Center for International Forestry Research

A CGIAR Consortium Research Center

CIFOR

RESEARCH DATA MANAGEMENT

GUIDELINES AND PROCEDURES

1 July 2013

These Guidelines and Procedures follow the principles outlined in the CIFOR *Research Data Management Policy* (1 July 2013)[1], and explain the implementation of research data management (RDM) practices. This document provides guidelines and procedures on managing research data — from applying for a research grant to depositing the findings into an archive.

1. Scope

Throughout this document, the term data is used to refer to the research data that is produced or collected during research activities. In the context of data archives, data means digital data.

Research data can be generated for different purposes and through different processes. Types of data that should be considered in RDM include:

- Observations: raw data, including that from field data collection;
- Derived data: resulting from processing or combining 'raw' or other data;
- Models and simulations: both the model with associated metadata, and the computational data arising from the model.

Data can be classified according to the way the researcher obtains the data:

- Primary Data
  - The data collected by the researcher himself/herself.
  - The data has never been gathered before, whether in a particular way, or at a certain period of time.
- Secondary Data
  - Data that has been collected by other organizations.
  - These data can be accessed directly from other researchers, or through published materials, a data portal, or data repository.

The categories of research data addressed by this policy include:

- Cleaned raw data. Some data in this category must be kept safe, such as when the observation cannot be repeated. Other data in this category may be discarded after the project is complete, such as when the experiment can easily and cheaply be repeated at will.
- Intermediate data. This data can usually be discarded at the end of the computation, as soon as the computation has been certified as having been completed correctly. For projects that develop and utilize new software and/or computational methods for processing the data, those resources also need to be retained as part of the data for purposes of final data validation and for reproducibility.
- Final data, including results to be uploaded to repositories, or to be made available to other parties, e.g. through CIFOR or publisher websites.

Data may be subject to access restrictions, such as:

- temporary restrictions to preserve intellectual property/copyright claims
- permanent restrictions due to ethical considerations, e.g the privacy of research subjects
- temporary or permanent restrictions due to proprietary considerations of research sponsors and/or partners.

---

[1] https://my.cifor.org/Documents/Data_Information_Management/CIFOR RDM Policy.pdf

2. Roles and responsibilities

Clarity in data management roles and responsibilities helps to establish and maintain a consistent RDM framework across CIFOR.

Research Director
- Oversees the management of research data within their area of responsibility

Project Leader
- Has overall accountability for managing research data in the project, and ensures that a RDM plan is developed and implemented
- Sets the RDM culture of practice based on guidance, and delegates specific responsibilities as appropriate
- Makes decisions about issues such as data access, data sharing, and long-term retention of data
- Obtains the necessary resources to ensure RDM consistent with CIFOR policy, e.g. provision of Information Technology resources, or allocation of responsibility for data control to a team member
- Is accountable for ensuring that, should they leave CIFOR during the life of the project, the data stays with the project.

Researcher
- Maintains records of research data and primary materials, and ensures that these records and the research data and materials are securely stored
- Makes the data available to other researchers via open or negotiated access, as appropriate and in accordance with the requirements of research funding bodies and CIFOR's Research Data Management Policy
- Ensures that, where projects span several institutions, an agreement is developed at the outset covering the ownership and storage of research data and primary materials within each institution in accordance with CIFOR policies and guidelines
- Ensures that adequate back-up, archival and monitoring strategies are in place to prevent the loss of research data and primary materials, and associated delays in completing research.

Data Expert (e.g. Statistician, GIS specialist)
- As a subject specialist, the data expert must understand the technical framework underlying data capture, processing, manipulation and management activities in their respective area
- Ensures overall data integrity and conformity of information gathered by their portfolio

Data and Information Services Unit
- Manages data repository
- Maintains RDM procedures and provide RDM training and support
- Protects rights of data contributors
- Responsible for institutional use of data, in both uploading and downloading data

Information Technology Unit
- Provide facilities for the safe and secure storage of research data and records, as appropriate
- Responsible for developing and maintaining a centrally-supported institutional repository for the provision of open and/or controlled access to secure storage of research data

Data User
- Responsible and accountable for all data access made through their user account and the subsequent use and distribution of the data.

3. Data lifecycle

To ensure research data is managed to the highest standard, it must be included and integrated into five stages of project management: proposal management; contract management; project start-up; project ongoing (tracking and reporting); and project closure. Each stage is characterised by a set of activities carried out by staff involved in the project, with responsibilities as described above, in Section Two of this document.

3.1. Proposal management

Funding agencies increasingly require that applications for support include data sharing and dissemination plans or data management plans. Thinking ahead during this early stage enables researchers to take into account important issues from the very beginning, which can simplify the process and avert problems later on at the data submission stage.

Here are the suggested steps to follow regarding RDM at the proposal management stage. Annex 1 also provides information on the process workflow for this stage:

1. Conduct a review of related existing datasets. Please visit http://dvn.iq.harvard.edu/dvn/dv/cifor for background information what statistical data is available within CIFOR, and its copyright status, and the CIFOR geo portal http://gislab.cifor.cgiar.org/geoportal/catalog/main/home.page for spatial data. Responsibilities for these repositories within CIFOR rests with the Data and Information Services Unit and the GIS Unit, respectively; both can be reached at <cifor-rdm@cgiar.org>.
2. Determine whether a new dataset will be produced.
3. Describe any special challenges that might arise when archiving the data.
4. Describe the potential users of the dataset.
5. Determine the costs of preparing the data and documentation for archiving.
   a. Do you need to digitise the paper based research data?
   b. Will you transcribe qualitative data to be able to deposit in an archive?
   c. Will you need to hire someone to clean the data (cleaned, checked, or spell-checked)
6. Do other parties hold copyright in the data; if so, do you need to seek copyright clearance?

3.2. Contract management

These are the suggested steps to follow regarding RDM at the contract management stage (Annex 2):

1. Visit the Letter of Agreement (LOA) Policy (https://my.cifor.org > Tools and Resources > Policies)
2. Understand funding agency requirements for data management
3. Ensure clarity of ownership of data, consistent with CIFOR Policy
4. Develop the rights and obligations of all parties with respect to their roles and responsibilities for the management and retention of research data
5. Develop the concept of how the data will be retained after the project ends or the grant expires
6. Determine how the data will be disseminated and verify that it will be available for sharing.

Confidentiality and disclosure

CIFOR claims ownership rights to intellectual property, including data generated through grant funding. Researchers and scientists have a responsibility to maintain research data and make that data available for preservation by CIFOR — both as a matter of research integrity, and because of CIFOR's ownership rights.

The Project Leader (PL) must familiarize themselves with all grant terms and conditions, including those relating to intellectual property and data responsibilities. The PL has a responsibility to cooperate in the deposit of research data to the CIFOR repository for archiving purposes.

Rights and ownership

Collaboration agreements with external organisations may have an impact on ownership of the copyright in research data. At the commencement of any collaborative research project, parties should reach an agreement on matters of ownership. The Letter of Agreement is CIFOR's standard instrument for this agreement.

This agreement, which should be in writing, should cover the following issues:
▪ authorship
▪ ownership of any intellectual property that is produced
▪ ownership of copyrighted material.
▪ management of the research data, including roles and responsibilities, confidentiality, privacy, access, and
▪ dissemination of results.

An agreement will help resolve any conflicts over ownership that may arise later in the project. Further information on collaborative agreements can be found in the Letter of Agreement (LOA) Policy.

3.3. Project start-up

During the start-up phase (Annex 3), key information needs to be entered into relevant databases, and all administrative units that will be working on or with the project need to be notified of its existence and of their respective RDM roles.

Once a project code is received from Finance and Administration, the Project Management Office (PMO) will initiate an electronic project start-up form and email notification to the Project Leader (PL). The electronic project start-up form will be pre-filled with what information can be obtained from the project proposal and the donor agreement or contract.

Information that needs to be added by the PL will include:
▪ description of the research data and primary materials that will be acquired
▪ secondary materials that will be required (if any)
▪ level of access needed for workspace set-up (if needed)
▪ Confidentiality restrictions.

This information will be then recorded into a project data pipeline (Annex 8)

3.4. Project ongoing

These are the suggested steps to follow regarding RDM at the project ongoing stage:
1. Data collection and file creation:
   a. document all the relevant materials
   b. follow best practices
   c. organise files, back-ups and storage
2. Data analysis:
   a. maintain a master version of the dataset
   b. establish documentation versions similar to those used for the data
   c. keep a separate change file that tracks changes to the documentation
   d. conduct a review of the final files to make sure the data and documentation are harmonised and correspond to the final version of the data.
3. Data or materials must not be removed from the storage provided without authorisation during the active phase.
4. Sensitive records/data must be appropriately protected from unauthorised access.

Center for International Forestry Research
A CGIAR Consortium Research Center

Researchers must ensure that all research data, regardless of format, is stored securely and backed up or copied regularly. The data must be stored on network drives managed by IT staff. Personal computers, laptops and external storage devices such as USBs, CDs and DVDs are not recommended for research data storage.

File format
Researchers should record information about which file format(s) they expect to use to capture and store their research data, and which software programs will be used to create and manipulate the data and retrieve the data.

It is recommended that the file format should be:
- accessible in the future
- non-proprietary
- open, documented standard
- commonly used by the research community.

CIFOR will provide a StatTransfer for those who want to use the data and when the user would like to transfer of statistical data between different application/program. The responsibility to transfer is fall in to the user who want to utilize the data.

File naming
You should develop file naming conventions early in a research project, and agree these with colleagues and collaborators before data is created. A good documentation on file naming convention should be provided for further use.

Conventions will differ depending on the nature and size of a research project. In all cases, filenames should be unique, persistent and consistently applied, if they are to be useful for finding and retrieving data.

In deciding on digital file naming conventions, you should consider:
- Using a brief file names
- Using lower-case characters only - some computer operating systems are case-sensitive.
- Identify the project name
- Avoid special characters (e.g., \ / : * ? " < > | [ ] & $, .)
- Use underscores rather than spaces
- Include date of creation or modification in a standard format (e.g. YYYY_MM_DD or YYYYMMDD)

Data back-up
When considering your back-up strategy you need to know:
- whether all data, or only changed data, will be backed up
- how much hard-drive space will be required to maintain this back-up
- if the data is sensitive, how the data will be secured and (possibly) destroyed
- if data is password protected, where a file listing all passwords will be kept, and who will be allowed access to the passwords

IT will provide two back up methods for the data:
- In campus back up
- Off campus back up, using a third party service provider.

Data security
It is important to consider data security to prevent:
- accidental or malicious damage/modification to data

- theft of valuable data
- breach of confidentiality agreements and privacy laws
- premature release of data, which can void intellectual property claims
- release before data has been checked for accuracy and authenticity.

Research data stored in the CIFOR server will be held in a manner that allows for controlled access where this is required. Project Leader along with the project team, is responsible for controlling access to the data.

3.5. Project closure
These are the suggested steps to follow regarding RDM at the project closure stage:
1. Prepare data for archive
2. Assign relevant metadata
3. Comply with relevant standard and formats as suggested by CIFOR repository
4. Contact Data and Information Unit (cifor-rdm@cgiar.org) to get an assistance on the archiving process
5. Sensitive records/data will be appropriately protected from unauthorised access.

Where data has been obtained from limited access databases, or is owned by a third party, it may not be possible to store the data in the repository. In such cases: (i) a written description of the location of the original data, or (ii) key information regarding the limited access database from which it was extracted, must be recorded.

4. Submission of data
There should be no additional costs for archiving data, other than researcher(s) time to prepare data and documentation for deposit. This time should be budgeted, based on the prior consultation on the process.

4.1. Qualitative data
Types of qualitative data that may be archived include:
- in-depth/unstructured interviews transcription
- semi-structured and structured interviews transcription
- focus groups documentation
- unstructured or semi-structured diaries/log book
- observation field notes/technical fieldwork notes
- case study notes
- minutes of meetings

In order for qualitative data to be used in secondary analysis, it is important that the data is well documented. Any information that could provide context and clarity to a secondary user should be provided. Documentation for qualitative data should include:
- research methods and practices
- interview guide
  - o details on setting of interviews
  - o details on selection of interview subjects
  - o instructions given to interviewers
- data collection instruments such as interview questionnaires
- anonymisation guideline and documentation
- any problems that arose during the selection and/or interview process and how they were handled
- information on confidentiality, access and use conditions.
- transcription and digitization procedure and guideline

![Center for International Forestry Research - CIFOR logo]

Center for International Forestry Research
A CGIAR Consortium Research Center

Confidentiality and anonymisation

Before archiving the data, the depositor has an obligation to ensure that data is protected through anonymisation and/or the withdrawal of sensitive data.

Please ensure to do the following before archiving your data:

1. Establish anonymisation guideline
2. Make clear in the guideline about what to anonymise. The key question to be considered whether you should remove details such as address, places and occupation is whether or not there is a significant attachment that would link an individual clearly to one place and occupation, and thereby alllow them to be identified.
3. Replace details with descriptions that reflect the significance of the original text within the context of the transcript
4. Create a tracking table to record changes and to link real names with pseudonyms.

Highly sensitive information should be flagged and will be archived with the following options:
▪ closure of materials for a specified period, as agreed with the depositor, funder or partner;
▪ restricted access, where certain materials are only available to particular researchers, and/or where requests for access are granted by the depositor.

Transcription and digitization

1. Establish transcription procedures
   a. What to transcribe?
   b. Who should transcribe? Interviewer, professional transcription services, or the individual consultant hired specifically for the task?
2. Establish a transcription convention
3. Create a manual to guide the transcription process
4. For languages other than English (or the language of reporting), the transcription can be done in the original language with a summary of each interview in English (or the language of reporting).

4.2. Quantitative data

Project documentation for quantitative data that should be archived could include:
▪ context of data collection
▪ data collection methods
▪ structure and organisation of data files
▪ data sources used
▪ data validation, quality assurance
▪ transformations of data from the raw data to information through analysis
▪ methodology and tools (software, programs, etc.) used for analysis
▪ variable names, and descriptions
▪ explanation of codes and classification schemes used
▪ information on confidentiality, access and use conditions.

4.3. Spatial data

CIFOR's spatial data archive is managed in ArcGIS, a desktop application. Please send an email to cifor-rdm@cgiar.org to have the application installed in your desktop if you do not already have it.

4.4. Secondary data

For projects that do not involve original data collection or may involve combining data from one or more existing sources, here are some guidelines to consider in the decision of whether or what to archive:

- If the existing data used for analysis is not already publicly available, researchers are encouraged to submit the data for archiving, with the permission of the original data producer.
- If the researcher collects some primary data and also appends existing or secondary data to it for analyses, the guiding questions on whether to submit the whole dataset or just the primary dataset to the archive are: (a) how easily the existing data can be linked to the primary data, and (b) whether the existing data is publicly available.

Linked census data

When the primary data is linked to census data, the linked census data should be archived also, even though the link between the data files is straightforward and the census data is publicly available.

Since the original census files are large in size and contain a large number of variables, determining which census variables to use and at what level to extract the data for the subsets can be time-consuming. Archiving the linked census data makes it unnecessary for other users to repeat these sub-setting steps.

Straightforward link

If the link is straightforward and the existing data is publicly available, then users can easily obtain the existing data themselves and link them to the primary data submitted by the researcher. The source of the existing data (including version) should be clearly documented so that other users know which data to obtain. Information about the variable (or combination of variables) that constitutes the unique identifier used to link the data should also be provided.

Links that are not straightforward

If the link between datasets is not straightforward, then the linked data should be archived. Examples of this include: (a) link requiring judgements about combinations of non-unique variables, such as age, sex, and race; and (b) an understanding of local geographic factors is needed to link correctly, especially over a range of years when boundaries shift. Here, the redundancy of having data stored twice at the archive is outweighed by the usefulness of providing others with the data already linked.

Derived variables

Often, after the data is linked, the researcher may compute new variables based on the linked data (e.g. new categories are created, areas are extended, or scales are developed). All useful derived variables should be archived also, especially if they are used in analyses included in publications. The derived variables may be deposited in a data file that includes the primary data collected for the project and the existing data from another source, or the derived variables may be deposited with the primary data alone. The code or set-up file used to link the files and create the derived variables should also be provided.

Programming code

If the project involves only analysis of data already publicly available and the product of the project is the analysis alone, then data may not need to be submitted for archiving. However, researchers are encouraged to deposit their programming code that created new variables or scales, especially if the derived variables are not deposited within a data file and are cited in publications.

Documentation that is need to be submitted should include:
- project aims and objectives (to provide context)
- catalogue of data collected
- description of lifecycle of key data elements (procedures for collection/creation, validation, transformation, processing, analysis, publication, archiving/destruction)
- description of instruments, calibrations, etc.
- description of how data is structured (data model, coding schemes, controlled vocabularies)
- details of any quality control processes

- confidentiality agreements and consent forms (if any)
- manuals, code books and procedure documents.

5. Metadata

Metadata is structured information that describes data, information, services or other resources. The information in the metadata helps people to find, manage, control and understand their data assets and other resources.

Metadata is the key management mechanism for data information. It provides:
- discovery — enabling users to locate and evaluate information
- management — enabling custodians to better manage their data information
- utilisation — enabling users to access and manipulate information by means of automated/distributed systems.

It is critical to begin to document your data at the very beginning of your research project, even before data collection begins; doing so will make data documentation easier and reduce the likelihood that you will forget aspects of your data later in the research project.

Please refer to Annexes 6 and 7 for the minimum metadata component that you need to assign to your dataset.

6. Licensing

Licences can be exclusive or non-exclusive. An exclusive licence means that only the recipient of the licence has the right to use the data in the manner covered by the licence – to the exclusion of the copyright owner. A non-exclusive licence means that the recipient has the right to use the data in the manner covered by the licence – but not to the exclusion of the copyright owner, who is free to use the data in the same manner and to grant non-exclusive licence to others. Data owners generally grant non-exclusive licence to repositories and other end-users of their data.

Licences can also be:
- time limited
- limited to geographical regions
- limited to specific purposes
- contractual
- non-contractual.

7. Access and reuse of data

Whenever possible, research data should be made available to other researchers[2]. Therefore, the default position should be to allow sharing of the data, unless there are compelling reasons as to why this is not appropriate.

Sharing data does not necessarily mean making data openly and publicly available, although this is the default expectation of the CGIAR Intellectual Asset Principles. When sharing data, researchers have the option to determine access conditions and to set a desired level of access to their datasets

Options for sharing include:

---

[2] See CGIAR Intellectual Asset principles:
www.cgiarfund.org/sites/cgiarfund.org/files/Documents/PDF/fc7_cgiar_ia_principles_inclusion_COF_Feb16_2012.pdf

- Controlled/Mediated Access. This means that requests to access the data may be directed to a nominated contact, for example the Project Leader or a Data Manager. This option is also means sending the data directly to the requester.
- Restricted Access. Access to the data is restricted. However, the metadata is available to be browse.
- Open Access. The data is published and freely available in the public domain.

Access to research data may be provided through an access agreement. The agreement allows the data owners to specify the terms of access. For an access agreement, the following information would be required:

- description of the data
- identity of those who are permitted to access the data
- the limits of the access rights (e.g. what cannot be done with the data)
- a provision for consequences for failure to comply with the agreement.

To prevent a misuse of data, the following precautions will be applied:

- A specific citation for each data will be provided
- A permission statement will be provided
- Data user should provide information regarding proposed use and that they send a copy of any publication that results from use of the data.

A template of a data Access Agreement is included in Annex 9 of these Guidelines.

Annex 1. Proposal development

```
┌─────────────────────────────┐
│ Calls for proposals are     │
│ circulated within CIFOR.    │
│ The research is designed by │
│ the appointed team          │
└─────────────────────────────┘
              │
              ▼
        ╱───────────╲                    ╭─────────────────────╮
       ╱ Will the    ╲        No         │ Continue with the   │
      ╱ project       ╲ ───────────────▶ │ usual proposal      │
      ╲ acquire data? ╱                  │ development workflow │
       ╲             ╱                    ╰─────────────────────╯
        ╲───────────╱
              │ Yes
              ▼
        ╱───────────╲                    ┌─────────────────────────────┐
       ╱ Will the    ╲        Yes        │ Contact cifor-rdm@cgiar.org │
      ╱ project needs ╲ ───────────────▶ │ to check whether the data   │
      ╲ secondary     ╱                  │ is available at the         │
       ╲ data?       ╱                   │ repository                  │
        ╲───────────╱                    └─────────────────────────────┘
              │ No                                    │
              ▼                                       ▼
   ┌──────────────────────┐              ╱───────────╲
   │ Prepare cost         │     No      ╱ Is the data ╲
   │ allocation for data  │ ◀────────── ╲ available?  ╱
   │ acquisition and      │              ╲           ╱
   │ archiving            │               ╲─────────╱
   └──────────────────────┘                    │ Yes
       ▲                                        ▼
       │                                 ╱───────────╲
       │                    Yes/No      ╱ Is there any╲
       └──────────────────────────────── ╲ restrictions╱
                                          ╲ and/or     ╱
                                           ╲conditions ╱
                                            ╲ of re-use?╱
                                             ╲────────╱
```

Annex 2. Contract management

```
┌─────────────────────────────────┐
│        Proposal accepted        │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│   Understanding funding agency  │
│      requirement for data       │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│ Revisit the Partnership and     │
│ consultancy policy and guideline│
│ at My CIFOR page                │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│ Develop the rights and obligation│
│ of all parties with respect to  │
│ their roles and responsibilities│
│ for the management and retention│
│ of data                         │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│ Retain some ownership of data   │
│ for CIFOR                       │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│  Develop a Data Management Plan │
└─────────────────────────────────┘
```

Annex 3. Project start up

```
┌─────────────────────────────┐
│   PMO sends electronic email │
│  notification form to PL to  │
│   initiate project start up  │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│ PL fill in project start up  │
│ form about the data that the │
│     project will acquire     │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│ Project start up form        │
│ circulated to relevant       │
│ parties for tracking progress│
└─────────────────────────────┘
         │              │
         ▼              ▼
┌──────────────┐  ┌──────────────────┐
│ Project data │  │ Working space    │
│ pipeline     │  │ developed by IT. │
│ updated      │  │ PL needs to      │
│              │  │ advise IT on     │
│              │  │ access level and │
│              │  │ back up plan     │
└──────────────┘  └──────────────────┘
```

Annex 4. Project ongoing and tracking

```
Data acquired/created  →  PMO send an email notification remind you on the milestones  →  PL fill in the form to inform the latest status of your data

        ↓

Have you document all the relevant materials?  — No →  Document all the relevant materials and follow the best practices as instructed in guideline

        ↓

Have you inform IT to back up the data and documentation?  — No →  Inform IT to back up your data; manage your files by versioning

        ↓

Preparing your data for archive  →  Contact cifor-rdm@cgiar.org to assist with the archiving preparation
```

Annex 5. Project closure

Data
acquired/
created

↓

Check, validate, clean data
(anonymise data when necessary)

↓

Contact cifor-rdm@cgiar.org to get
assistance on the archiving process

↓

Assign relevant metadata

↓

Deposit the data

Annex 6. Data Documentation Initiative (DDI) metdata requirements – for social sciences data

Data citation details	
Metadata	**Description***
Title	Title of data collection. This element is required in the Study Description citation.
Study Global ID (Dataverse Catalog Number)	The unique identifier by which this study will be known. This archive's handle will be prefixed to the study identifier to make a unique system identifier
Author	Person, corporate body, or agency responsible for the work's substantive and intellectual content. Repeat the element for each author, and use affiliation attribute if available. Invert first and last name and use commas.
Author's Affiliation	Organization with which the author is affiliated.
Producer	The producer is the person or organization with the financial or administrative responsibility for the physical processes whereby the document was brought into existence. Use the role attribute to distinguish different stages of involvement in the production process, such as original producer.
Affiliation	Organization with which the producer is affiliated.
Abbreviation	Abbreviation by which the producer affiliation is commonly known.
URL	Producer URL points to the producer's web presence, if appropriate. Enter an absolute URL where the producer's web site is found, such as http://www.my.org
Logo URL	URL for the producer's logo, which points to this producer's web-accessible logo image. Enter an absolute URL where the producer's logo image is found, such as https://my.cifor.org/Documents/Editorial%20and%20Multimedia/CIFOR%20logos/CIFOR_green_print.eps
Production Date	Production or Published Date (if the distributor date is not filled-in, this date is used for the DVN study citation).
Funding Agency	Sources of funds for production of the work. If different funding agencies sponsored different stages of the production process, use the role attribute to distinguish them.
Distributor	Organization designated by the author or producer to generate copies of the particular work including any necessary editions or revisions.
Affiliation	Organization with which the distributor contact is affiliated.
Abbreviation	Abbreviation by which this distributor is commonly known.
URL (Enter full URL, http://...)	Distributor URL points to the distributor's web presence, if appropriate. Enter an absolute URL where the distributor's web site is found, such as http://www.my.org
Logo URL (Enter full url for image, http://...)	URL of the distributor's logo, which points to this distributor's web-accessible logo image. Enter an absolute URL where the distributor's logo image is found, such as https://my.cifor.org/Documents/Editorial%20and%20Multimedia/CIFOR%20logos/CIFOR_green_print.eps .
Contact	Contact information
Affiliation	Organization with which the distributor contact is affiliated
E-mail	Email address of the distributor contact.
Description and scope	
Description	An unformatted summary describing the purpose, nature, and scope of the data collection, special characteristics of its contents, major subject areas covered, and what questions the PIs attempted to answer when they conducted the study. A listing of major variables in the study is important here. In cases where a codebook contains more than one abstract (for example, one might be supplied by the data producer and another prepared by the data archive where the data are deposited), the source and date attributes may be used to distinguish the abstract versions. Inclusion of this element in the codebook is recommended. The date attribute should follow ISO convention of YYYY-MM-DD.
Keyword	Words or phrases that describe salient aspects of a data collection's content. Can be used for building keyword indexes and for classification and retrieval purposes. A controlled vocabulary can be employed. ** CIFOR and other CGIAR centre are using the CAB Thesaurus for controlled vocabulary
Related Material	Any related material
Related studies	Any studies that are relevant to this one, such as prior research on this subject
Date of Collection (Start and End)	Dates when the data were collected. Use the event attribute to specify start, end, or single for each date entered. The ISO standard for dates (YYYY-MM-DD) is recommended, although this form will accept YYYY or YYY-MM as well.
Country/Nation	Country where the data was collected. If more than one, they can be separated by commas

Geographic coverage	All countries included in the study should be listed here. Inclusion of this element in the codebook is recommended. Use this element to provide information on the areas that were covered, and highlight the reasons why the survey did not cover the entire geography of the country
Geographic unit	Information on the geographic coverage of the data. Inclusion of this element in the codebook is recommended.
Kind of data	Type of data included in the file: survey data, census/enumeration data, aggregate data, clinical data, event/transaction data, program source code, machine-readable text, administrative records data, experimental data, psychological test, textual data, coded textual, coded documents, time budget diaries, observation data/ratings, process-produced data, or other.
Term of use	
Confidentiality declaration	nature of any confidentiality or non-disclosure agreement signed before data access can be granted.
Special Permission	Used to determine if any special permission are required to access a resource.
Restriction	Any restriction on access to or use of the collection such as privacy certification or distribution restriction should be indicated here. This can be restriction applied by the author, producer, or disseminator of the data collection. If the data is restricted to only a certain class of user, specify which type
Contact	Detail contact person of whom the request should be addressed. Provide the information if this is different from the data distributor
Citation requirement	Text of requirement that data collection should be cited properly in articles or other publication that are based on analysis of the data
Conditions	Additional information that will assist the user in understanding the access and use condition of the data collection
Disclaimer	Information regarding responsibility for uses of the data collection

*The descriptions were taken verbatim from the Dataverse Network Project at Harvard: The Dataverse Network Project. Institute of Quantitative Social Science, Harvard <http://thedata.org/book/metadata-references>.

See **www.ddialliance.org**

Center for International Forestry Research
A CGIAR Consortium Research Center

Annex 7. Spatial metadata requirements

Name	Description	Obligation*
1. Content Citation	**Basic information required to uniquely identify a resource or resources**	
1.1 Title	Name by which the dataset is known. Free text	M
1.2 Publication date	Date identifies when the dataset was issued. Fill-in the year (and optionally month, or month and date) in the format of DD/MM/YYYY	M
1.3 Corresponding date	Date(s) identifies when the resource brought into existence or when the resource examined. Fill-in the year (and optionally month, or month and date).	C
1.4 Originator	Person or authoritative body primarily responsible for the intellectual content of the dataset.	
1.4.1 Name	Name of the responsible person	O
1.4.2 Organization	Name of the responsible organization	M
1.4.3 Homepage	Location for on-line accessing of the responsible organization using a Uniform Resource Locator address or similar addressing scheme. Example: http://www.cifor.cgiar.org/	O
1.4.4 Email	Address of the electronic mailbox of the responsible organization or the individual	O
1.4.5 Mailing address	Name of street and number or P.O. Box of the responsible organization	O
1.4.6 City	City of responsible organization located	M
1.4.7 Country	Country of responsible organization located	M
1.4.8 Telephone	Telephone number of responsible organization in a format of + (country code) - (area code) - (number)	O
1.4.9 Facsimile	Facsimile number of responsible organization in a format of + (country code) - (area code) - (number)	O
1.5 Publisher	Agent or agency responsible for making the dataset available in the current form	
1.5.1 Name	Name of the person in charge of publication	O
1.5.2 Organization	Name of the publisher	M
1.5.3 Homepage	Location for on-line accessing of the publisher using a Uniform Resource Locator address or similar addressing scheme	O
1.5.4 Email	Address of the electronic mailbox of the publisher or the individual	O
1.5.5 Mailing address	Name of street and number or P.O. Box of the publisher	O
1.5.6 City	City of publisher located	M
1.5.7 Country	Country of publisher located	M
1.5.8 Telephone	Telephone number by which individual can speak to the publisher of dataset or a contact person, in the format of + (country code) - (area code) - (number)	O
1.5.9 Facsimile	Telephone number of a facsimile machine of the publisher or a contact person in the format of + (country code) - (area code) - (number)	O
1.6 Metadata Contact	Person or organization who owns the dataset and primarily responsible of the metadata content	
1.61 Name	Name of the person who owns the dataset	O
1.62 Organization	Name of the organization who owns the dataset	M
1.63 Homepage	Location for on-line accessing of the owner of dataset using a Uniform Resource Locator address or similar addressing scheme	O
1.64 Email	Address of the electronic mailbox of the owner of the dataset or the individual	O
1.65 Mailing address	Name of street and number or P.O. Box of the owner	O
1.66 City	City of owner of dataset located	M
1.67 Country	Country of owner of the dataset located	M

1.68 Telephone	Telephone number by which individual can speak to the owner of dataset or a contact person, in the format of + (country code) - (area code) - (number)	O
1.69 Facsimile	Telephone number of a facsimile machine of the responsible organization or contact person in the format of + (country code) - (area code) - (number)	O
2 Content Description	Narrative information about the content and physical form of the dataset	
2.1 Abstract	Brief narrative summary of the content of the dataset. Free text.	M
2.2 Data presentation	Physical manifestation of the dataset	M
2.3 Specific content	Other information required to complete the citation that is not recorded elsewhere	O
2.4 Process	Information about the events, parameters and source data or technique used in constructing the dataset	O
2.5 Purpose	Brief narrative summary of intentions with which the dataset was developed	O
2.6 Supplemental information	Any other object or descriptive information that have direct or indirect relationship with the dataset	O
2.7 Content Status	Narrative information about the status and readiness of the dataset	
2.8 Status	Status of the dataset. Choose one from the following: ▪ Unknown ▪ Published ▪ Finished ▪ Draft ▪ In process	M
2.9 Maintenance	Information about maintenance and update frequency of the dataset. Choose on of the following: ▪ Unknown ▪ Monthly ▪ Continually ▪ Quarterly ▪ Daily ▪ Annually ▪ Weekly ▪ Bi-annually ▪ Fortnightly ▪ As needed ▪ Irregularly	M
3 Keyword information	Description of the dataset by using formalized system	
3.1 Category	Subject matter used to group dataset into similar or its main theme	M
3.2 Thesaurus	Name of the formally registered thesaurus or a similar authoritative source of keywords	M
3.3 Keywords	Commonly used word(s) or phrase(s) used to describe the dataset	M
4. Constrain information	Restriction on the access and use of the dataset	
4.1 Access	Access constrains applied to assure the protection of privacy or intellectual property, and any special restrictions or limitation on obtaining the dataset. Choose one of the following: ▪ No security restriction ▪ Copyright ▪ Unclassified ▪ Patent ▪ Restricted ▪ Trademark ▪ Confidential ▪ License ▪ Secret ▪ Intellectual Property Rights	M
4.2 Limitations	Brief description of limitation(s) affecting the fitness for use the dataset	M
5 Distribution Information	Information about distribution and option of obtaining the dataset	
5.1 Distribution Type	Provide description of the format and/or media by which the dataset can be obtained	O
5.2 On-line connection	Information about protocol and connection for on-line access	
5.2.1 Protocol	Connection protocol to be used	O
5.2.2 Connection string	File name and location of the dataset that can be used with on-line access	O
5.3 Off-line connection	Information about off-line media on which the dataset can be obtained	
5.3.1 Connection string	File name and location of the dataset that can be usde with off-line access	O
5.4 Transfer size	Estimated size of the dataset in a specified transfer format. Expresses in MegaByte	O
6 Spatial Representation Information	Information about spatial representation and reference system of the dataset	
6.1.Geographic extent	Description of geographic name or location within which data is available. Example: Asia, Borneo, etc	M

6.2 Spatial Domain	Information about spatial reference system used by the dataset	
6.2.1 Projection and Datum	Name of projection system and/or datum used to project the dataset into a plane	M
6.2.2 Bounding Box	Minimum bounding rectangle within which the data is available , expressed by set of coordinates of the polygon	
West coordinate	Western-most coordinate of the limit of the dataset extent	M
East coordinate	Eastern-most coordinate of the limit of the dataset extent	M
South coordinate	Southern-most coordinate of the limit of the dataset extent	M
North coordinate	Northern-most coordinate of the limit of the dataset extent	M
7 Data Quality	General assessment of the quality of the data set	
7.1 Horizontal accuracy	Estimate of accuracy of the horizontal positions of the spatial objects	O
7.2 Vertical accuracy	Estimate of accuracy of the vertical positions of the spatial objects	O
7.3 Scale	Scale of the source map	O
7.4 Cloud cover	Percentage of cloud coverage in a dataset	O
8 CIFOR Information	Specific information for CIFOR about which project who use this dataset as an input or produce the data as an output of a project and any other internal related information.	
8.1 Input	Projects who use the data as an input of its project activity. This could be obtained/purchase from third party or other projects at CIFOR	
8.1.1 Project name	Name of the project	M
8.1.2 Project description	Description of the project	O
8.2 Output	Project who produce the dataset as an output of its research activity	
8.2.1 Project name	Name of the project	M
8.2.2 Project description	Description of the project	O

* Note: Obligation: M = Mandatory; O = Optional; C = choice between 2 options, at least 1 option is mandatory and must be documented.

Center for International Forestry Research

A CGIAR Consortium Research Center

Annex 8. Research data pipeline

Project code	Title of the project	Funding agency	Name of the data set	Description of the data	Data manager/ author	Licensing and re-use conditions
PC	The official name of the project	The organisation who fund the project	Official name	Basic description of the purpose and content of the dataset	Name and/or position, e.g. Project Manager	Information about embargoes, license restrictions, conditions of re-use

Center for International Forestry Research

A CGIAR Consortium Research Center

Annex 9. Sample Access Agreement

Research Data Access Agreement

CIFOR grants the access to the data set mentioned below solely for the purpose specified. The data recipient will:

1. take all necessary precaution to ensure that the data are used only in the manner stated and for the research purposed specified
2. will exercise proper safeguards to prevent any breach of confidentiality or privacy
3. will not use or disclose "data set" or "information" for any purpose other than the research project identified below and will not distribute any part of the data set to anyone who is not part of the data recipient's research team

Title of the data set to which access is required

_____

_____

_____

Title of the research project

_____

_____

_____

Proposed uses of data set

_____

_____

_____

[Data recipient information]
Name   :
Organisation   :
Address of organisation   :
Position   :
Email address   :
Date   :

[CIFOR approval information]
Name   :
Position   :
Email address   :
Date   :