RESEARCH PROGRAM ON
Forests, Trees and Agroforestry

CGIAR

SPATIAL METADATA

Guidelines and Procedures

(July 2021)

These guidelines and procedures has been developed consistent with the CGIAR Principles on the Management of Intellectual Assets and by the CGIAR's commitment to Open Access, and explain how to implement spatial metadata practices. This document provides guidelines and procedures on developing spatial metadata – from entering metadata to providing spatial metadata documentation.

foreststreesagroforestry.org

SCOPE

This is a supplementary document and provides practical information to better understand and implement the spatial metadata profile specified in the CIFOR RDM Guidelines and Procedures (15 January 2020).

Spatial metadata or geographic metadata can be defined as metadata applicable to geographic data and information. This metadata can be stored along with geospatial data (vector and raster format) or may be listed in a separate document. There are many standards for geospatial metadata, including the International Organization for Standardization (ISO) and Federal Geographic Data Committee (FGDC) metadata standards. This document adopts (ISO) 19115-1:2014 – Geographic Information - Metadata as the standard for developing geographic information or spatial metadata. According to ISO, core elements of information that should be provided in spatial metadata are as follows:

1. **Spatial representation**, this information supports the provision of metadata identifying the mechanism of modelling real world phenomena in a digital dataset.
2. **Lineage,** this information supports the provision of metadata concerning the sources and production processes used in producing a dataset or resources.
3. **Distribution information**, this record supports the provision of metadata about the distributor and options for obtaining a resource.
4. **Content information**, this record supports the provision of metadata identifying the content of a resource.
5. **Constraint information**, this record supports the provision of metadata concerning the legal and security constraints placed on resources and metadata about resources.
6. **Reference system information**, this data supports the metadata identifying the spatial, temporal, and parametric reference system(s) used by a geographical resource.

ROLES AND RESPONSIBILITIES

Project leaders

- Set the culture of practice based on guidance, and delegate specific responsibilities as appropriate;
- Are accountable for ensuring that, should they leave their organization during the life of the project, the data stays with the project.

Researchers

- Maintain records of geospatial metadata and ensure these records and the geospatial metadata are securely stored;
- Make the spatial metadata available to other researchers via open or negotiated access, as appropriate and in accordance with the requirements of research funding bodies and CGIAR Open Access and Data Management Policy;
- Ensure that, where projects span several institutions, an agreement is developed at the outset covering the ownership and storage of geospatial metadata within each institution in accordance with CGIAR Open Access and Data Management Policy;
- Ensure that adequate back-up, archival and monitoring strategies are in place to prevent the loss of geospatial metadata, and associated delays in completing research.

GIS specialists (metadata creators)

- As subject specialists, must understand the technical framework underlying spatial metadata records development;
- Responsible for creating and maintaining the metadata and for its quality;
- Oversee and implement geospatial metadata development procedures;
- Ensure overall validity of the geospatial metadata and information gathered to achieve a complete metadata profile.

Metadata managers (metadata validators)

- Compile and validate metadata that are stored in data or separate documents to ensure metadata follow rules and metadata profiles;
- Upload and generate services from data and provide information about meta-service for accessing and downloading data – (if data is approved for publishing in the geoportal).

Geospatial metadata users

- Responsible and accountable for all data access made through their user accounts and the subsequent use and distribution of the metadata.

SPATIAL METADATA ENTRY

It is mandatory for GIS specialists to create metadata for geospatial data during data production. The initial step for metadata creation is to enter information by using a Metadata Entry Tool (MET) such as ArcCatalog or CatMDedit. MET provides the ability to create, read, update and delete metadata records for the geospatial data based on the ISO metadata standard and profile. Following the rules for creating metadata (Annex 1) and core metadata profile (Annex 2), suggested spatial metadata entry procedures are as follows:

a. Check the geospatial data and determine data type (standalone or time series);
b. Understand the metadata profile and framework of the metadata standard, including mandatory and optional elements;
c. Collect additional information that should be essential for completing the metadata profile prior to metadata entry;
d. Enter metadata based on the profile and fill in information for mandatory and optional components. Mandatory elements to be filled in are as follows:
    - Title, abstract, and data presentation;
    - Date of creation and publication;
    - Geographical extent and reference system;
    - Keywords, topic, theme, flagship, priorities, and project relating to geospatial data;
    - Distribution, responsible person, organization, and contact information;
    - Data constraint, limitation, status, and maintenance information;
    - Sources and procedures for creating the geospatial dataset (lineage).

e. Recheck metadata entry and generate metadata in xml format.

   In order to carry out metadata profile validation and compliance procedures, xml files are important for validation against the ISO 19115 standard.

SPATIAL METADATA VALIDATION

Metadata validation is the testing and checking of metadata to ensure compliance with the standard's requirements. MET provides the facility for validating metadata based on a selected scheme, such as ISO 19115. The metadata creator and/or metadata validator is responsible for conducting metadata validation. Suggested steps for validating metadata are as follows:

1. Select the scheme or standard to be used for metadata validation. METs such as CatMDedit provide several schemes, but it is recommended you select ISO 19115.
2. Validate the item's metadata with the metadata style's XML scheme to ensure the metadata complies with the standard.
3. If any errors are reported in the tool's messages, identify the problem. If a validation error indicates there is a problem with a metadata profile (Annex 2), look up the appropriate section in the metadata standard document.
4. Edit the item's metadata. Add any missing information and correct any values that have the wrong data entry. Recommendations for correcting errors can be found in Annex 3.
5. Click the Save Metadata Edits button to save changes and stop editing the item's metadata.

Repeat steps 1 to 5 until no validation error messages are reported.

SPATIAL METADATA MAINTENANCE AND DOCUMENTATION

On an ongoing basis, ensure the metadata is adequately maintained and kept up to date. Documentation is required if any interventions and/or improvements are made to the geospatial metadata during maintenance processes. If the metadata is created and stored separately from the dataset, the metadata file's storage location should be documented to ensure it can be accessed and displayed together with the spatial data.

Annex 1. General rules for creating spatial metadata

Rule	Description
Free text	To provide easily understood information for potential users of CGIAR FTA sources, free text entries are available for some metadata fields, such as abstract, lineage and description of data quality. Hence, for some data elements, such as theme and FTA's flagship, code lists are specified.
Obligation	Obligations for data elements are mandatory or optional. Mandatory elements must be completed, whereas doing so is not always necessary for optional elements. The optional status may be applicable for several reasons. For instance, if the element is not relevant, or if its value is unknown.
Element domain	Some metadata elements have a range of allowable values called domains. For example, range of elevation (positive integer).
Spatial references	Spatial references are the key for geographic data presentation in the FTA geoportal, and the basis for searching by location. A consistent set of spatial references enables data for searching.
Date fields	Dates of data, including creation and publication dates are the main information determining the data's lifecycle.

Annex 2. Metadata profiles

M = Mandatory O = Optional V = Variable

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). Use the format DD/MM/YYYY	M
1.3 Corresponding date	Date(s) identifies when the resource was brought into existence or when the resource was examined. Fill in the year (and optionally month, or month and date). Use the format DD/MM/YYYY	O
1.4 Originator	Person or authoritative body primarily responsible for the intellectual content of the dataset	V
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 online 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 individual	O
1.4.5 Mailing address	Street name and number or PO Box of the responsible organization	O
1.4.6 City	City where the responsible organization is located	M
1.4.7 Country	Country where the responsible organization is located	M
1.4.8 Telephone	Telephone number of the responsible organization using the format: + (country code) -- (area code) -- (number)	O
1.4.9 Facsimile	Facsimile number of the responsible organization using the format: + (country code) -- (area code) -- (number)	O
1.5 Publisher	Agent or agency responsible for making the dataset available in the current form	V
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 accessing the publisher online using a Uniform Resource Locator address or similar addressing scheme	O
1.5.4 Email	Address of the electronic mailbox of the publisher or individual	O
1.5.5 Mailing address	Street name and number or PO Box of the publisher	O
1.5.6 City	City where the publisher is located	M
1.5.7 Country	Country where the publisher is located	M
1.5.8 Telephone	Telephone number through which an individual can speak to the dataset publisher or a contact person, using the format: + (country code) -- (area code) -- (number)	O
1.5.9 Facsimile	Facsimile number of the dataset publisher or a contact person using the format: + (country code) -- (area code) -- (number)	O

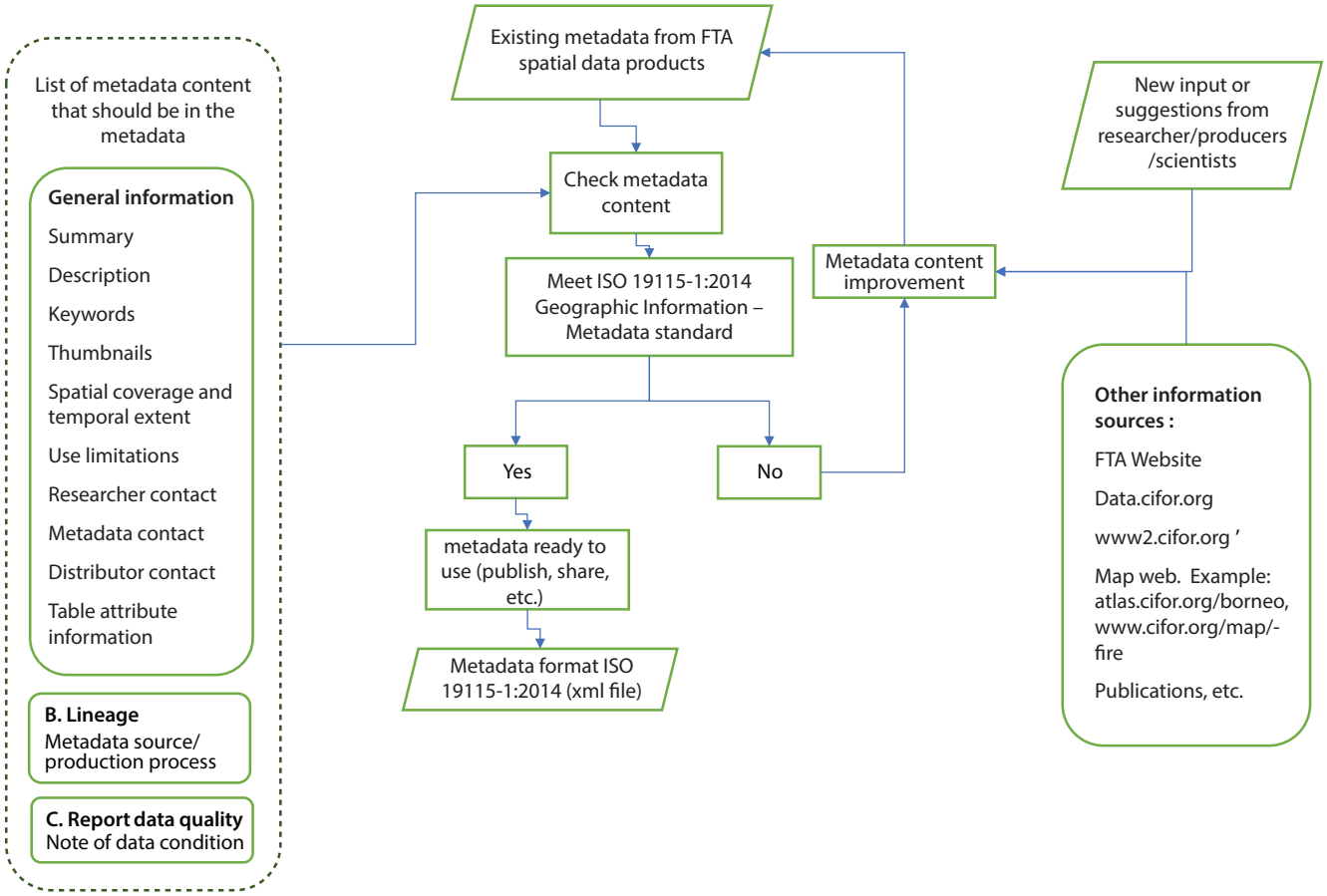
Name	Description	Obligation
1.6 Metadata contact	Person or organization who owns the dataset and is primarily responsible for the metadata content	V
1.61 Name	Name of the person who owns the dataset	O
1.62 Organization	Name of the organization that owns the dataset	M
1.63 Homepage	Location for accessing the dataset owner online using a Uniform Resource Locator address or similar addressing scheme	O
1.64 Email	Address of the electronic mailbox of the organization or individual that owns the dataset	O
1.65 Mailing address	Street name and number or PO Box of the dataset owner	O
1.66 City	City where dataset owner is located	M
1.67 Country	Country where dataset owner is located	M
1.68 Telephone	Telephone number through which an individual can speak to the dataset owner or a contact person, using the format: + (country code) -- (area code) -- (number)	O
1.69 Facsimile	Facsimile number of the dataset owner or a contact person using the format: + (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	V
2.8 Status	Status of the dataset. Choose one of the following: • Unknown • Published • Finished • Draft • In process	M
2.9 Maintenance	Information about dataset maintenance and update frequency. Choose one 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
3.4 Theme (New)	Theme information of dataset	O
3.5 FTA flagship (New)	FTA flagship information about the dataset	O
3.6 FTA priority (New)	Information on FTA priority	O

Name	Description	Obligation
4. Constraint information	***Information on restrictions on dataset access and use***	
4.1 Access	Access restrictions applied to ensure protection of privacy or intellectual property, and any special restrictions or limitations 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 fitness for dataset use	M
5 Distribution information	***Information about distribution and options for obtaining the dataset***	
5.1 Distribution type	Provide a description of the format and/or media by which the dataset can be obtained	O
5.2 Online connection	Information about protocol and connection for online access	V
5.2.1 Protocol	Connection protocol to be used	O
5.2.2 Connection string	File name and location of the dataset for online access	O
5.3 Offline connection	Information about offline media through which the dataset can be obtained	V
5.3.1 Connection string	File name and location of the dataset for offline access	O
5.4 Transfer size	Estimated size of the dataset in a specified transfer format. Expressed in megabytes	O
6 Spatial representation information	***Information about the dataset's spatial representation and reference system***	
6.1. Geographic extent	Description of the geographic name or location for which data is available. Example: Asia, Borneo, etc.	M
6.2 Spatial domain	Information about the spatial reference system used for the dataset	V
6.2.1 Projection and datum	Name of the 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 a set of coordinates for the polygon	V
West coordinate	The westernmost coordinate of the limit of the dataset extent	M
East coordinate	The easternmost coordinate of the limit of the dataset extent	M
South coordinate	The southernmost coordinate of the limit of the dataset extent	M
North coordinate	The northernmost coordinate of the limit of the dataset extent	M
7 Data quality	***General assessment of the quality of the dataset***	
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 cover in the dataset	O
7.5 Lineage (New)	Sources and procedures used for creating the geospatial dataset	M
8 CIFOR-ICRAF information	***Specific information for CIFOR about projects using the dataset as input, or producing the data as a project output, and any other internally-related information***	
8.1 Input	Projects using the data as input for project activities. This could be obtained/ purchased from a third party or from other projects	V
8.1.1 Project name	Name of the project	M
8.1.2 Project description	Description of the project	O
8.2 Output	Project producing the dataset as output from its research activities	V
8.2.1 Project name	Name of the project	M
8.2.2 Project description	Description of the project	O

Annex 3. Error prevention and correction

Common errors	Description	Example of error	Guidance for preventing and correcting errors
Metadata or documentation too general	The metadata lacks detail despite covering diverse topics or themes that require detailed information.	A topographic map containing a variety of different scales and specifications is documented in only one general metadata record.	• Metadata creators should read procedures for filling in metadata carefully, so they conform with the standard. • Metadata validators should check and recheck to ensure consistency between the data and metadata. • Where possible, metadata creators should collect the additional information necessary for ensuring metadata completeness.
Extent over-generalized	The extent of the dataset has been standardized with names of regions or countries, but still contains inconsistencies in its use of names covering the extent of the data.	Coverage of Borneo is referred to as Kalimantan. Data only covering Indonesia is written as world coverage.	• There should be clear rules on naming the extent of a dataset, including instructions for naming data covering specific regions. • Metadata validators should check for consistency. • Metadata creators should correct any errors and follow guidelines.
Dataset theme(s) or topic(s) underreported or not reported at all	The metadata still contains inconsistencies or errors in determining the individual topic(s) or theme(s) of the data. *Determining the topic(s)/ theme(s) of data is vital for dataset categorization.	Errors in providing theme(s)/topic(s) – for instance, fire data is assigned a health theme. Land cover data has deficiencies in reporting data boundaries.	• Have clear instructions on determining themes or topics. • Use closed lists to reduce errors when determining theme(s)/topic(s). • Metadata creators should correct any erroneous theme(s)/topic(s) as appropriate.
Incorrect or inconsistent date entries	Inconsistencies in date entry	Date of publication is inconsistent or erroneous	• Metadata validators should recheck dates to ensure consistency. • Metadata creators should correct any errors in filling in dates.
The same items are given different names	Name or data descriptor is inconsistent	Keywords for the same items are written differently (e.g., sentinel landscape and sentinel-landscape).	• Have instructions on standard nomenclature. • Checks and rechecks are needed to ensure consistency. • Correct erroneous names/descriptors in the metadata by following the instructions.

Annex 4. Metadata check and validation workflow.



RESEARCH PROGRAM ON Forests, Trees and Agroforestry
CGIAR

The CGIAR Research Program on Forests, Trees and Agroforestry (FTA). FTA is the world's largest research for development program to enhance the role of forests, trees and agroforestry in sustainable development and food security and to address climate change. CIFOR leads FTA in partnership with ICRAF, the Alliance of Bioversity International and CIAT, CATIE, CIRAD, INBAR and TBI.

FTA's work is supported by the CGIAR Trust Fund: cgiar.org/funders/