

# CGIAR Open and FAIR Data Assets Policy

Approved by the System Management Board with effect from 16 April 2021 (Decision reference SB/M19/EDP3). This Policy supersedes and replaces in its entirety the 2013 Open Access and Data Management Policy. It addresses funder and publisher requirements, the recommendations of a 2018 external assessment undertaken pursuant to the recommendation of the then CGIAR Independent Evaluation Arrangement and responds to a 2020 Data Management Maturity Assessment.

## Table of contents

1. CGIAR’s Commitment.....	3
2. Purpose .....	3
3. Scope.....	4
4. Policy.....	4
4.1 Management of data assets according to FAIR Data Principles.....	4
4.2 Responsible management to safeguard data assets, including those containing personally identifiable information.....	6
5. Implementation .....	6
Annex 1 Relevant Defined Terms .....	7
Annex 2 Embargo periods: Timeframes for open and FAIR data assets.....	8

## 1. CGIAR’s Commitment

- 1.1 As set out in the [CGIAR Principles on the Management of Intellectual Assets](#), CGIAR adheres to the principle that the results of its research and development activities are international public goods and is committed to their widespread dissemination and use to achieve the maximum impact to advantage the poor, especially smallholder farmers in developing countries. CGIAR considers Open Access (as defined in Annex 1) to be an important practical application of this commitment as it enhances the visibility, accessibility and impact of its research and development activities. CGIAR is therefore committed to sharing outputs of its research that are as open as possible and always Findable, Accessible, Interoperable and Reusable (FAIR)<sup>1</sup>, advancing CGIAR’s aspirations to digital transformation and data-driven innovation.
- 1.2 Open and FAIR data assets improve the speed, efficiency, and efficacy of research; they facilitate interdisciplinary research; assist data aggregation, computation, and the derivation of new insights; and allow the global public to benefit from CGIAR Research. They enable CGIAR to collectively leverage the infrastructure, data pools, and new data science capacities necessary for innovation and for effective and agile responses to global challenges. They facilitate text and data mining and analysis to derive insights, recognizing that these are dependent not only on access to high quality data, but also on that data being well-contextualized (through rich metadata and relevant open materials), interoperable, and reusable.
- 1.3 CGIAR is also committed to managing data assets responsibly, with regard to privacy, ethical, and any other nationally or internationally determined considerations, in accordance with the [CGIAR Research Ethics Code](#).

## 2. Purpose

The purpose of the CGIAR Open and Fair Data Assets Policy (“this Policy”) is to clarify expectations regarding the management and dissemination of data assets (as defined in Annex 1) to ensure that they are as open as possible, always FAIR, and managed responsibly. This Policy should be read in conjunction with the CGIAR Open and FAIR Data Assets Implementation Guidelines which may be updated from time to time to reflect current recommended practices<sup>2</sup>.

---

<sup>1</sup> Mark D. Wilkinson et al., ‘The FAIR Guiding Principles for Scientific Data Management and Stewardship’, *Scientific Data* 3, no. 1 (15 March 2016): 160018, <https://doi.org/10.1038/sdata.2016.18>.

<sup>2</sup> Currently under development.

### 3. Scope

- 3.1 This Policy applies to the management and dissemination of data assets (as defined in Annex 1) produced by CGIAR entities<sup>3</sup> and CGIAR implementing partners in the context of CGIAR Research<sup>4</sup>, and by CGIAR's advisory bodies.
- 3.2 Indicative types of data assets include peer-reviewed journal articles; non-reviewed articles, reports, briefs, extension and training and other materials; books and book chapters; data and databases; data collection and analysis tools (e.g. models and survey tools); video, audio and images; computer software (e.g. models, APIs); web tools (e.g. data portals, dashboards). This Policy also applies to the metadata associated with data assets.

### 4. Policy

#### 4.1 Management of data assets according to FAIR Data Principles

Data assets shall be managed in compliance with the FAIR Data Principles<sup>2</sup>, as appropriate for the particular data asset type and as elaborated upon below, per agreed targets<sup>5</sup>.

##### a) Findability: data assets shall be made findable within and beyond CGIAR.

- i) The permanence of institutional repositories shall be maintained and, barring extenuating circumstances, data assets and metadata shall be curated and kept up-to-date therein with permanent, stable identifiers, and in keeping with the embargo periods and timeframes set out in Annex 2.
- ii) All repositories used shall be stable, permanent, Open Access and offer application programming interfaces (APIs) or similar means to enable users, web sites and search engines to locate assets or metadata.
- iii) Metadata to describe data assets shall conform to the [CGIAR Core metadata schema](#) (based on the Dublin Core standard schema) as revised from time to time in consultation with CGIAR data and information managers and other relevant stakeholders.
- iv) Metadata for all data assets, regardless of their accessibility, shall be published in the relevant CGIAR institutional repository (per the embargo period set out in Annex 2).
- v) Data Management Plans shall address how data assets will be made Findable, and in general, managed throughout their life cycle.

---

<sup>3</sup> CGIAR entities include all CGIAR Research Centers and the CGIAR System Organization.

<sup>4</sup> As defined by the CGIAR System Framework and the Charter of the CGIAR System Organization, as amended from time to time. On the date of approval of this Policy, "CGIAR Research" is defined as "the research carried out by the Centers and the CGIAR System Partners in support of the CGIAR Strategy and Results Framework.

<sup>5</sup> In development.

**b) Accessibility: access levels to data assets shall be consciously considered according to content and made as open as possible.**

- i) Data assets shall be published Open Access (per the embargo periods in Annex 2) with default permissive Creative Commons (or similar standard) machine-readable licenses (e.g. [CC 0](#) or [CC BY](#)), except where the circumstances in ii) below are satisfied. [MIT](#) or [GNU](#) licenses shall be applied for software, algorithms, databases etc. Metadata shall record the license delineating the use of data assets.
- ii) Data assets may only be published without Open Access if they are associated with a Restricted Use Agreement or Limited Exclusivity Agreement (both as defined in Annex 1), or other contractual obligation to retain confidentiality. In such instances, the associated metadata shall be published in a CGIAR repository.
- iii) Translations of key documents and other media into pertinent languages are encouraged. All versions shall be deposited in suitable open repositories.
- iv) Data assets that require minimal data download to view and use are encouraged to assist those with limited internet connectivity.
- v) Ways in which users may request access to the data assets shall be clearly and openly published alongside the metadata. This access protocol shall conform to a reliable, timely process and decision-making will be made transparent to potential users.
- vi) Creators shall justify an embargo extension<sup>6</sup> to keep data assets closed beyond the embargo period.

**c) Interoperability: data assets shall be made easy to interpret by humans and machines.**

- i) Relevant data assets (e.g. datasets) and metadata shall be interoperable and fit for reuse.
- ii) Data Management Plans shall be developed at project formulation to ensure that data assets conform to accepted interoperability standards (e.g. ontologies, controlled vocabularies).
- iii) FAIR data workflows shall be used to make high-value legacy data interoperable. Where possible, data collection shall be digital, using tools that conform to interoperability standards.

**d) Reusability: the conditions for reuse of data assets shall be clearly established upon publication.**

- i) Machine-readable permissive licenses shall be applied to data assets, software, and allied products (per Section 4.1.b.i).
- ii) Open file formats shall be used to the maximum extent possible to archive data assets.
- iii) Metadata shall record data lineage, including versioning.

---

<sup>6</sup> Criteria to extend embargo periods for data assets are in development.

#### **4.2 Responsible management to safeguard data assets, including those containing personally identifiable information.**

Data assets shall be managed responsibly, with due regard to privacy and ethical considerations in accordance with the relevant sections<sup>7</sup> of the [CGIAR Research Ethics Code](#) and relevant policies on personal data protection<sup>8</sup>.

### **5. Implementation<sup>9</sup>**

- 5.1 This Policy and the CGIAR Open and FAIR Data Assets Implementation Guidelines shall be publicly available on the CGIAR website. All relevant CGIAR staff shall be made aware of these documents through periodic updates or trainings, and they shall be incorporated into relevant induction material.
- 5.2 Appropriate arrangements shall be put in place to ensure the efficient and consistent implementation of this Policy, including but not limited to:
- a) clarifying associated responsibilities of research delivery leadership;
  - b) incentivizing, rewarding, monitoring and reporting of FAIR and open data assets, ensuring integration with discovery platforms and reporting frameworks (e.g. GARDIAN, MARLO and MEL);
  - c) ensuring the allocation of an appropriate proportion of project budgets to ensure FAIR and open data assets, and, to the extent funders allow, as pooled resources or other mechanisms to sustain FAIR and open data management beyond the life cycle of relevant projects;
  - d) handling deviations from this Policy and embargo period extension requests.
- 5.3 The effectiveness of this Policy shall be reviewed periodically and at least once every six years, through consultation and feedback from relevant stakeholders.

---

<sup>7</sup> At the date of adoption of this Policy, the relevant section of the [CGIAR Research Ethics Code](#) is Section 4.1.

<sup>8</sup> Currently under development.

<sup>9</sup> At the time of adoption of this Policy, the CGIAR Managing Director, Research Delivery and Impact shall be responsible for Policy implementation as set forth in Section 5.

## Annex 1 Relevant Defined Terms

Certain terms and concepts used in the CGIAR Open and Fair Data Assets Policy are defined below to ensure their shared understanding across the System.

- i. Data assets: Intellectual Assets that are raw or manipulated datasets and information products produced by CGIAR, as outputs of CGIAR's research and development activities. These may include peer-reviewed and non-reviewed materials including journal articles, reports, briefs, extension, training and e-learning content and other materials, books and book chapters, data and databases, data collection and analysis tools (e.g. models and survey tools), video, audio and images, graphics, maps, and other GIS outputs, computer software, models and code, digital and mobile applications, and web-based services (e.g. websites, data portals, online platforms).
- ii. Consent: As defined in the [CGIAR Research Ethics Code](#).
- iii. Creator: The individual/team who collected or developed the data asset.
- iv. Database: An organized collection of data stored as multiple datasets, that are generally stored and accessed electronically from a computer system that allows the data to be easily accessed, manipulated, and updated.
- v. Data lineage: The history of a data asset as it is used and reused, including versioning.
- vi. Data management: A process that includes acquiring, validating, storing, protecting, and processing data to ensure accessibility, reliability, and timeliness of the data for its users.
- vii. Embargo period: The duration of time in which a given data asset may remain outside of the public domain.
- viii. Human subject: As defined in Annex I of the [CGIAR Research Ethics Code](#).
- ix. License: Agreed legal framework for the exchange of data assets between two or more parties, the permitted use of data assets for a project, the access terms for those data assets and associated commercial use, conferring the right of use without which such use could be illegal.
- x. Limited Exclusivity Agreement: As defined in the [CGIAR Principles on the Management of Intellectual Assets](#).
- xi. Metadata: Data describing data assets to make finding and working with data assets easier.
- xii. Open Access: A set of principles and a range of practices through which any user online has immediate, irrevocable, unrestricted and free online access to research outputs, with legal clarity regarding re-use of the data assets (which could be granted subject to appropriate licenses and proper attribution).
- xiii. Personal data/personally identifiable information (PII): As defined in Annex I of the [CGIAR Research Ethics Code](#).
- xiv. Repository: An archive for collecting, preserving, and disseminating digital copies of the intellectual output of an institution, particularly a research institution.
- xv. Restricted Use Agreement: As defined in the [CGIAR Principles on the Management of Intellectual Assets](#).

## Annex 2

### Embargo periods: Timeframes for open and FAIR data assets

The following list of data assets are indicative, not exhaustive.

- i. Peer-reviewed journal articles: Peer-reviewed versions of scholarly articles reporting research should be deposited in a suitable repository and made Open and FAIR as soon as possible, ideally at the time of publication, and no later than 6 months from the date of publication. Authors are free to choose the journal that is most appropriate. Where an author publishes in a closed access journal, he/she shall self-archive in an Open Access repository a digital version of the final accepted manuscript (the “postprint” version), to the extent permitted by the journal.
- ii. Reports and other papers: Information products that are not intended for peer-review journals, such as reports, conference papers, policy briefs and working papers, shall be deposited in suitable repositories and made Open and FAIR as soon as possible and in any event within 3 months of their completion.
- iii. Books and book chapters: The full digital version of books and book chapters shall be made Open and FAIR as soon as possible after publication and in any event within 6 months either through self-archiving or other suitable publication arrangements.
- iv. Datasets and databases: Datasets (and any relevant data collection and analysis tools) shall, subject to any additional donor requirements, be deposited in a suitable repository and made Open and FAIR as soon as possible and in any event within 12 months of completion of the data collection or appropriate project milestone, or within 6 months of publication of the information products underpinned by that data, whichever is sooner. Data deposited shall be prepared in a manner consistent with the aims of this Policy. Existing and future databases shall be made Open and FAIR.
- v. Video, audio and images: Complete final digital versions of video and audio outputs, and image collections must be stored appropriately and made Open and Findable, Accessible, and Reusable within 3 months of their completion.
- vi. Computer software: Where a data asset is software developed internally, the associated source code must be deposited in a free/open software archive upon completion of the software development unless doing so would be incompatible with third-party rights. Access to such information products may be granted subject to appropriate licences (e.g. Copyleft).
- vii. Metadata for all data assets: regardless of accessibility. The metadata associated with a data asset must be deposited in a suitable repository before or on publication of the products themselves, shall be published in data asset. If the relevant CGIAR institutional asset cannot be widely shared, its metadata must be deposited in a suitable repository within 6 months after completion of project for which they were collected.