**TERMS OF REFERENCE**

**Data Consultant**

**Background**

The environment in which data is generated, used, and exchanged for research purposes is changing rapidly throughout the world. Governments, charitable foundations, scientific organizations, curators, funders, and other users of research are changing the policies, regulations, and tools that govern and manage data. Many are pursuing policies to make data freely available and easily accessible so that they can be used without restriction, while others are focusing on the protection of privacy and prevention of data misuse.

These issues are particularly important to the International Food Policy Research Institute (IFPRI), which produces, shares, and uses data in multiple forms to provide research-based policy solutions to sustainably reduce poverty and end hunger and malnutrition in developing countries. Data forms the main ingredients of any research processes that the institute undertakes. Open data, data that are free to access, use and share, are critical to achieving the Institute’s vision of a world without hunger and malnutrition.

IFPRI has started sharing some of the data the Institute has produced since the 1990s. The institute has a data repository that allows users to download data easily and freely. IFPRI has codified these efforts in its 2010 data policy. IFPRI has also endorsed the Open Access and Open Data Policy of the CGIAR (<http://www.cgiar.org/resources/open-access/>), of which IFPRI is a member. While policies are in place, these policies have relied on individual initiative to make data freely and openly accessible as public goods. As a result, a considerable number of datasets may not have been made publicly available in line with the Institute’s policy.

On the other hand, donors are enhancing and consolidating their efforts around open data and imposing stricter standards and policies for data management, data publishing and sharing, and making data available as public goods. For example, the Bill & Melinda Gates Foundation (BMGF) requires making data public immediately — as soon as the data collection ceases in the field — unless otherwise stated in the contract. There is also an increasing level of concerns and demands for compliance around data privacy and security from donors as well as the country governments where IFPRI operates. Failure to comply with requirements may mean not only losing the funds that IFPRI receives but also damaging the Institute’s reputation in the donor community.

The Institute works with both quantitative and qualitative data from primary and secondary sources. IFPRI researchers collect primary data from household, business, and institutional surveys. In addition, the researchers also combine, aggregate, simulate, and analyze existing data from secondary sources in various ways, and create secondary datasets as a product. Examples include many social accounting matrices (SAM), Statistics on Public Expenditures for Economic Development, and simulated projections under different scenarios generated from the IMPACT Model among others. These datasets vary from small surveys (less than 100 observations) to large data portals. Management of such a wide range of datasets present numerous challenges particularly during data collection, distribution and sharing, and dissemination. Among the challenges are issues related to data security and privacy of respondents, and possibly reduced interoperability and non-compliance with data quality standards.

To address all of these concerns and be more forward-looking, IFPRI needs to revisit both IFPRI’s Data Policy and the institutional procedures designed to manage data generation, curation, use, and exchange. To this end, IFPRI recently established the Data Governance Team (DGT) with the following objectives:

1. Take a leadership role in the creation, implementation, and oversight of an IFPRI-wide data management goals, standards, practices, and processes;
2. Foster an organized system to manage data effectively and smoothly;
3. Reduce costs and increase effectiveness through coordination of efforts across various IFPRI divisions and units; and
4. Ensure the transparency of the processes and adherence to open access policies, privacy acts, and the other prevailing policies as imposed by country governments in which IFPRI operates.

The DGT constitutes staff members from each of the Institute’s four research divisions, the two CGIAR research programs hosted by the Institute, the Information Technology (IT) Services unit, the Finance and Administration (F&A) division, the Communications and Public Affairs (CPA) division, and the Institutional Review Board (IRB).

**Terms of Reference**

This Terms of Reference describes a consultancy designed to assist the DGT by addressing the following:

1. Identification and description of the types of data resources and products that are generated, curated, used, or exchanged by IFPRI research and the related challenges in governing and managing these products.
2. Recommendation of appropriate policies for governing and managing these products, taking into consideration:
	1. the needs of IFPRI, including its overall mission and mandate as well as the needs of its research, communications, and corporate services staff;
	2. the needs of IFPRI’s collaborators in research and communications activities;
	3. benchmarks, standards, and best practices used in comparable public policy research organizations;
	4. the CGIAR Open Access and Open Data Policy;
	5. the protection of private information and ethical research on human subjects; and
	6. the policies and provisions set forth by IFPRI’s donors.
3. Recommendation of appropriate procedures and practices for managing these data products, taking into consideration:
4. rules or other mechanisms for defining ownership and use-rights for data;
5. standardization of metadata and other supporting documentation as required by donors and data users;
6. protection of privacy, private information, and other proprietary data;
7. processes and procedures for exit clearance for data when staff leave IFPRI;
8. include data management plan in research proposal that involves generation of primary data

**Expected Deliverables/Outputs**

The Consultant is expected to produce a Data Governance Report that provides

1. Detailed identification and description of IFPRI’s full range of data resources and products, as well as the challenges associated with each;
2. Recommendations for a revised IFPRI Data Policy that addresses the points described above;
3. Recommendations for procedures and practices to accompany the IFPRI Data Policy that addresses the points described above.

**Methodology**

The Consultant will follow the workplan and the time schedule agreed with Indira Yerramareddy, manager of Knowledge Management and Web and will work closely with her and other individuals as determined necessary. The appropriate methodology for soliciting opinions and thoughts from IFPRI staff and other concerned stakeholders regarding processes, protocols, and policies (if the needs arise) should be determined in consultation with DGT members. The consultant should consider/undertake all the required data/information from research divisions and regional offices, donors, and other relevant sources to suggest the process improvement and update policies that can sustain the changing landscapes of research and open data.

**Duration of the work**

This contract shall cover 20 working days spanning 3 months, beginning April 1, 2018 and ending on June 30, 2018.

**Required expertise and qualifications**

1. Advanced degree (Masters level or above) in a relevant field such as information, data, and knowledge management; library science; informatics or data science or equivalent experience.
2. At least 7-10 years of prior work experience in the governance and management of data, particularly research data or scientific data and full awareness of the issues in this field.
3. Demonstrated knowledge of, and experience with, policies, legislation, and procedures related to open data, open data platforms and repositories, privacy and confidentiality protections in data, and other rules and regulations related to data management.
4. Familiarity with household survey data and geospatial data, as well as machine and human readable metadata and data files.

**Contact:** Indira Yerramareddy (email: i.yerramareddy@cgiar.org)

Enclosed: 1. IFPRI data policy

2. CGIAR open access and open data policy

**International Food Policy Research Institute (IFPRI) Policy on Dataset Management**

This memorandum represents IFPRI policy on household data access. It replaces the policy outlined in a memorandum dated July 6, 2000.

**Principles**

IFPRI’s mandate is to conduct policy research and outreach that have positive impacts on the lives of poor and malnourished people. Increasing access to the data assembled by IFPRI staff and collaborators, while safeguarding the privacy of participants and protecting confidential and proprietary information, will enhance IFPRI’s ability to carry out this mandate. The purpose of this policy is to assist IFPRI research divisions and researchers in managing the storage of, retention of, and access to data and records associated with research activities.

Data assembled by IFPRI staff and collaborators represent a vital element of the Institute's work. IFPRI as an institution has sole proprietary claim or right to datasets assembled by IFPRI staff and collaborators, subject to any contractual agreement between IFPRI and a donor agency, a collaborating institution, or a government entity of the country in which the data were collected. Any rights and obligations with respect to data and their release established by such individual contractual agreements will supersede this policy.

**Definitions**

*Research data*: facts and observations on which a test is based. The data may be numeric or descriptive.

*Primary dataset*: raw data and other research data that are obtained directly from respondents. The data are generally captured through surveys, interviews, focus groups, or other direct interactions with individuals in the field. Primary datasets are collected by researchers to accomplish a project’s objectives. Primary data should be stored internally at IFPRI and be accessible to IFPRI researchers.

*Secondary dataset*: pre-existing data not gathered or collected for the current research project. Usually it has been collected by another organization or source or data collected from government publications.

*Major publication*: research reports, books, or journal articles. Discussion Papers are *not* considered major publications.

*Researcher*: the individual who conducts, directs, and/or facilitates research activities and carries the primary responsibility for the research. The researcher is referred to as the *principal investigator* (PI) when acting as the leader of a research team. When IFPRI commissions and oversees research performed by collaborators, the Institute’s research manager for the project has the responsibilities listed below for researchers.

**Responsibilities**

Researchers are responsible for ensuring the following:

* Research methods, data, and records are reliable, accurate, and complete.
* Data are appropriately recorded and referenced following the documented procedures provided by IFPRI’s Communication Division and available on the Institute’s intranet.
* Where confidentiality provisions apply, data comply with privacy protocols for personal privacy protection. It is desirable that the data are kept in a way that can be referenced by third parties without breaching confidentiality.

All data are subject to IFPRI’s Institutional Review Board (IRB) guidelines.

* The division director is informed if confidentiality agreements apply and what obligations they entail. This should be reported in the documentation provided to the Communication Division.
* When they apply, permission requests are submitted and approval received in writing for data obtained from outside databases, a collaborating agency, or a government. The permission release should be included in the documentation provided to the Communication Division.
* Consent forms and permissions are obtained when a project involves human participants.

**Practices**

IFPRI will make all primary and value-added secondary datasets collected after January 1, 1999, publicly available two (2) years after all data collection ceases or, before two years, at the time of a major publication (see definition above) by the lead data collector. Before the end of two years or the time of such publication, datasets should be available for IFPRI’s Publications Review Committee (PRC) or any other reviewer in case of publishing with external publishers.

Unless otherwise stated in a contract or agreement, IFPRI does not need permission to use data gathered by collaborators whom IFPRI has paid for collection or surveys. The collaborators also have the rights to use the data. They need IFPRI’s permission, however, to release the data jointly collected with IFPRI.

Prior to the end of the two-year release time, access to datasets—for IFPRI and non-IFPRI researchers alike—will be granted if the IFPRI project leader responsible for assembling the dataset and the project leader’s division director agree. If there is a disagreement between these two parties on data access prior to the end of two years, the perspectives of both will be recorded in writing, and the final decision relating to access will be made by IFPRI’s director general. If the project leader has left IFPRI prior to the end of two years, the decision to grant access to the dataset during that period rests with the division director alone.

In the event of an IFPRI researcher leaving the Institute, he or she may negotiate with the division director to take analysis files/ other files that are not yet public, for their own use, but original data and records are to remain in the division

Some datasets collected prior to the implementation of IFPRI’s data documentation policy on January 1, 1999, may not be sufficiently well documented for public release. If these datasets are requested, release procedures will be identical to those for datasets requested prior to the end of their two-year period.

IFPRI’s Communications Division will be responsible for facilitating the implementation of this policy, including listing datasets on IFPRI’s public Web site, filling requests for datasets, and monitoring which datasets are requested. Permission request letters for access to the data will be provided by the Communication Division and be available from IFPRI’s intranet.

An alert from the Project Database will be sent to the IFPRI’s Communication Division when a project includes a dataset as a deliverable. It is the responsibility of the research divisions to ensure that all datasets are delivered and properly documented, following the guidelines provided by the Communications Division, at the end of the two years, before that time if the project leader and division director agree, or when a major publication using the data is about to be published. A new item will be added in the exit form for research staff that will refer to datasets that are pending and that need to be deposited in safe storage in the appropriate research division.

For research publications that are reviewed and approved by IFPRI's PRC, to the extent the manuscript relies on IFPRI primary or value-added secondary data, the PRC will request the author(s) to provide the data electronically along with the final accepted draft of the publication. A copy of this request will also be sent to the Communications Division.

IFPRI’s Communication Division will provide a standard citation to be employed by all external users of publicly available IFPRI datasets and require external users to use the citation in a document’s text and in the list of references and/or bibliography.

Division directors, with the help of IFPRI Computer Services, will provide storage space that meets security and confidentiality requirements for research data and records while in preparation.

IFPRI staff members and consultants will be provided a copy of this policy at the time their relationship with IFPRI is formally established, and acceptance is a condition of that relationship. The policy may be changed by the Institute through revision of this document.



CGIAR Open Access and Data Management Policy (the “Policy”) 2 October 2013

# CGIAR Open Access and Data Management

Policy

## 2 October 20131

1 This Open Access and Data Management Policy is effective as of 2 October, 2013. It was endorsed by the System Management Board on 13 July 2016 ([SMB/M1/DP7](http://cgiarweb.s3.amazonaws.com/wp-content/uploads/2016/07/SMB1-20_MeetingSummary_1stMeeting_July2016.pdf)). It was first approved by the Consortium Board on 2 October, 2013 and endorsed by all Centers on 20 November 2013, prior to the transition to the revised CGIAR System that took effect on 1 July 2016. While nomenclature has been amended to reflect these governance changes, the historical context of the document has not been updated.

## Preamble

CGIAR regards the results of its research and development activities as international public goods and is committed to their widespread dissemination and use to achieve the maximum impact to advantage the poor, especially smallholder farmers in developing countries. CGIAR considers Open Access (defined below) to be an important practical application of this commitment as it enhances the visibility, accessibility and impact of its research and development activities. Open Access improves the speed, efficiency and efficacy of research; it enables interdisciplinary research; assists novel computation of the research literature; and allows the global public to benefit from CGIAR research. Furthermore, CGIAR recognizes the benefits that accrue to individual researchers and to the research enterprise from wide dissemination, including greater recognition, more thorough review, consideration and critique, and a general increase in scientific, scholarly and critical knowledge. CGIAR further recognizes that, in implementing this Policy, it can more easily and collectively build the infrastructure necessary to be at the forefront of the open access and open data for agriculture movement.

This Policy stems from – and complies with – the *CGIAR Principles on the Management of Intellectual Assets* (“**CGIAR IA Principles**”)1, which is the umbrella document for this Policy. In particular, this Policy expands on Article 6.1 of the CGIAR IA Principles which provides that “*The [CGIAR] System Organization and the Centers shall promptly and broadly disseminate their research results, subject to confidentiality as may be associated with [certain] permitted restrictions, or subject to limited delays to seek IP Rights [(patents, etc.)]*”.

## Scope and implementation

This Policy was approved by the CGIAR System Organization2 on October 2, 2013 and is effective as of this date (the “**effective date**”). Implementation of and compliance with this Policy by CGIAR System Organization, CGIAR Centers and their implementing partners within the scope of the Strategy and Results Framework (“**SRF**”) and the CGIAR Research Programs (“**CRPs**”) will be phased over a transition period. The transition period runs from the effective date of the Policy for an initial period of 5 years, with comprehensive implementation by the end of 2018. This Policy should be read in conjunction with the CGIAR Open Access and Data Management Implementation Guidelines3, which may be updated from time to time to reflect current recommended practices.

1 The CGIAR IA Principles are available at <https://library.cgiar.org/bitstream/handle/10947/4486/CGIAR%20IA%20Principles.pdf?sequence=1>

2 This Policy was approved by the CGIAR System Organization when it was called the Consortium of International Agricultural Research Centers and operated under the name “CGIAR Consortium”

3 Available at [http://library.cgiar.org/bitstream/handle/10947/4489/Open%20Access%20Data%20Management%20Implementa](http://library.cgiar.org/bitstream/handle/10947/4489/Open%20Access%20Data%20Management%20Implementation%20Guidelines.pdf) [tion%20Guidelines.pdf](http://library.cgiar.org/bitstream/handle/10947/4489/Open%20Access%20Data%20Management%20Implementation%20Guidelines.pdf)

## Information products

This Policy sets common expectations with respect to Open Access to the following indicative types of information products (“**information products**”): peer-reviewed journal articles; reports and other papers; books and book chapters; data and databases; data collection and analysis tools (e.g. models and survey tools); video, audio and images; computer software; web services (e.g. data portals, modeling on-line platforms); and metadata associated with the information products above.

## Policy statement

### General

* + 1. Openness. Best efforts shall be used to make all information products Open Access, subject always to the legal rights and legitimate interests of stakeholders and third parties, including intellectual property rights, confidentiality, sensitivity (including price and politically sensitive information), farmers’ rights and privacy.

Information products may not always be of value to others, for example because those outputs are draft, poor quality or incomplete. Open Access arrangements should consider the characteristics of the information product, their potential impact, the level of data processing required, and whether the information products generated are within the scope of this Policy. Some judgment therefore needs to be made over the information products that will be made Open Access.

* + 1. Suitable Repositories. Stable, permanent, Open Access repositories shall be utilized, to enable users and other sites and search engines to access or locate information products, including application programming interfaces (APIs) or other mechanisms enabling those information products to be available from the CGIAR website and associated web-based products. Preference should be given to existing repositories to minimize the number of repositories in use (and the interoperability challenges presented by multiple incidences of repositories).
		2. Interoperability. Syntactic and semantic interoperability is a key consideration in enabling and promoting international and interdisciplinary access to and use of information products. Information products must therefore be described with standardized metadata, and stored and delivered using appropriate protocols and formats to ensure that their content can be discovered, shared and incorporated across different technological platforms.
		3. Data storage and preservation for future use. Information products must be stored where users can find them and where they will be preserved for future use. As time goes by, they will need to be managed, maintained and curated.
		4. Copyright and Open Licenses. Suitable open licenses shall be used that recognize the legal rights to information products and encourage their use and adaptation.
		5. Incentives and professional expertise. Incentives and the development of professional expertise in all areas of Open Access and data management shall be devised, adopted and promoted.
		6. Translation. Translations of key documents and other media into pertinent languages are encouraged. All versions should be deposited in suitable repositories and made Open Access.
		7. Limited internet connectivity. To assist those with limited internet connectivity, designing easily accessible information products (e.g. websites, PDFs) or making available alternative versions that require minimal data download to see and use is encouraged.
		8. Open Access and Data Management Plans. Open Access and Data Management Plans should be prepared in order to ensure implementation of this Policy. Such Plans shall, in particular, outline a strategy for maximizing opportunities to make information products Open Access.

### Open Access for indicative types of information products

* + 1. Peer-reviewed journal articles. Peer-reviewed versions of scholarly articles reporting research should be deposited in a suitable repository and made Open Access 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 to their needs. 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).
		2. 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 Access as soon as possible and in any event within 3 months of their completion.
		3. Books and book chapters. The full digital version of books and book chapters shall be made Open Access as soon as possible after publication and in any event within 6 months either through self-archiving or other suitable publication arrangements.
		4. Data and databases. Data (and any relevant data collection and analysis tools) shall, subject to any additional donor requirements, be deposited in a suitable repository

and made Open Access 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 Access.

* + 1. Video, audio and images. Complete final digital versions of video and audio outputs, and image collections must be stored appropriately and made Open Access within 3 months of their completion.
		2. Computer software. Where an information product is software developed internally, the associated source code must be deposited in a free/open software archive upon completion of the software development. Access to such information products may be granted subject to appropriate licences (e.g. Copyleft).
		3. Metadata. The metadata of an information product must be deposited in a suitable repository before or on publication of the information product. Where an information product is not deposited in a suitable repository, the deposited metadata must include a link to the information product.

## Review

The System Management Office will carry out an evidence-based review of the implementation of this Policy on an annual basis. The reviews will be used to devise appropriate institutional tools and guidelines for the implementation of this Policy.

The System Management Office (in consultation with the Centers) will review this Policy in 2015 and every two years thereafter in light of experiences gained. This Policy may be amended at any time by agreement of the System Organization, in consultation with the Centers.

## Definitions

For the purposes of this Policy:

**Data** means the digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings, including data sets used to support publications and/or that have been prepared and validated but that do not support publications. This does not include laboratory notebooks, preliminary analyses, drafts of

scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as laboratory specimens.4

**Database** means a collection of independent works, data or other materials, which are arranged in a systematic or methodical way and which are individually accessible by electronic or other means.5

**Open Access** means the immediate, irrevocable, unrestricted and free online access by any user worldwide to information products, and unrestricted re-use of content (which could be restricted to non-commercial use and/or granted subject to appropriate licences in line with the CGIAR IA Principles), subject to proper attribution.

4 Adapted from the Office of Science and Technology Policy Guidelines.

5 From Directive 96/9/EC of the European Parliament of 11 March 1996 on the legal protection of databases.