Ethiopia.

- Some policies inhibit geospatial data sharing due to concerns over the inability to prevent data misuse or liability claims[43] or national security concerns.[44]

- The country is not reaping the full potential benefits from the data it has. This is mainly due to challenges associated with quality and maintaining consistency; resource scarcity and wastage; lack of accessible spatial information.[45] Duplication in production and procurement of spatial information has resulted in time and resource waste.

- INSA acknowledges lack of common vision and activities regarding collection, management and utilization of spatial information among stakeholders.

- The awareness level of researchers and existing preventive culture of access to data is a constraint to data sharing. Many researchers, even in the international research organizations resist sharing their data, and this happens at their expense, missing access to data generated by others.

39 Allemang, D., and Teegarden, B. 2016. A Global Data Ecosystem for agriculture and food.

40 Gelagaye, H.S. 2017. Geospatial Data Sharing Barriers across Organizations and the Possible Solution for Ethiopia, International Journal of Spatial Data Infrastructures Research, Vol.12, 62-84.

41 ibid.

42 CGIAR. 2014. CGIAR Open Access and Data Management Implementation Guidelines.

43 Gelagaye, H.S. 2017. Geospatial Data Sharing Barriers across Organizations and the Possible Solution for Ethiopia, International Journal of Spatial Data Infrastructures Research, Vol.12, 62-84.

44 ibid.

45 ibid.

Photo: CIAT/Lulseged Tamene

2.4. Possible restrictions and categories of data

Guidelines regarding research that involve human or some national security elements detail the required protocol and the level of confidentiality of the data. Such guidelines provide strict rules in data management and sharing.[46] [47]If access is allowed to such data, applicants are required to sign an agreement of confidentiality. Research Data Center at the CDC's National Center for Health Statistics (NCHS) is one such organization which follows strict rules for confidentiality. If access is allowed conditionally, the user is expected to submit the paper or report for publication to NCHS for disclosure limitation review.[48]

In line with this, AAUDE guidelines categorize data into different levels and restrict sharing of sensitive data.[49] However, they promote sharing of data among

the members of the association. These guidelines also require staff members of the association to ensure that the data are safely and securely maintained in a manner that is consistent with industry standards.[50] Accordingly, member institutions rely on their representatives to ensure safety of the data. These representatives have the responsibility to ensure that data sharing guidelines are followed strictly and data are shared in the manner the association guidelines requires. In addition, these guidelines require the storage to be safe and protected in a manner that prevents data from being stolen or accidentally accessed.

The AAUDE categorizes the data so that standards are applied for sensitive data. The aim is to balance the need for free, unimpeded use of data with the need for ensuring confidentiality. Accordingly, the data is categorized into four exchange categories:

---

46 ibid.

47 NACO, 2015. Data Sharing Guidelines: National AIDS Control Organization, Ministry of health and family welfare, India.

48 National Institutes of Health, Data Sharing Workbook (2004). http// grants1.nih.gov/grants/policy/ datasharing/data_sharing_workbook.pdf

49 AAUDE, 2018 AAUDE Data Sharing Guidelines and Confidentiality

Rules, http://aaude.org/system/files/documents/public/reference/data-sharing-confidentiality-rules.pdf

50 ibid.

- Publicly reported data: publicly available and are not subject to data sharing guidelines and confidentiality rules. These are open to any user.

- Ad hoc/special requests: member institutions regularly send various inquiries and special requests to exchange participants. Such inquiries and special requests are not deemed confidential unless specified by the inquiring or the responding institution.

- Confidential exchange items: contains sensitive information that exchange participants have defined as confidential. Establishment and observance of data-sharing rules and guidelines for these items are essential. Detail data sharing rule is provided for this category.

- Confidential exchange items with additional rules: this contains data that are considered highly confidential and/or sensitive by contributing institutions. Additional rules are needed to govern the dissemination of such data and users of these data should abide by these rules.

The ODI also classifies data access levels as closed, shared, or open.[51] The ODI further states approaches to identify and prioritize data for release; data licensing and reuse rights; data publishing standards to ensure that data is shared in well-structured, machine-readable formats, with clear metadata and documentation; how the organization will work with external stakeholders to help guide the release of data and ensure data can be easily used; what the organization is committing to do over the time span of the policy, and how the policy and the processes it describes will be reviewed based on feedback from stakeholders and lessons learned.

2.5. Data sharing options

There are different options for data sharing in UK Data Archives.[52] Accordingly, data owners or researchers can share their data using one of these options:

- depositing the data with a specialist data center, data archive or data bank;

- submitting them to a journal to support a publication;

- depositing data in an institutional repository;

- making them available online via a project or institutional website;

- Making them available informally between researchers on a peer-to-peer basis.

These guidelines detail the advantages and disadvantages of each option. In addition, these options assume that data of a single entity will be shared. However, in planning to bring data from different institutes and organizations to a central repository, two stage guides are needed. The first is how data is shared with the central repository system. The second layer is the mechanism/guidance to share the data with end users from the repository.

Despite the difference in data sharing options they provide, most guidelines agree that data management and sharing should fulfill the criteria of interoperability and technological compatibility.

2.6. Benefits of data sharing and developing guidelines

The UK archives states the benefits of data sharing in their guidelines as:[53]

- encourage scientific enquiry and debate;

- promote innovation and potential new data uses;

- lead to new collaborations between data users and data creators;

- maximize transparency and accountability;

- enable scrutiny of research findings;

- encourage the improvement and validation of research methods;

- reduce the cost of duplicating data collection;

- increase the impact and visibility of research;

- promote the research that created the data and its outcomes;

- can provide a direct credit to the researcher as a research output in its own right;

---

51 Dodds, L. (2015). How to write a good open data policy – The ODI. [online] Theodi.org. Available at: https://theodi.org/article/how-to-write-a-good-open-data-policy/

52 Eynden, V., V., Corti,L. Woollard, M., Bishop, L. and Horton, L. 2011. Managing and sharing data: Best practice for researchers, UK data archive.

---

53 ibid.

- provide important resources for education and training;

- improved access to, and use of data at grassroots, local, national and global levels holds the potential to transform both long-standing and emerging problems.

Similarly, the American Psychological Association summarizes the benefits of data sharing guidelines as:[54]

- promote scientific progress;

- encourage culture of openness and accountability among larger scientific groups;

- allow geographically dispersed individuals and those with limited resources to investigate scientific questions of their interest, facilitate replication and validity checks;

- promote aggregation for the purpose of knowledge synthesis;

- scientific process is enhanced by managing and sharing research data;

- data sharing enhances the quality of research outputs and the re-use of research data.

2.7. Case experiences (interview)

Interviews were conducted with some governments and international organizations to understand their experiences in terms of data sharing. For instance, the central statistical authority produces huge data every year, in addition to the census that is conducted every ten years. According to the IT section which manages the data, there are data/information/reports which are uploaded on the website. These are freely accessed. Other printed materials or raw data not available on the website are accessed through request by letters. Individuals or institutions have to pay for such data. Access has to be officially approved.

On the other hand, as part of the CGIAR open data policy, discussions were held with ILRI communication and knowledge management section. This section is mandated to disseminate the outputs of research in the forms of reports, journal articles, policy briefs etc. The raw data is managed by the Nairobi office. The basic principle here is to make the research output accessible to the public through open access publication, and also make the data available, linked to the publication.

Even though the intention is to share raw data after publication, there are still many issues to be solved, including the perceptions of the researchers with regard to data sharing. During the interview, one staff member of the CGIAR mentioned a case in which one of the member organization's partners located in a different part of Ethiopia refused to share the data they generated. This means they also compromised the chance of sharing data generated by others.

Another case is that of N2Africa, financed by the Gates Foundation. This research program has eleven countries, each working with different partners. The server is based at Wageningen University, with the main administrator having full responsibility for managing the data, letting others access or allocate access rights. The national researchers have passwords and usernames to upload data, access the data and use if they need, and do certain modifications.

Even though the Gates Foundation is advocating for open access, there is no protocol developed to share the data outside the researchers and partner organizations. There are templates, data standards and formats that each partner has to use either to upload or access the data. Researchers can access data from other countries, but they have to properly acknowledge using the data from other countries.

Practically, open access works more for research outputs in the form of reports/publication rather than raw data. The main reasons presented are preventing abuse, protecting intellectual property rights, and issue of costs involved to generate the data.

---

54 American Psychological Association. 2015. Data sharing: Principles and consideration for policy development, data sharing working group.

Partners: