



Review of CGIAR's Open Access/Open Data Policy and implementation support

Commissioned by
Independent Evaluation Arrangement

March 2018

Review Team:
Johannes Keizer
Devika Madalli



Independent
Evaluation
Arrangement

This review has been commissioned by the Independent Evaluation Arrangement (IEA) of CGIAR.

The Independent Evaluation Arrangement (IEA) of CGIAR encourages fair use of this material provided proper citation is made.

Correct citation: CGIAR - IEA (2018), Review of CGIAR's Open Access/Open Data Policy and implementation support. Rome, Italy: Independent Evaluation Arrangement (IEA) of CGIAR
<http://iea.cgiar.org/>

Table of Contents

Abbreviations	ii
Executive Summary	1
Recommendations	3
1. Introduction	5
1.1 Background	5
1.2 International context	6
1.3 Objectives and focus	7
1.4 Methodology	8
2. Appropriateness of the Policy and Implementation Guidelines	9
2.1. Introduction	9
2.2 Origins of the OA/DM Policy and Guidelines	9
2.3 Scope and adequacy	10
2.4 Conclusions	12
3. Effectiveness and Efficiency	14
3.1 Effectiveness	14
3.1.1 Introduction	14
3.1.2 Awareness	14
3.1.3 OA/DM implementation plans	15
3.1.4 Data management challenges	16
3.1.5 Status of implementation	19
3.1.6 Promotion and adoption of best practices	21
3.2 Efficiency	23
3.2.1 Introduction	23
3.2.2 Communication	23
3.2.3 Harmonization of donor requirements	24
3.2.4 Infrastructure	25
3.3 Conclusions	27
4 Implementation Support	29
4.1 Introduction	29
4.2 Support Pack	29
4.3 Communities of practice	30
4.4 Resources	32
4.5 Conclusions	33
5 Governance and management	34
5.1 Introduction	34
5.2 System Council role	34
5.3 Roles and Responsibilities	34
5.4 Center-level governance and management	35
5.5 Reporting	35
5.6 Conclusions	37
Annex: Team Bios	38

Abbreviations

CCAFS	CRP for Climate Change, Agriculture, and Food Security
COP	Community of Practice
CRP	CGIAR Research Program
DFID	Department for International Development of UK
DM	Data Management
DMTF	Data Management Task Force
FAIR	Findability, Accessibility, Interoperability, and Reusability
FC	Fund Council
GFAR	Global Forum on Agricultural Research
GODAN	Global Open Data for Agriculture and Nutrition
IA	Intellectual Assets
IEA	Independent Evaluation Arrangement
IP	Intellectual Property
ISPC	Independent Science and Partner Council
MARLO	Managing Agricultural Research for Learning and Outcomes
OA	Open Access
OAWG	Open Access Working Group
OD	Open Data
OECD	Organization for Economic Co-operation and Development
SMB	System Management Board
SMO	System Management Office
USAID	United States Agency for International Development

Executive Summary

The CGIAR Open Access and Data Management (OA/DM) Policy is one of the System-wide policies applicable to the 15 CGIAR Research Centers that operate as independent research organizations in implementing CGIAR research activities with their partners. The OA/DM Policy came to effect in 2013. It is based on CGIAR's Intellectual Asset Principles, approved in 2012, expanding the sections that concern international public goods and broad dissemination of CGIAR's research results. The Independent Evaluation Arrangement (IEA) undertook the review of the OA/DM Policy following the review of Intellectual Asset Principles as these policies are linked.

CGIAR is a major producer of information products that are international public goods. In addition to a wide range of publications and data that account for the largest part of these information products and are the focus of this review, the OA/OD Policy covers many types of information technology products and tools.

Moving towards open access is essential for maximizing CGIAR's impact through expanding the range of user and beneficiaries of research outputs, enhancing the effectiveness and efficacy of research, and increasing the visibility and recognition of the scientists' work and CGIAR as a System. The OA/DM Policy foresees CGIAR at the forefront of open access and open data for agricultural research for development.

The OA/DM Policy is accompanied by Implementation Guidelines. Implementation of the Policy is proceeding through a five-year transition period and comprehensive implementation is expected at the end of 2018. The review focused on assessing the Policy against CGIAR needs and the international context, as well as implementation guidance, implementation support, progress and challenges as CGIAR is in the final year of the transition period. As full implementation of the Policy is not scheduled until 2018, the review did not assess implementation at Center level nor progress at Centers, but focused instead on System-level support efforts towards compliance.

The review team found the Policy to be concise and appropriate in terms of scope and detail, which is also the view of stakeholders interviewed and surveyed. The Policy is in line with "FAIR" principles (referring to findability, accessibility, interoperability and reusability) ¹, commonly accepted as a standard, particularly for open access of data. The team found that these principles are thoroughly incorporated in the design, implementation and reporting of OD efforts in CGIAR although "FAIR" as a concept was formally introduced after the Policy was approved. The Implementation Guidelines were initially prepared with focus on implementation planning and support. They have served that purpose well, but the team found several areas where the Guidelines need revision and expansion.

At this early stage of moving towards comprehensive implementation of the OA/DM policy, the review team found that the Policy has already resulted in a cultural shift in Centers and CRPs. It has resulted in community building towards achieving OA resources and data, and greater awareness of benefits of and requirements for open access. This is not only true for the data and information managers, but also researchers across the System recognize the importance of open access to their work and mission. It was not scope of this review to analyze the effects and impacts open access is having; this is too

¹ <https://www.nature.com/articles/sdata201618>

early given that the Policy is not yet fully implemented. However, feed-back from stakeholder groups attests that progress is being made and needs of infrastructure are being addressed.

The review team identified several areas where it is important to intensify the efforts to be able to reach comprehensive implementation of Policy by end of 2018. These relate particularly to financial investment and resources needed for implementation of open access across CGIAR, and data management aspects of open access, which have both technical and resource implications.

Open access to scholarly publication is an established and agreed practice and CGIAR has since long been moving towards providing open access to its research publications. Unlike open access publications, open data is a relatively new trend and requirement. Among the data communities globally, there are open questions regarding open access to datasets, which relate to data quality, data prioritization, common ownership of data, data storage procedures and other data management practices. Data management necessitates an investment of time, and elements of data management are necessary all along the data collection and publishing pathway. Overall guidance can be developed, but change and improvements need to become part of the day-to-day work of staff in order to ensure high quality, relevant data being made open access.

Given that the volume of publications and data in particular is growing, it is very important to take resource needs into account. Regarding peer reviewed publications, publishing has costs irrespective of the venue of publishing. For data, resources are needed in the processes related to open data, for assuring quality and relevance of the final data outputs, and therefore prioritization of what will be open access is important. As Centers bear the costs of open access often well-beyond project lifetime, resourcing of open access and open data is a challenge which should be addressed at both Center and System levels.

For providing central support for implementing the Policy a Phase I project was funded. The review team found that the activities of the Phase I project significantly supported early implementation of the Policy, although there are no comprehensive reports available. The Phase II project which aimed to strengthen the volume, variety and quality of openly accessible research outputs was later subsumed as Module I into the Big Data Platform, which has operated since the beginning of 2017.

The two main areas where Centers and programs have received central support for implementing the OA/DM Policy are communities of practice and a repository of support materials in Open Access Support Pack. Two communities of practice were set up, namely the Open Access Working Group and the Data Managers Task Force. These have been crucial for sharing information, exchanging experiences and consolidate views of the professionals involved in developing best practices and solutions to challenges related to technical and managerial aspects of open access implementation. The review team found that the approach of tasking smaller working groups within the communities to focus on specific issues was very appropriate and potentially effective, as long as results are communicated to the wider community, establishing linkages also to the network of legal experts who deal with intellectual assets. The COPs have enhanced capacity through workshops, and capacity development remains an important need also in the future. The review team found, overall, that's CGIAR has well-qualified and engaged champions and leaders in Centers and CRPs, which has strengthened the awareness of open access across CGIAR. .

The Open Access Support Pack was intended as a resource to provide support to the implementation process. The review team observed that the Support Pack has not been sufficient for providing comprehensive, updated guidance material. The team considers that alongside with updating the Implementation Guidelines, the Support Pack needs to be restructured and updated as a dynamic resource that is complementary to the formal Guidelines.

The Implementation Guidelines state that the Centers have primary responsibility and accountability for day-to-day implementation of open access, and several Centers have their own policies for publications as well as for data; most Centers have developed a implementation plan following the Guidelines. By the end of 2018, central systems need to be in place for governance and management for enforcing full implementation and monitoring progress. This is not yet the case, although the team found work on indicators for assessing the extent of “FAIRness” of data laudable. There needs to be common understanding about what full implementation means, particularly concerning data. Since 2014, there has been massive growth in the number of published datasets in most Centers, with high variance in these datasets being open access. Therefore, monitoring progress on open access of datasets is important.

The review team concludes that open access needs to be mainstreamed in all management, with further support and promotion from the System Organization. This includes championing open access requirements and benefits by senior leadership; promotion and incentives among all researchers; communication across communities of practice, including data managers, knowledge managers and legal experts; and planning and budgeting for open access in project development, as well as in Center business plans as open data has costs beyond the project life-cycle. A stronger push is needed from Center leadership to make sure that the importance of the Policy is clearly communicated within Centers and that open access issues and challenges are kept on the agenda of Center senior management. Governance of the Policy and its implementation need strong involvement of the System Council and continuous central support from the System Organization.

Recommendations

Recommendation 1: Taking into account lessons from the implementation to date, the Implementation Guidelines should be expanded to reflect the current phase of implementation and future needs at Centers and programs. The additional content should include the following:

- clarity in System-level governance and management roles for both OA and OD oversight and reporting reflecting CGIAR’s governance transition in 2016 and establishment of the Big Data Platform;
- aspects of data management that are particularly pertinent to open data, such as data prioritization, curation and quality;
- templates for legal compliance and process for legal clearance (complementing guidelines for intellectual assets policy);
- guidance for workflows for tracking progress towards OA/OD, and sharing lessons;
- guidance for compliance to ethics in data publishing;
- minimum requirements of repositories for open access;
- guidance on promotion and incentives.

Recommendation 2: The System Management Office should promote greater interaction among data managers and intellectual property and legal experts through joint activities and capacity development events of the respective communities of practice so as to enhance best practices and workflows that address all aspects of OD management.

Recommendation 3: The System Organization should engage with donors to harmonize on requirements OA and OD and agree on common principles.

Recommendation 4: The System Management Office should oversee restructuring of the OA/OD Support Pack to enhance its usefulness as the main source of guidance to different kinds of users within CGIAR, and to make it more suitable as a communication and dissemination channel and tool.

Recommendation 5: For resourcing OA implementation, action is needed at Center and System levels, and should include the following:

- Centers should make budget planning for OA as part of normal project formulation for all research and consider earmarking institutional financial resources for peer-reviewed publication which generate costs long after project termination.
- SMB should consider approaches to OA financing, considering that Centers have limited resources beyond project lifetime for guaranteeing OA.
- Central funding, necessary for supporting Policy implementation and facilitating the communities of practice and capacity development, should to be planned in an inclusive manner and transparently reported.

Recommendation 6: The System Organization should assume an active role in championing OA across CGIAR through the following ways:

- taking on a coordinating role for promotion and enhancement of implementation the OA/DM Policy across CGIAR;
- inclusion of OA issues regularly in SMB meetings;
- developing reward mechanisms for exemplary OA/OD management;
- supporting System-wide sharing of OA/OD best practices and promotion;
- improving online visibility and information sharing of CGIAR efforts on OA.

1. Introduction

1.1 Background

CGIAR is a global partnership comprised of 15 Research Centers that implement development-oriented agricultural research through CGIAR Research Programs (CRPs) producing public goods results. The CGIAR Open Access and Data Management (OA/DM) Policy² is one of the System-wide policies applicable to the 15 CGIAR Research Centers that operate as independent research organizations in implementing CGIAR research activities with their partners.

Approved in October 2013, the OA/DM Policy is based on CGIAR's Principles for Management of Intellectual Assets (IA) Principles;³ the System-wide policy from 2012 that provides CGIAR a common position and framework for governing the production, acquisition, management and dissemination of IA and specifically concerns intellectual property (IP) rights. The OA/DM Policy expands on the Principle 1 that states CGIAR's commitment to international public goods, and Article 6.1 that emphasizes prompt and broad dissemination of CGIAR's research results. The OA/DM Policy aims to provide a common position and expectations for open access (OA) of a wide range of information products across CGIAR, publications and research data being the major types. The Policy is accompanied by Implementation Guidelines⁴ that were approved in July 2014. Transitional period for implementation of the Policy is five years with full implementation expected by the end of 2018.

The Policy presents a vision of benefits from OA, where it *"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"*. CGIAR considers OA to be an important practical application of CGIAR's commitment to wide-spread dissemination of its international public goods results, as stated in the Policy. OA enhances the visibility, accessibility and impact of CGIAR's research and development activities. In the Policy, benefits to researchers are considered to include *"greater recognition, more thorough review, consideration and critique, and a general increase in scientific, scholarly and critical knowledge"*.

Following the endorsement of the Policy, the Fund Council (FC) approved a proposal⁵ for funding submitted by the Consortium Office to assist Centers in their implementation efforts. Initially, the project was approved as a Phase I to focus on setting up the systems and foundation for implementing OA in CGIAR. The objectives of the project included setting up templates, framework and guidance to support implementation efforts. The new Big Data Platform has assumed a subsequent phase of central implementation support in its Module 1.

OA isn't a new concept to CGIAR but it has been discussed within CGIAR for more than a decade. OA practices have been introduced in many Centers over several years; Centers have invested in

²<https://cgspace.cgiar.org/bitstream/handle/10947/4488/Open%20Access%20Data%20Management%20Policy.pdf?sequence=1>

³ <https://cgspace.cgiar.org/bitstream/handle/10947/4486/CGIAR%20IA%20Principles.pdf?sequence=1>

⁴<https://cgspace.cgiar.org/bitstream/handle/10947/4489/Open%20Access%20Data%20Management%20Implementation%20Guidelines.pdf>

⁵ CGIAR Open Access and Open Data Phase I, Nov 2014:

<https://cgspace.cgiar.org/bitstream/handle/10947/3737/CGIAR%20Open%20Access%20and%20Open%20Data%20Phase%20I.pdf?sequence=4>

publishing OA and open repositories have been in place. Eight Centers had a policy for OA concerning publications prior endorsement of the Policy. There have been many System-level initiatives to support and encourage OA, such as the System-wide Information, Communication, Technology and Knowledge Management (ICT-KM) program. However, the adoption and support of a System-wide Policy and implementation support for these efforts have the potential to create even greater opportunities for CGIAR research and its use for on-the-ground impact. This review therefore covers both the Policy and its Implementation Guidelines, as well as the implementation support provided during the transition period to date.

1.2 International context

Whilst openness has been one of the accepted norms of scientific practice, a paradigm of 'Open Science' is emerging⁶, which encompasses a more collaborative scientific enterprise including OA to scientific data and scientific journals. In parallel, the availability and scale of data that are available for and produced by science has massively increased as has the potential for analysis and wide use of those data. 'Big data' and data driven research are now ubiquitous across all scientific disciplines and opening up new possibilities for addressing previously intractable scientific challenges. Open Science creates a potential environment for seamless knowledge sharing and acceleration of research. Shared and openly available information products, especially publications and research data, are some of the essential factors towards Open Science.

These two elements: OA to journals and scientific publications, and OA to data have their own opportunities and challenges, which for CGIAR are discussed in this report. In CGIAR, OA approach to scientific publications has become commonplace and widely understood. On the contrary, open data (OD) in the context of Open Science is still being defined. In 2014, a group of academic and private stakeholders interested in knowledge sharing formulated, a set of guiding principles; the "FAIR principles"⁷. FAIR stands for: Findable, Accessible, Interoperable and Reusable and the principles define criteria for "FAIRness" with focus on automated processes. Figure 1 presents the four FAIR principles in detail, as they apply to data. These principles aim to identify and break-down the elements of what is needed for data to be considered "open", with a focus on accessibility to research data for maximizing the use and re-use of data. While the OA/DM Policy precedes the launch of the FAIR principles, it provides a similar emphasis for data usability as in the principles.

As stated by OECD, the global landscape for data sharing in science is increasingly complex.⁸ Many international data networks for OD already exist with different aims and structures responding to specific needs of different research communities. Some are linked to large intergovernmental research infrastructures, whereas others are distributed with more flexible governance structures, and provide access to data from many different domains (for example the Australian National Data Service).

⁶ OECD (2016), "Open science", in OECD Science, Technology and Innovation Outlook 2016, OECD Publishing, Paris. http://dx.doi.org/10.1787/sti_in_outlook-2016-37-en

⁷ Fair principles at <https://www.force11.org/group/fairgroup/fairprinciples>

⁸ OECD (2017), "Co-ordination and support of international research data networks", OECD Science, Technology and Industry Policy Papers, No. 51, OECD Publishing, Paris.

Figure 1: The Fair Principles to OD



According to OECD⁹, challenges associated with data sharing in science include the following:

- As massive amounts of data are being generated in an unprecedented scale, their reliability and statistical validity is not yet fully understood.
- New forms of personal data, including social networking data, are constantly being created and their use may create risks to the individuals' privacy.
- Researchers often lack the resources and skills needed to make sure that the data they gather and produce are available for re-use.
- Researchers need to have the right set of incentives to ensure effective data sharing.
- The issue of inadequate quality of open datasets is a challenge to their re-use¹⁰.

The value from OD is often constrained due to missing or incomplete metadata, lack of descriptions and other issues related to the comprehensibility and origin of data.

1.3 Objectives and focus

The purpose of this review was to assess the appropriateness of the OA/DM Policy in terms of achieving its intended purpose of maximizing global accessibility of CGIAR research. Given that the Policy is still being rolled out and full implementation is not expected until end of 2018, the focus was to assess the intent, relevance, clarity and coverage of the Policy and its Implementation Guidelines and the effectiveness of support provided during the transitional period. The review provides lessons from implementation of the Policy since its approval and includes recommendations for enhancing the efficiency and effectiveness of the Policy and its implementation.

The review examined the Policy in a comprehensive manner with focus on the following aspects:

- **Appropriateness:** The scope, adequacy and clarity of the Policy, and the extent to which it reflects latest developments in OD and knowledge management. Also, the Implementation Guidelines accompanying the Policy were considered.
- **Effectiveness and efficiency:** Progress in implementation of the Policy, the extent to which good practices have been promoted, relevance and adequacy of support for achieving implementation in the agreed timeline (2018)
- **Implementation support:** Adequacy and efficiency of guidance, documentation and

⁹ OECD (2015), "Promoting data-driven scientific research", in Data-Driven Innovation: Big Data for Growth and Well-Being, OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264229358-11-en>.

¹⁰ Open data: Quality over quantity, International Journal of Information Management, Volume 37, Issue 3, Pages 150-154, Shazia Sadiq, Marta Indulska.

technical support extended to the Centers and stakeholders

- **Governance and Management.** Given that the full compliance is not expected until in 2018, the assessment of governance and management at Centers and at the System level was limited to what is currently in place and foreseen.

1.4 Methodology

The review team followed mainly a two-pronged approach for eliciting evidence for its analysis including document review and interviews. The review of CGIAR documents includes the following:

- OA/DM Implementation Guidelines.
- Implementation support proposal and progress report.
- OA Support Pack.
- Minutes and notes from Data Managers Task Force (DMTF) and the Open Access Working Group (OAWG) meetings.
- Fund/System Council and Consortium Board meeting minutes related to OA/DM Policy, its support and implementation.
- Background literature and resources including “Open Access and Open Data: challenges and solutions”
- CG Core Metadata Schema and Application Profile.
- OA and DM Plan template and sample of Center plans.
- A 2015 survey report on the OA/DM practices in each CGIAR Center based on the Inventory of CGIAR Consortium Members’ Open Access & Data Management Practices.
- External resources such as the report of the GODAN Donor Open Data policy and ‘Open access to scientific data and literature, the assessment of research by metrics’ report by the International Council for Science¹¹ and reports by OECD¹².

Feedback was solicited through interviews and surveys covering the following stakeholder groups: Center focal points, researchers, data managers, donors, System Management Office (SMO) and external experts. In addition, team members attended the DMTF 2017 annual meeting and presented the review virtually to the OAWG at its meeting. As the CGIAR IA Principles review was taking place in parallel, the two teams had interaction.

¹¹ <https://www.icsu.org/publications/open-access-to-scientific-data-and-literature-and-the-assessment-of-%20research-by-metrics>

¹² [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/EAS/STP/NESTI\(2015\)4/FINAL&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/EAS/STP/NESTI(2015)4/FINAL&docLanguage=En)

2. Appropriateness of the Policy and Implementation Guidelines

2.1. Introduction

The team assessed the OA/DM Policy for clarity, scope, and adequacy, and the extent to which latest developments and challenges in OD and knowledge management are reflected. The team also assessed the Implementation Guidelines that were finalized a year after endorsement of the Policy but are presented in conjunction with the Policy.

2.2 Origins of the OA/DM Policy and Guidelines

Following the approval of the IA Principles of CGIAR, FC members highlighted the need for expanding on the policy governing global accessibility of research results for impact. The FC noted that principles for OA would be “timely” especially as many Centers had already introduced aspects of OA in their work. The initial FC discussion focused on scope of such policy, linking it with other OA initiatives, and taking a staggered approach to OD to ensure that cost increases would be incremental. The Policy was developed and approved in 2013, and its implementation was phased over a transition period of five years from the approval with full implementation expected end of 2018.

The Policy defines Open Access as *“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 licenses in line with the CGIAR IA Principles), subject to proper attribution”* and commits CGIAR to making all information products OA. The main provisions of the Policy are given in Box 1. The Policy focuses on final research outputs – *“those information products (regardless of format) that are “stable” and unlikely to undergo further change (e.g., post-publication materials, datasets collected over the life of a project that has ended, etc.)”*.

The Policy commits CGIAR to openness, organized access to and dissemination of information products, ensuring interoperability and equitable access immaterial of internet limitations or language barriers while abiding by the privacy and copyright requirements over resources.

Importantly, although the name of the Policy refers to data management, the review team observed that the Policy does not govern data management at Centers in a comprehensive manner, but concerns data management only to the extent required for OD and OA.

Shortly following the adoption of the Policy, CGIAR hosted a Data Summit¹³ bringing together key stakeholders from within CGIAR and partner organizations, for developing a roadmap for OA in CGIAR, especially for data. This led to development of the Implementation Guidelines that, through consultation with key partners such as FAO and the Centre for Agriculture and Biosciences International, were finalized in July 2014. The Guidelines focus on planning and coordination to guide and empower Centers’ development of their own OA/DM implementation plans in the first phase of the transition. The Guidelines aimed to be broad in nature and *“to offer as much flexibility as possible for Centers in planning for and preparing their own implementation plans”*. Beyond that, they present the elements of implementation support to include an Open Access and Open Data Support Pack

¹³ Meeting Report available online: <https://cgspace.cgiar.org/handle/10947/2892>

(OA/OD Support Pack)¹⁴ for detailed guidance and recommendations to support the implementation of and compliance to the Policy, communities of practice (COP) on OA access and DM (see section 4.3), and capacity development workshops. The Guidelines also define the roles of the Centers, System Organization and partners in relation with OA.

2.3 Scope and adequacy

The OA/DM Policy addresses all types of information products including 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 online platforms); and metadata associated with the information products.

Prior to finalization, the draft Policy was formally reviewed by CGIAR's Independent Science and Partnership Council (ISPC) that provided a commentary.¹⁵ While the ISPC found the intent of the Policy laudable, it stated that the Policy fell short of describing the means, limitations, penalties and potential costs of full OA in CGIAR. Examples of this include the cost and risk of making all data OA and concern related to lower quality publications and data.

The team considers that the Policy is generally adequate and comprehensive given the CGIAR mandate and the variety of the CGIAR information products. This was confirmed by most of the interviewed stakeholders who agreed that the policy is clear and concise, compiled as a policy document should be, presenting the fundamental principles and concepts. Most of the researchers surveyed considered that the Policy is beneficial for their research objectives and CGIAR's objectives in general (see also section 3.1 on effectiveness). The review team considers that the issues and concerns raised by the ISPC are best addressed in the next version of the Implementation Guidelines, which should provide guidance on governance, enforcement and monitoring.

The Policy was seen by data managers, in particular, as weak in terms of data management, but the team considers that this is intentional. While the Policy covers all types of resources, including data, it addresses data management only to extent that it is relevant and necessary for OD, which is clearly specified in the Implementation Guidelines. The team, however, agrees that the Policy name is misleading, and this is the cause, along with the fact that OD depends on a solid data management approach, of the high expectations of the data managers concerning the policy addressing data management issues. The team concurs with data managers that data management is a prerequisite for OD and although it cannot be fully addressed in the context of the Policy, it is crucial for its implementation (see section 3.1.4 for discussion of data management in the OD context). This is especially true in light of the FAIR principles that set specific requirements for data, including adequate mechanism for their re-use. Current OD practices in CGIAR Centers indicate that there is high variability across Centers in data management workflows, practices and data repository use.¹⁶ It is clear to the team that data management policies should remain Center specific, but with System-wide input and guidance on what the end result should be: high quality, prioritized data outputs following

¹⁴ <https://sites.google.com/a/cgexchange.org/oad-support-pack/>

¹⁵ https://ispc.cgiar.org/sites/default/files/ISPC_Comments_OpenAccessPolicy_0.pdf

¹⁶ Devare et al. 2017. Open Access and Open Data at CGIAR: challenges and solutions. Knowledge Management for Development Journal 13: 7-21. (<http://journal.km4dev.org/index.php/km4dj/article/viewFile/354/425>)

FAIR principles. The review team notes that the Big Data Platform (see section 3.2.4) is charged to address issues of OD in particular.

Box 1. Policy Statement, general points

Openness: The policy mandates efforts to make all information products Open Access, subject to legal issues as applicable.

Suitable repositories. The policy directs utilizing stable, open access repositories to enable users and other sites and search engines to access or locate information products. Application programming interfaces (APIs) or other mechanisms enabling the information products to be available from the CGIAR website and associated web-based products should be deployed. It further encourages the use of existing repositories in order to overcome technical barriers such as interoperability issues when multiple repositories are deployed.

Interoperability: prescribes implementation of standardized metadata for resource description. The aim is to achieve syntactic and semantic interoperability as one of the keys to enabling international and interdisciplinary access to information products. It also prescribes for the information products to be stored and delivered using appropriate protocols and formats to ensure that their content can be discovered, shared and incorporated across different technological platforms.

Data storage and preservation for future use: the policy also covers the need for preservation of data and Information products. It prescribes that data and information must be stored where users can find them and where they will be preserved for future use. The resources need to be managed, maintained and curated to be transient over time.

Copyright and Open Licenses: Suitable open licenses shall be used that recognize the legal rights to information products and encourage their use and adaptation (ideally without violation of any rights that may be applicable).

Incentives and professional expertise: This is one of the important provisions of the policy where it encourages offering incentives to the professionals that help them work for compliance with the Open access policy. It states that incentives towards the development of professional expertise in all areas of Open Access and data management shall be devised, adopted and promoted.

Translation: The policy encourages catering to multilingual communities across the world and that at least the content of key documents must be conveyed. It states that *“translations of key documents and other media into pertinent languages are encouraged. All versions should be deposited in suitable repositories and made Open Access”*.

Limited internet connectivity: The policy tends to be inclusive of those communities that may not have the required bandwidth to access OA resources. It states that *“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”*.

Open Access and Data Management Plans: The policy encourages the CGIAR Centers and CRPs to draw up their Open Access and Data Management Plans (OADMPs). It states that *“OA and DMP 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”*.

The main content of the Guidelines concerns Centers' development of OA/DM implementation plans. The Guidelines prescribe that *“Centers should develop their own specific and verifiable implementation guidelines and plans, according to the Policy. Such plans should be tailored by each Center to their specific research programs, resources, and previous experience with Open Access and Data Management.”* The team considered that the Guidelines are comprehensive and complete in prescribing the approach to developing Center/CRP implementation plans. The detail in the Guidelines is mostly adequate including a number of concrete action points for each area of OA implementation. The plan includes both essential elements and minimum requirements, with explanatory text for each area. For example, each Center, at a minimum, needs to include information on how all types of information products are collected, stored, and disseminated, and also to include information on

types of repositories used, and links (URL) to each. The Guidelines also provide a detailed overview of the minimum set of requirements concerning Policy implementation such as technical infrastructure including repositories, metadata, and interoperability, which the team considered adequate and useful. Given that the Centers' role as having primary responsibility and accountability for day-to-day implementation of OA was clearly acknowledged in the Guidelines, the review team found the emphasis on guiding Centers to be appropriate and in line with a perspective where individual Centers implement a policy tailored to their needs under the unifying umbrella of the System Organization.

The team also assessed the extent to which data management (in the context of OD), licensing and ethical issues are covered in the Guidelines. The team concurs with the views expressed by data managers that rather than revising the Policy itself, the data management issues related to OD can be addressed through expanding the Implementation Guidelines. Any development of a System-wide practices and support for data management in OD context should be considered through experiences from the Big Data Platform.

The interviews revealed that the Guidelines were not considered sufficiently detailed on legal requirements for the Centers to assess whether datasets can be made open or not (due to restrictions), and how to deal with any restrictions about sharing data. In this regard there needs to be complementarity between the Implementation Guidelines and guidance material for the IA Principles.

Regarding data quality (see also section 3.1.4) and quality of information products in general, the Policy highlights that some outputs may be in draft form, incomplete or of poor quality. It prescribes that *“some judgment therefore, needs to be made over the information products that will be made Open Access”*. However, the review team considers that Guidelines need to be expanded to provide more concrete guidance for enhancing data quality uniformly across CGIAR.

The Guidelines called for development of the OA/OD Support Pack, which was seen as the living, digital, online resource for making all guidance and reference documents to OA in CGIAR available. As a result, the Policy at the moment is not accompanied by a specific, updated guidelines document focusing on implementation issues beyond the early preparatory phase. The review team considers this as one of the reasons that several stakeholders reported uncertainty in issues like compliance measurement and metrics, reporting and governance of the Policy.

The team considers that, in addition to updating the OA/OD Support Pack as needed (see section 4.2 about the OA/OD Support Pack), the Guidelines need updating, given that they were initially prepared for first stages of transition and approved before Phase I of the implementation. The concerns raised by the ISPC regarding mechanisms for policy enforcement, incentives, sanctions and monitoring can to some extent be addressed in revised Guidelines, but as discussed later in this report, there needs to be clearer understanding of what full compliance means before the mechanisms of enforcement can be put in place.

2.4 Conclusions

The OA/DM Policy which is presented and to be read in conjunction with the Implementation Guidelines, is sufficiently broad in terms of scope and coverage given CGIAR's mission, range of information products and aspects of operationalizing the policy. It is appropriate in terms of laying

down the expectations for different types of information products. The Policy aims at rationality and efficiency in making data products open access by emphasizing interoperability and ease of use. Although the Policy became effective before the FAIR principles were introduced among OA/OD global communities, the Policy is fully aligned with those principles to be enforced in CGIAR and to be used for guiding implementation and monitoring of the Policy. It is clear that since approval of the Policy and Guidelines, CGIAR discourse about OA/OD has evolved reflecting what the Policy intends to achieve, and this is a clear indication of its appropriateness. The discussion in the OD domain has progressed from the notion of just putting resources online to ensuring compliance with FAIR principles and offering high-quality information products. Currently FAIR principles are thoroughly incorporated in the design, implementation and reporting of OD efforts in CGIAR.

OD is the most complex area within the Policy and it has implications on how data are managed. Despite the name of the Policy, which the team found misleading, it does not govern all data management. The team does not think the Policy should cover data management in full, because it is part of regular research process and management at Centers. However, without proper data management, OD cannot be realized.

From the evidence and perceptions received, the review team concludes that the Guidelines need to be revised and extended to cover many more aspects of OA/OD where CGIAR researchers, and knowledge and data managers need best practice and implementation guidance. Initially, the Guidelines had a very narrow purpose of facilitating each Center's preparation of its own OA/OD implementation guidelines and plans. In particular, further clarity on aspects of data management that are essential for OD, such as data quality requirements, needs to be included in the Guidelines.

Recommendation 1: Taking into account lessons from the implementation to date, the Implementation Guidelines should be expanded to reflect the current phase of implementation and future needs at Centers and programs. The additional content should include the following:

- clarity in System-level governance and management roles for both OA and OD oversight and reporting reflecting CGIAR's governance transition in 2016 and establishment of the Big Data Platform;
- aspects of data management that are particularly pertinent to open data, such as data prioritization, curation and quality;
- templates for legal compliance and process for legal clearance (complementing guidelines for intellectual assets policy);
- guidance for workflows for tracking progress towards OA/OD, and sharing lessons;
- guidance for compliance to ethics in data publishing;
- minimum requirements of repositories for open access;
- guidance on promotion and incentives.

3. Effectiveness and Efficiency

3.1 Effectiveness

3.1.1 Introduction

Given that the Policy is still being rolled out and full compliance is not expected until end of 2018, the review team in its assessment of effectiveness considered progress made in the implementation of the Policy, the extent to which best practices have been promoted, relevance and adequacy of support for achieving full compliance in the agreed timeline (2018), effects of the Policy implementation in Center operations and communications, and relations with partners. It also considered the extent of awareness of the OA/DM Policy and OA in general, Centers' OA/DM implementation plans, and aspects of infrastructure, such as data and information repositories.

3.1.2 Awareness

It is evident from the review that the Policy, through its implementation, has raised awareness about OA and OD across CGIAR and, specifically, amongst CGIAR researchers. Practically all researchers who responded to the review questionnaire were aware of the Policy and most of them felt very familiar with it. Most researchers considered that the Policy is beneficial for their research objectives and CGIAR's objectives in general.

The review team found that researchers strongly recognize the benefits of the Policy related to open access for publications. The benefits they mentioned include: higher visibility of their research, both for their own benefit (higher citation index, the possibility to access more publications by other scientists without a fee) and also that of other researchers, especially those with limited resources and access, mostly in developing countries, which is particularly important for CGIAR's mission. The researchers consider that higher visibility means more feedback for their work and the prospect of new collaborations and funding, and consequently increased impact of their work. To this extent, the policy has efficiently brought about awareness of benefits of OA in the CGIAR research community.

Concerning OA related to publications, researchers expressed some concerns of OA journals becoming the preferred publishing option. The review team emphasizes that quality of a journal and effectiveness of publishing are determined by multiple parameters (credible peer review practice, impact factor, open access). When selecting a journal, researchers will consider all these parameters. Making a publication OA at the end does not depend on the choice of the journal, as nearly all journals agree with the common practice of depositing pre- or post-prints of the article in open institutional repositories.

The team notes that publishing both in OA journals and in journals that have subscription fees have costs for authors. OA journals charge authors/their institutions rather than users. Some donors have started to cover the costs of making publications OA.

Some researchers commented that the access policies of journals and book publishers vary and change and researchers would be well served if a regularly updated reference were made available centrally so that budgets and dissemination plans are easier to make. In addition to costs of

publishing, some researchers considered that maintaining an institutional repository adds to their workload.

The team considers that an OA strategy should be included already at the project formulation and design stage. This would help the organization to address the issues of OA regarding scientific publishing (as well as OD), setting up plans for necessary support, incentives and budgeting already at project application stage.

With regard to OD, the team found that the researchers do not recognize its benefits and potential (for instance publishing their datasets) to the same extent as for OA research publishing. Although some feel that there is a large number of underutilized datasets at Centers and recognize the potential of the broader use of those data, some researchers consider OD as an additional resource requirement and burden.

Data managers confirmed that for researchers making OD an integral part of their work is a challenge. The review team concurs with them that in order for the Policy to be effective, researchers' mind set towards OD should change, as it has started to change for making publications OA. Of all stakeholders interviewed, data managers and open access professionals in particular feel that while this "culture shift" towards OA and OD is happening, it will take some time for researchers to recognize that openly sharing all knowledge assets resulting from research, publications as well as data, is important.

A donor representative interviewed defined awareness of OA/OD as *"the idea that the researchers as part of their day to day work naturally and consistently feel and believe that the management, collection, curation and dissemination of data is as natural part of their work as they would now think about publishing reports and articles. Making OA and OD an integral part of the story that we tell ourselves that our work is - that is the target endpoint we should be looking for."*

To this end, the review team emphasizes the need for both the Center and System Organization management to continue the positive push towards raising awareness among researchers of the whole spectrum of the Policy, both OA and OD, through coordinated advocacy and incentivization mechanisms.

3.1.3 OA/DM implementation plans

The Policy encourages Centers, and also CRPs, to prepare OA/DM implementation plans to ensure implementation of this Policy by each Center and across CGIAR. The Guidelines state that: *"Plans should address all of the elements specified by the Policy. Likewise, all aspects of implementation – particularly in terms of the technical infrastructure including repositories, metadata, and interoperability – should comply with the minimum parameters set forth in the Policy and these Guidelines."* The Guidelines offer detailed instructions and a template for compilation of the plans, including guidance on technical infrastructure, metadata, copyright and licensing, resourcing and budgeting, staffing issues and even considerations on advocating and incentivizing the Policy, metrics and reporting (see assessment of the Guidelines in section 2.2).

The (then) Consortium Office issued an OA/OD implementation plan template to help Centers draft their implementation plans, which was expected to happen by the end of 2014. Regional workshops were organized for using the template to draft implementation plans. Most Centers completed and

submitted to the Consortium Office their implementation plans. The team found that for the large majority of the Centers, the guidelines and templates have helped standardize the process of formulating implementation plans. In addition to pursuing this matter in the interviews, the team sampled three Centers' implementation plans for more detail. Each followed the template and guidelines, indicating that the reference and guidance for formulating the implementation plans had been useful.

As mentioned earlier, two aspects were challenging for the Centers. First, the Centers seemed to still lack sufficient understanding and know-how for dealing with (legal) restrictions/compliance to opening up data. Second, the Centers did not have sufficient guidance on governance and management of OA/OD, which has financial, manpower, capacity and information technology implications.

The review team concludes that preparing implementation plans has been very useful for those Centers that have completed their plans, not only for producing an actionable document tailored to the needs of each Center but also as a useful exercise that can reveal weaknesses and challenges in implementation of the Policy. It has also helped highlight needs for further elaboration and guidance on specific topics and management of the Policy. The team therefore considers that it would be useful for also those Centers not yet having a plan to develop one. The plans should not be seen as guiding a static implementation of a set of rules but as living documents of guidelines to promote OA to scientific publications and implementation of the FAIR principles in data management.

3.1.4 Data management challenges

Effective data management processes are a prerequisite for OD and amongst the most challenging issues for implementation of the Policy. Data managers interviewed acknowledged that there are several issues concerning OD that need to be resolved. These have been extensively discussed in the COPs and are being addressed. The review team focuses in this section on the main challenges that are relevant for data management in the context of OD and implementation of the OA/DM policy. These include data quality, data prioritization and co-ownership of data.

Delivering high-quality data products

The emphasis of the OA/DM Policy is on final research outputs. Therefore, data management in the context of the policy needs to be discussed only to the extent that it concerns final outputs, and data management across all stages of research implementation remains a Center management matter.

So far, focus on implementation of the policy has been on making datasets that can be considered final research outputs available and open, according to the FAIR principles. According to the data managers interviewed the COPs should move forward to considering the quality requirements for publishing high-quality datasets and promoting rapid, seamless and secure exchange of useful, standards-based information. The challenge is not to simply openly publish as many CGIAR datasets as possible, but rather, as some information and communication managers have noted, to offer usable and accessible information products that are meaningful for targeted stakeholder groups, thus developing "high-value" information products.

Data managers interviewed felt that consensus with the scientists and subsequently clear guidelines are missing on what can be considered a full dataset and how it should be described for publication. Quoting summary of the results of the 2014 OA/DM practices survey (see also discussion in section 3.1.5) *“practices for data quality and/or data cleaning are not consistent among Centers, and responsibilities for these also vary. Several Centers (5) indicated that no one is specifically responsible for assessing data quality or cleaning data. Five other Centers indicated that researchers, project leaders, or science theme leaders are responsible, while 6 Centers indicated that they have a research method or similar unit that is involved.”* This seems to be the situation still today.

FAIR principles also imply that preparing datasets for publishing involves investment in data curation, including cleaning and full metadata descriptions. Data papers may need to be prepared.¹⁷ The OA/OD Support Pack (discussed in section 4.2) offers only minimum guidance on data quality. However, dataset preparation guidelines by individual Centers, where such exist, may help others; the OA/OD Support Pack contains one example of such Center-specific guidelines. The team concludes that, while its interviews indicate that progress has been made in the past three years, data quality matters remain an issue that Centers/CRPs need to pay attention to.

The Policy defines the timeframe to publish datasets to be *“within 12 months of completion of data collection or appropriate project milestone, or within 6 months of publication of the information products underpinned by that data”*. Some interviewees were concerned that 12 months is insufficient to clean, analyze and publish data with the manpower and resources available to them and proposed 24 months as a more realistic timeframe. Similar concerns apply to other products like software, expected to go through several versions and upgrades before a stable version of the software may be released. The review teams suggests that some flexibility to go beyond 12 months may be needed.

Data prioritization

The Policy and Guidelines emphasize selection of “stable” and “final” datasets for publishing. The requirement for extensive data curation, which is part of preparing data for publishing, poses a limit to the number of datasets that it is feasible to make FAIR. Prioritization of datasets is therefore essential for policy implementation, but it is also a challenge in the review team’s view. This was acknowledged in the Phase I project that proposed to *“develop a framework to prioritize legacy data, customizing for each data stream as necessary”*.¹⁸ The project progress report in January 2016¹⁹ reports that *“a data prioritization framework is being drafted [...] with members of the [DMTF] soon to be involved to ensure that the framework will hold true for varied data streams”*. Data prioritization was a major topic in the DMTF meeting in 2017, which resulted in an early draft document of a data prioritization framework. The team considered that the draft prioritization framework is much needed

¹⁷ A data paper is a peer reviewed document describing a dataset, published in a peer reviewed journal. It takes effort to prepare, curate and describe data. Data papers provide recognition for this effort by means of a scholarly article. (<https://www.gbif.org/data-papers>)

¹⁸ CGIAR Open Access and Open Data Phase I (Years 1-2): Assessment, Prioritization, and Coordination of CGIAR’s Current Open Environment, Phase 1 Proposal to Fund Council, November 2014, (<http://hdl.handle.net/10947/3737>)

¹⁹ CGIAR Open Access and Open Data Phase I (2015): Assessment, Prioritization, and Coordination of CGIAR’s Current Open Environment, CGIAR OA-OD Phase 1 project report – Jan 2016, <http://cgiarweb.s3.amazonaws.com/wp-content/uploads/2016/03/CGIAR-Open-Access-and-Open-Data-Phase-I-2015-Progress-Report.pdf>

and considers its finalization as an important step towards supporting the implementation of the Policy for data.

Licensing and ethics

As per the OA/DM policy, openness is always subject 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”*.

Article 5 of the Implementation Guidelines enlists the legal aspects guiding the management of IA and IP rights covering agreements and contracts, legal compliance, maintenance of IP portfolio, regular audits and IP diligence. However stakeholders interviewed considered that these aspects need in-depth guidance on what exact legalities are involved and how to deal with them.

Article 6.2 of the Guidelines allows for *“Limited Exclusivity Agreements”*. Exclusivity refers to excluding datasets from being open for a limited time period, or geographical/domain purview. The clause, however, includes research exemption which means that the data and information resources should be made available for research purposes even though they may be under *“Exclusivity”* exemption. At the moment the Guidelines highlight exclusivity as an exception to making the data open but offer no additional guidance as to how these datasets are to be identified by researchers as such and how they should be handled.

While there is a consensus in CGIAR that much of CGIAR data needs to be open, stakeholders interviewed highlighted that data privacy is a crucial issue and a challenge for OD. This is also reflected in some Centers’ implementation plans. There are also ethical aspects to consider. CGIAR Centers work in an international context and face a variety of national legislatures concerning the collection of personal data. Some Centers have policies and processes in place (e.g. Institutional Review Boards for ethics approvals) but this differs from Center to Center according to need and capacity.²⁰ Publishing data could in some political contexts be even dangerous to the local population surveyed. Anonymization of the data set is not enough and removing all information on the population could render the data useless for the specific purposes it was collected for. The review team considers it important that the requirement for openness does not concern data for which publishing could have negative implications on the subjects of the research in question. In revising the Guidelines, clear guidance should be provided to apply robust ethical rules for deciding on confidentiality of data, use of confidential data in research, and under which conditions data considered restricted might be shared. In these matters, the two OA/DM COPs (see section 4.2 for discussion of these communities) and the CGIAR Consortium Legal/IP Network (CLIPnet)²¹ concerned with the IA Principles should interact with each other. This will give Centers the possibility to improve the implementation plans, which, judging by the three plans reviewed by the team, have not dealt with these issues in depth.

The recent review of CGIAR’s IA Principles²² stresses the need for an overall IA management and explicitly mentions *“the System Management Board should oversee the development and deployment*

²⁰ The Big Data Platform has commissioned a study to map privacy and ethics capacity/needs across CGIAR Centers, forthcoming in 2018.

²¹ CLIPnet is comprised of mainly Center Legal/IP Focal Points, and representatives from the System Management Office

²² <http://iea.cgiar.org/wp-content/uploads/2017/02/Final-Report.Principles-of-Intellectual-Assets-Review.Oct-2017.pdf>

of a communications program to create a more uniform understanding across the CGIAR System”. Furthermore, the review experts point out an important aspect of training in the form of formal services to be provided to the Centers. Trainings through workshops and resources in the form Q&A are also recommended. The review team concurs with these observations with respect IA and legal aspects of the OA/OD policy.

3.1.5 Status of implementation

Although the objective of the review team was not to assess compliance with the Policy but rather the supporting processes that promote it, the team in this section briefly discusses the extent of implementation to date drawing from recent CGIAR-wide data. Currently, data on OA implementation (for both publications and data sets) are presented only in aggregate form²³. The review team had access to an unpublished report “Discoverability of Publications and Datasets across CGIAR” which included overall statistics for each Center.

As part of the Phase I OA/OD project, the Consortium Office conducted a survey in 2015, designed as a 58-question questionnaire to staff responsible for OA/DM matters at Centers. The aim was to determine the OA and OD landscape across CGIAR, including data management and quality practices; and gaps and needs in human resources and enabling environments for OA/OD. The summary findings from the survey were reported in the CGIAR OA-OD Phase 1 project report in 2016²⁴ and in the recent publication by Devare et al (2017). According to the results, Centers were approaching their OA/OD operations in different ways through a wide range of approaches, priorities, and practices. Recent CGIAR-wide statistics show that there are still very large differences among Centers in terms of accessibility, particularly of data, but also regarding research publications.

With implementation not scheduled until 2018, the review team did not conduct a survey of implementation progress. The team, however, sought to review readily-available information and data on OA, and especially OD efforts. Information on data and publications were compiled through harvesting OA records from CGIAR Centers, was available in the above mentioned report commissioned by the Big Data Platform. Since 2014 (and before), the number of datasets has increased very rapidly across CGIAR (Figure 2) in most Centers, except four Centers that in 2017 reported less than 10 published datasets. The increase has been particularly rapid in four Centers that published an average of 9 datasets in 2014 and nearly 220 (on average) in 2017. There are two Centers that already in 2014 had >100 datasets and in 2017 reported 367 and 458 published datasets, respectively, and about half of them completely OA.

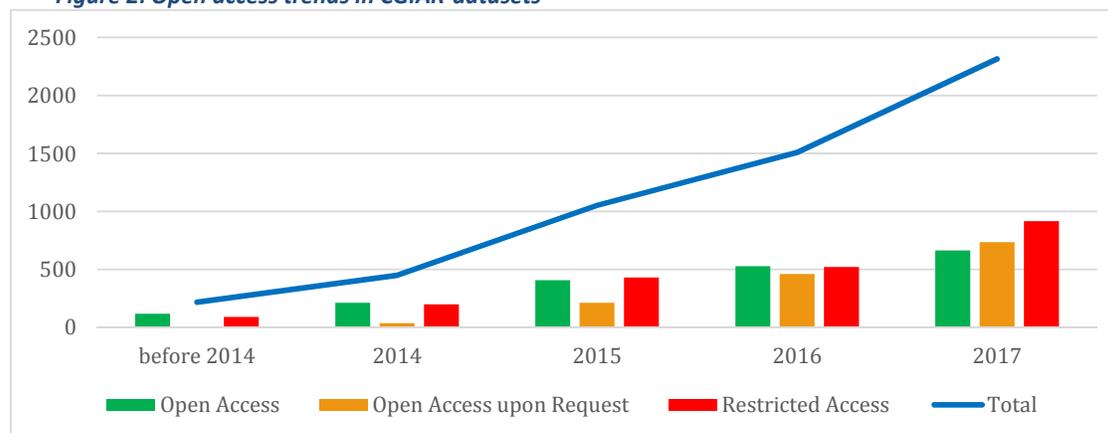
Corresponding with the overall increase in datasets, the number of OA datasets (either fully open or accessible upon request) has also grown across CGIAR from 119 in 2014 to 664 in 2017. Yet, the proportion of OA datasets of all published datasets has not changed but remains at about 60 percent if datasets fully OA and OA upon request are both included. The percent of fully OA datasets has actually gone down, from 48 in 2014, and 55 before that, to 29 percent in 2017. Although the 2015 survey report concluded that almost every Center was taking steps towards making publications and data open, the situation concerning the total, increasing number of datasets does not attest a

²³ <http://ceres.bigdata.cgiar.org/statistics.php>

²⁴ <http://cgiarweb.s3.amazonaws.com/wp-content/uploads/2016/03/CGIAR-Open-Access-and-Open-Data-Phase-I-2015-Progress-Report.pdf>

uniformly positive trend. One Center that previously had no accessible datasets, reported in 2017 only one OA dataset – a mere one percent of its datasets.

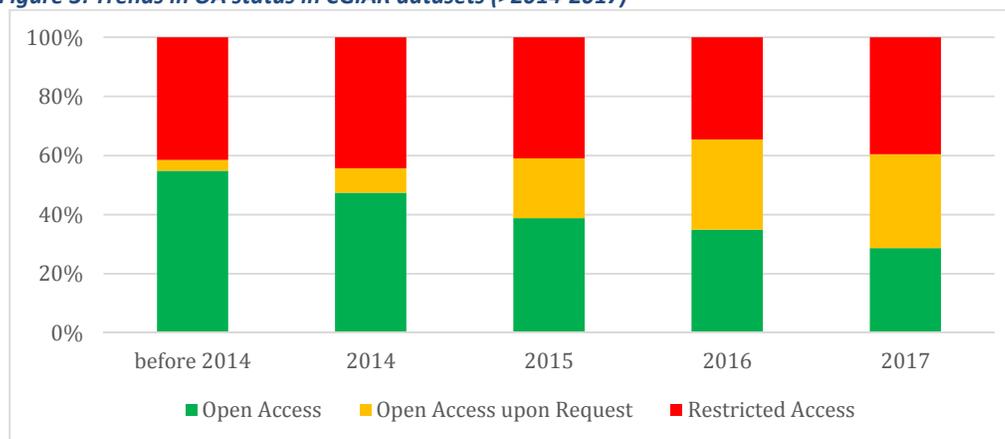
Figure 2: Open access trends in CGIAR-datasets



Source: Unpublished CGIAR report on discoverability of publications and datasets across CGIAR, collected as part of the Big Data Platform review of Center trends in OA

In 2017, the situation in nine Centers was quite positive regarding OA datasets. One Center had 78 percent of its published datasets fully OA and four other Centers had about 40-50 percent of their datasets fully OA. A further four Centers had also large proportions of their datasets OA (about 70-90 percent), but mostly upon request. Figure 3 shows that the category “OA upon request” has grown fastest. The number of datasets in that category in 2017 was over 90 times more than before 2014. However, only five Centers included above reported considerable number of datasets in this category and others practically none.

Figure 3: Trends in OA status in CGIAR datasets (>2014-2017)



Source: Unpublished CGIAR report on discoverability of publications and datasets across CGIAR, collected as part of the Big Data Platform review of Center trends in OA

At the time of the 2015 survey, eight Centers reported having an OA publications policy in place (separate from the OA/DM Policy). The percentage of peer-reviewed CGIAR publications downloadable without restriction was growing but still highly variable across Centers. Two Centers reported that over three-quarters of their peer-reviewed publications were fully-downloadable and most other Centers reported that close to half (+/- 10 percent) were OA.

In the last three years (2015-2017) the percentage of OA peer-reviewed publications across CGIAR has been close to 80 (in 2017 73 percent were OA). However, as the review team considered only the aggregate representation of OA publications, it did not estimate individual Centers' progress in making their publications OA, nor the variability among Centers.

As with datasets, number of peer-reviewed publications with CGIAR authorship has also grown steadily, which has meant increased cost in meeting OA requirement. In 2015, only one Center reported as having a centralized fund to pay article processing charges when publishing in OA journals. The financial cost of making publications and articles OA is still a concern for many. One researcher offered an example stating that he publishes 10 articles a year, which would result in USD 30,000-40,000 per year for OA rights. The budgetary implications of OA need to be planned and addressed accordingly, and supported by donors and Center senior management.

For full implementation of the Policy, it would be important for Centers and those managing and governing the implementation of the Policy to clearly define what is understood by full implementation. As indicated above, full implementation is easy to determine for (scientific) publications where the expectation is that all (100 percent) are downloadable without restrictions. For OD, full implementation is more difficult to determine and interviews with data managers revealed a great deal of confusion. Data managers highlighted the need for uniform indicators of implementation to be reported against, and the team noted that such indicators are already being developed. The indicators that have recently been drafted for OA data for monitoring (as part of results-based management) appropriately reflect the FAIR principles and progress towards full implementation of the Policy.²⁵ (See further discussion in section 5.5 under Reporting). As part of the indicator development, degrees of "FAIRness" are also being introduced.

Most data managers interviewed felt skeptical as to whether full implementation concerning OD will be possible within 2018. The review team agrees that before assuming that full implementation is possible, there should be agreement and a clear understanding of what it means for data. The review team considers that if CGIAR, in addition to endorsing indicators for reporting on OA and OD, can solve the most pressing issues concerning OD, namely preparing a data prioritization framework and a plan for making the prioritized datasets public, including a timeframe, it will be able to define "full compliance" also for data.

3.1.6 Promotion and adoption of best practices

The OA/DM Policy includes a statement about incentives and professional expertise (see Box 1), but the Guidelines and other support material have little guidance to help Centers and CRPs to incentivize professionals for greater compliance with the Policy. The review team received consistent feedback that there is a need for specific promotion of the OA/DM Policy and concrete incentives for OA/OD (see section 4.3 for discussion of the COPs). The team found broad agreement that future culture shift towards OA/OD is considered necessary and that it will happen only when the management and scientists are fully convinced about the benefits of OA/OD and engaged in the implementation. The

²⁵ 2017 RBM Indicators for Open Access/FAIR (Findable, Accessible, Interoperable, Reusable), Prepared by: Medha Devare, Paola Camargo, David Abreu, Ryan Miller, Indira Yerramareddy, Rodrigo Sara, Leroy Mwanzia, Draft 14/8/2017

team considers that guidance materials alone have limited potential to promote and catalyze culture shift. As discussed below, this requires sharing experiences and management championing.

Incentives have been frequently discussed by the COPs (see discussion of the COPs in section 4.3), and in the latest meetings the participants discussed specifically how to incentivize data sharing and proposed different incentives for promoting the policy objectives. These included establishing competitions and OD awards. Some of the Centers reported that they have already used these particular incentives successfully, according to their view.

The review team looked at CIP's experience as an example of successful incentivization of OD. In October 2016, CIP launched the Open Data Sprint. This initiative aimed to encourage researchers to put data into CIP's Dataverse. The goal was to have about 100 datasets within five months. The division with the most datasets entered into Dataverse was offered to get funding for two OA articles. CIP also proposed incentives to individuals, which was one raffle ticket for each dataset entered into Dataverse. The raffle price was funding for one OA article and attendance at one scientific conference. CIP used simple metrics (charts, maps) to help divisions and scientists to track their progress in Dataverse.²⁶ The target of the Data Sprint was not only achieved, but surpassed to over 130 OA datasets within the time frame set.

Another proposed incentive was to take individual OA/OD efforts and results into account during staff performance evaluations. Resolving authorship of datasets and including researchers as dataset authors would also incentivize practices for OD. One researcher stated that the Center places "no value on publication of datasets" and that all current data are published in the name of Center with no researchers name, providing no personal incentive to publishing and sharing datasets. Furthermore, Digital Object Identifier citation of the dataset could be included in the publication to ensure that the data generators, who are not always the publication authors, have appropriate credit. Some Centers have also already incentivized peer-reviewed data papers authored by the dataset creators. Also donors, many of whom have their own OA requirements have considered incentives, such as rewards and resourcing (see section 3.2.3).

Those interviewed felt that the Policy should be promoted more strongly through success stories and by identifying "champions" at Centers. The review team considers that Centers should adopt suitable positive incentives, including the kinds already tested in some Centers and discussed by the COP members, and use them together with stronger endorsement by management enforcement of the Policy to reach compliance.

Overall, Center stakeholders interviewed felt that the Policy has enhanced good practice and brought forth very positive outcomes. One concrete result is the CG Core Metadata Schema introduced in CGIAR in 2015, and close to being finalized.²⁷ The CG Core Schema is intended to be "*the minimum set of elements applicable across CGIAR Centers, data streams, and formats*" and is expected to "*enable consistent annotation of final research products and adherence to OA-OD under the "FAIR" principles*". The CG Core Schema will also allow meta-searching and indexing across CGIAR repositories and databases and inter-linking across multiple resources. This schema is closely aligned with Dublin

²⁶ <https://docs.google.com/presentation/d/1wmaua5r1YCKBTcLlFSo4ochPuPUD2RuTffUcV7oLF8k/edit#slide=id.p13>

²⁷ <http://repo.mel.cgiar.org/handle/20.500.11766/4764>

Core²⁸, a generic and widely-adopted metadata schema that has been in use since the mid-1990s; and to the Data Documentation Initiative²⁹, which is an international standard for describing data, used by many CGIAR Centers. Nearly all Centers/CRPs reported using Dublin Core and/or planned to use CG Core Schema (see also section 3.2.5 on Infrastructure).

However, there is not yet a uniform approach to the use of CG Core Schema, and for that it would be helpful to include in the support material clearer guidelines on what can be considered a full dataset and how it should be described for publication (see also section 3.1.4 about data quality).

The review team recognizes the effort within CGIAR to bring its practice in line with recent developments. CGIAR, in working towards implementation of the OA/DM Policy, needs to stay connected to the global discussions. Technical issues such as data interoperability are being addressed by the CG Core Schema initiative, but other issues currently discussed in the data science community include privacy and data ownership, especially concerning field trial data. The CGIAR OA managers need to stay abreast of the latest advances made in the open science community for identifying and promoting good practices also in areas that currently remain challenging.

3.2 Efficiency

3.2.1 Introduction

The review team assessed the extent to which communications, processes and infrastructure have been enhanced for efficient implementation of the OA/DM Policy and information and data management associated to it. The team reviewed efforts to streamline practices and use of existing infrastructure to host and manage OA/DM resources, including CG Core Schema (see also section 3.1.5 “Status of Implementation”).

3.2.2 Communication

Implementation of the OA/DM Policy concerns a range of stakeholders in Centers, CRPs and the System Organization that need to align their activities to serve common objectives. The activities of the two COPs, OAWG and DMTF, are assessed in section 4 about Implementation Support. Apart from feed-back on communication and sharing across CGIAR facilitated by those COPs, data managers called for more interaction between data and knowledge managers and those Center staff dealing with legal issues and IP (CLIPnet). As mentioned earlier (section 3.1.4), legal expertise is needed for best practice guidance on issues such as licensing. The OA/DM Policy makes only a brief mention about copyright and licensing: *“Suitable open licenses shall be used that recognize the legal rights to information products and encourage their use and adaptation”*. However, CGIAR’s IA Principles and Guidelines govern these issues. Therefore, the CLIPnet members concerned with all aspects of IP should be consulted in cases where data managers fear that requirements for OA may conflict with Centers’ use of IP rights, especially in cases where third parties like companies are involved in the creation of the specific datasets. The communities also need to include ethical issues in their agenda, and connect with staff or committees with the responsibility and competence (see section 3.1.4).

²⁸ <http://dublincore.org/>

²⁹ <https://www.ddialliance.org/>

These concerns expressed by data managers highlighted to the review team the need to promote communication across the different communities and staff that deal with knowledge assets, IA and ethical issues related to their management. In particular, linkages between the OA and DMTF COPs and the IP managers COP is encouraged. Legal issues for data could be discussed through seminars/webinars and a working group consisting of IP Data/Open Access managers could be established to issue guidance.

3.2.3 Harmonization of donor requirements

Several donors have in recent years started to pay attention to OA/OD. In 2014, the Bill and Melinda Gates Foundation instituted a policy which, after a two year transition period, will require unrestricted access to and re-use of all research outputs including data from activities funded fully or in part by the foundation.³⁰ Also other major funders such as DFID, USAID and USDA have stated their positions towards making OA research outputs, including data, resulting from work they fund.

Center staff were concerned that donor requirements are conflicting with each other and with CGIAR's OA/OD Policy. For example, DFID focuses on transparency and accountability, whereas USAID focuses on the value of OD for decision making. None of the donor policies contains a clear definition of OD. Another concrete example is the deadlines imposed for publishing datasets, although this, to date, has not been a big concern as these deadlines are rarely policed. Donor agencies have already recognized issues related to consistency across requirements and policies concerning OD, and steps are being taken towards their resolution. For example, a recent GODAN report commissioned by DFID, Bill and Melinda Gates Foundation and USAID aims to identify opportunities for donors to make OD implementation more efficient and streamlined for their implementing research partners.³¹

The review team considers it important to identify patterns of good practice, which donors, including those outside agriculture, can build on and contribute to through further dialogue. The GODAN report identifies several possibilities (for example through common agreements, definitions, implementation plans and metrics, and ethical standards) to align the adoption of shared guidelines, tools and templates with the aim to reduce the time and cost of policy compliance, and to incentivize researchers to publish by rewarding good quality data production and resourcing data management and publication (see section 3.1.6 on incentives).

The review team considers that CGIAR funders need to be involved in regular discussion about the OA/DM Policy and progress in its implementation. This would allow them to consider their own approaches to OA/OD to seek consistency with the CGIAR Policy. This would also facilitate implementation of and compliance with the Policy at the Centers that currently find it difficult to comply with different requirements from different donors.

³⁰ <https://www.gatesfoundation.org/how-we-work/general-information/open-access-policy>

³¹ GODAN: Donor Open Data Policy and Practice: An Analysis of Five Agriculture Programmes.

<http://www.godan.info/documents/donor-open-data-policy-and-practice-analysis-five-agriculture-programmes>

3.2.4 Infrastructure

The review team assessed infrastructure to the extent that it matters to the efficiency of the policy implementation, not covering technical aspects in detail. It considered use of repositories, and development with the Big Data Platform.

The Policy states that suitable repositories should be utilized, i.e. *“stable, permanent, Open Access repositories, 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).”* The guidelines prescribe that *“Repository systems should meet current industry standards for interoperability and metadata”*.

Publications repositories

According to the OA/DM survey reflecting the status in CGIAR by 2015, most Centers were moving towards DSpace³² or other standards-compliant, interoperable OA repository platforms. Reasons for selecting particular repository systems vary, although being open source, good interoperability, and adoption by other Centers were cited as the top reasons. A number of the Centers have now moved to using CGSpace³³, a digital repository of agricultural research outputs and results that indexes journal articles, reports, conference papers, proceedings, presentations, posters, videos, audio, policy briefs and more from across the CGIAR Centers, research programs, and partners (Box 2). The International Livestock Research Institute (ILRI) developed and launched this repository in 2009, based on the DSpace digital repository. Soon after the debut of CGSpace at ILRI, other CGIAR Centers, research programs, and partners began joining the collaboration. Although this was not the first instance of DSpace in CGIAR, it is the first so-called “multi-tenant” installation³⁴. CGSpace is now being utilized as a central repository space for CGIAR, and is accessible on the CGIAR main webpage.

Box 2. CGSpace

CGSpace is a centralized online publically accessible repository of agricultural research outputs and results produced by different parts of CGIAR and partners. It indexes scientific papers, reports, articles, and press releases, as well as ‘grey literature’ and information products such as presentations, videos, policy briefs and more.

Using CGSpace as a tool, CGIAR outputs and products are made FAIR. CGSpace archives research outputs for referencing, reporting, and posterity, simultaneously serving both open access and publishing needs.

Developed and hosted by ILRI, CGSpace is now a collaboration of several Centers, the CGIAR System management office and CRPs.

(<https://cgspace.cgiar.org/>)

³² www.dspace.org

³³ <https://cgspace.cgiar.org/>

³⁴ Yabowork, A., Orth, A., Ballantyne, P. 2017. Making CGIAR outputs open and accessible: the CGSpace collaboration. Knowledge Management for Development Journal 13(2): 23-33, <http://journal.km4dev.org>

Data repositories

In the 2015 OA/DM survey across Centers and CRPs, the responses indicated a relatively good status regarding data repositories for OD (see section 3.1.5 on Status of Implementation). At that time eight Centers and two CRPs were adopting Dataverse (see Box 3) as the platform for at least some of their data repositories. Two others were using CKAN, which is an open-source data management system for powering data hubs and portals³⁵. As with publication and information product repositories, “good interoperability” was the most-often cited reason for selecting a particular system, followed by the repository being “open source”. Uploads/deposits of datasets appeared more or less evenly split between data managers and scientists.

Interviews with data managers revealed that the situation for data repositories has not changed significantly since 2015. The Centers continue using different tools for the management of data repositories. Some are using the Harvard installation of Dataverse, while others have their own installation. A few use customized versions of CKAN. There is a tradeoff between the easy use of a repository and its customizability. It was acknowledged in the interviews that the high grade of customizability of a CKAN development means a high cost of development and sustainability in terms of need to maintain and update the version. Generally the opinion was that a centralized approach to data storage in CGIAR would not cater for the very different needs of the Centers, but customization is needed. Nevertheless, the review team thinks that the DMTF should look for synergies and for cross-Center use of infrastructures.

Box 3. Dataverse

Developed and hosted at Harvard University, Dataverse is an open source web application to share, preserve, cite, explore and analyze research data. It is open to all scientific data from all disciplines worldwide.

Researchers, data authors, publishers, data distributors, and institutions all receive appropriate credit via a data citation with a persistent identifier (e.g., DOI, or Handle).

Institutions can either choose to use the Harvard University Dataverse for sharing research data or choose to locally hosting Dataverse at their own institution.

Use of Dataverse for sharing research data is on the rise in CGIAR, with Centers and CRPs opting to use Harvard Dataverse to publish their data or locally (Center based) installed application.

The Big Data platform

While this review did not assess the Big Data Platform, which has been operating for a year³⁶, it is briefly discussed here as it represents an important new development in CGIAR for addressing OD issues. This Platform is part of the CGIAR Portfolio approved for 2017-2022. It coordinates and provides leadership in organizing OD. Its objective is to harness the capabilities of big data to accelerate the impact of international agricultural research. The Big Data Platform therefore contributes to implementation of the OA/DM Policy. What was originally conceived as Phase II of implementation support focusing strengthening the volume, variety, and quality of openly-accessible

³⁵ <https://github.com/ckan/ckan>

³⁶ ISPC's assessment of the Big Data Platform revised proposal was positive
https://ispc.cgiar.org/sites/default/files/BIG%20DATA_ISPC%20assessment_14Sep16%20.pdf

research outputs from CGIAR and its partners, was later incorporated into the Big Data Platform as Module 1 “Organize”. A large portion of the Module 1 budget (about 50% of USD 14 million over six years) is dedicated to “address needs articulated by Centers during the OA/OD Phase I, in support of consistent institutional implementation of the OADM Policy and monitoring and evaluation of progress”³⁷ The Platform’s priorities have been developed with the involvement of the DMTF and include building a Harvester as a small pilot to make discoverable the data and publications available. This is the first step towards seamless data integration, analysis, visualization and mapping, which the program hopes to achieve.³⁸

While the Big Data Platform was seen as very positive development by those interviewed, some interviewees expressed concern that OA issues in general may become too much subsumed under the DM agenda brought about by the Platform. Their fear was that consequently investment might get diverted from strengthening the capacities of the individual Centers. The review team considers that the Big Data Platform is an important step forward for demonstrating the power of OD and using it in CGIAR. Nevertheless, this should not in any way diminish the attention, including investment, which the System, Centers and CRPs pay to OA issues and needs, and to the overall implementation of the OA/DM Policy.

3.3 Conclusions

During the transition period the OA/DM Policy has had a very positive effect on awareness of OA trends, benefits and requirements among CGIAR researchers. More work and promotion is nevertheless needed to demonstrate the benefits from OD which is less clear to researchers than making publications OA. For the Policy to realize its objectives, the OA and OD concepts would need to become an integral part of planning and implementing research in CGIAR, which requires wider use of appropriate incentives and promotion. That requires a culture shift, which the team found to be well recognized among CGIAR stakeholders.

The OA/DM Policy transition period focused on each Center putting in place an implementation plan, which was the focus of the Implementation Guidelines. All but a few Centers have developed such plans, but irrespective of having a formal plan, Centers across CGIAR are making progress in implementation. The biggest challenges relate to OD, which is where organizations elsewhere have also faced challenges. Therefore, determining what full compliance of the Policy means regarding data is also challenging and needs to be collectively agreed. These challenges include the mere volume in data and data sets that makes prioritization necessary for selecting data sets for OA. Quality requirements are particularly important. But, there are also ethical issues, for example considering interests of the subjects of surveys, and legal aspects to consider. In the latter, the IA Implementation Guidelines provide complementary guidance. Given the challenges, the timelines set for making data sets OA may have to be reconsidered. In this regard, many donors are demanding OA practices to satisfy for their own policies and harmonization of those policy requirements and the compliance with CGIAR policy would be needed.

³⁷ Big Data Platform full proposal, 2016: <http://hdl.handle.net/10947/4303>

³⁸ A prototype harvester is now online: <http://ceres.bigdata.cgiar.org/>

CGIAR was already making publications, in particular, OA at the time when the Policy was agreed. Since then, there has been considerable progress concerning both publications and data. At the same time as the volume of publications and datasets has increased many folds, the proportion of information products that are partly or fully open has also increased. CGIAR has set up central mechanisms and repositories that Centers are using although to variable extent and in variable applications. Launched in 2017, the Big Data Platform represents a laudable cross-CGIAR effort for OD and seamless data integration, analysis, visualization and mapping. Nevertheless, OA requires its own resources and central promotion and monitoring for the OA/DM Policy to be fully effective.

Recommendation 2: The System Management Office should promote greater interaction among data managers and intellectual property and legal experts through joint activities and capacity development events of the respective communities of practice so as to enhance best practices and workflows that address all aspects of OD management.

Recommendation 3: The System Organization should engage with donors to harmonize on requirements OA and OD and agree on common principles.

4. Implementation Support

4.1 Introduction

Efforts to support CGIAR-wide OA implementation were initiated with a concept note prepared by the Consortium Office in early 2014, with a financial request for USD 17 million. While both ISPC and FC considered support for implementation of the OA/DM Policy, both had reservations on the proposed support project in terms of prioritization, scope, and sequence of activities. A working group was established with members from FC, GFAR and ISPC to review and provide support and feedback on subsequent proposals. The members provided input to FC on two subsequent revisions proposing that focus should be on an initial phase of support with a limited budget.

In November 2014, the FC approved the Project proposal “CGIAR Open Access and Open Data Phase I (Years 1-2): Assessment, Prioritization, and Coordination of CGIAR’s Current Open Environment” for a funding of USD 2.38 million, a dramatically smaller budget for work of much smaller scope than originally proposed. The intention then was that a subsequent Phase II project would tackle longer term and wider OA efforts across CGIAR. The Phase I project focused on the following activities:

- Data prioritization framework
- Coordinated support to Centers and CRPs through a OA/DM support pack; supporting implementation plans; developing improved interlinkages and platforms, collaboration with partners)
- Plans for impact assessment

The review team assessed the project deliverables and achievements against what had been proposed with focus on three areas: the OA/OD Support Pack, support of the COPs and resource support.

4.2 OA/DM Support Pack

The Implementation Guidelines call for the (then) Consortium Office (since 2016 SMO) to provide OA/OD Support Pack. This resource was developed to provide access to useful and reusable resources and materials developed by Centers or external parties, and to include new resources in response to requests from Centers. The OA/OD Support Pack is thus designed as a “living” online open resource for CGIAR members to assist in the implementation of OA/OD. It includes resources, templates and examples from Centers concerning OA implementation, such as information policies and guidelines, sample processes/workflows, training materials, and staffing models.

The objective of the OA/DM Support Pack was to function as the main source for policy guidance – apart from the Implementation Guidelines. Developed with Google sites, the structure is that of a repository and there are two main sections:

- *Resources*. This includes templates for data implementation plans and budget planning, a limited number of Center policies, and example Terms of References for data managers. It also includes an incomplete list of data repositories in CGIAR Centers.
- *Meetings and Workshops*. This includes meeting records, minutes of the two COPs (OAWG and DMTF) and minutes of regional workshops.

The development and use of a central support repository was an important effort to provide necessary information and templates for implementing OA/OD in CGIAR. It served as a useful centralized space to share information and resources at the beginning of the implementation efforts, where none existed before. As noted in the Implementation Support proposal, the OA/DM Support Pack was based on the data management support pack developed by the CRP for Climate Change, Agriculture, and Food Security (CCAFS)³⁹. In terms of structure and usability, the CCAFS data management support pack offers users the possibility of filtering information depending on the type of user and use (example, data manager or researcher; and designing plans or curation), which can be very helpful in navigating the information.

Many of the data managers and OA focal points interviewed pointed to the usefulness of the Implementation Guidelines, but noted that the OA/DM Support Pack was only useful for those who knew where to access the information they were looking for. The review team considers that the current format and structure of the OA/DM Support Pack make it more like a repository than a guidance tool for implementation, which the CCAFS tool is. For the OA/DM Support Pack to continue to be useful to implementers and effective in the future, it should be upgraded by adding usability features (clearly differentiating among users and uses) and active content management. Content would need to be effectively developed, organized, updated, and curated. In addition to storing only meeting minutes, the “Meetings and Workshops” section should contain information also on main activities and work streams for serving users better.

While the OA/DM support pack was not foreseen initially as a communications tool for external stakeholders, it is currently the only public and centrally updated space for CGIAR OA/OD information on the Internet and its design and content should also reflect its communication and dissemination character. The OA/DM support pack, guidance and toolkits are also made available to external stakeholders, and are seen as “exemplars to guide others” as stated by one partner. A review of the material considering also external partners’ and users’ needs would be beneficial.

4.3 Communities of practice

As mentioned earlier, the Implementation Guidelines defined two COPs with direct involvement in the implementation and promotion of OA/OD efforts.

OAWG was established already in 2012 following the endorsement of the IA Principles, to assist with the drafting and development of the Open Access Policy for CGIAR. Upon endorsement of the OA/DM Policy, the working group focused on helping to create the enabling environment for OA implementation. It consists of Knowledge Managers from Centers and CRPs. The terms of reference for the OAWG include the drafting of OA/DM implementation plans to ensure compliance.⁴⁰

The DMTF was set up in 2014, to focus on data management issues of OA. It aims particularly at providing oversight and defining data standards and interoperability protocols to be implemented and

³⁹ <https://ccafs.cgiar.org/data-management-support-pack>

⁴⁰ <https://docs.google.com/viewer?a=v&pid=sites&srcid=Y2d4Y2hhbmdlLm9yZ3xvYXdndfGd4Ojc0MTJhN2RIYjJlMmNkM2I>

applied across CGIAR OA repositories⁴¹. DMTF is made up of data managers or equivalent from the Centers and CRPs. Both COPs have also representation from the SMO.

The overall purpose of the two COPs is to act as a strong harmonizing and guiding force for the implementation of the Policy. The communities have held annual meetings since 2015 (two each) to share information and updates in their respective work. To continue efforts between meetings, smaller working groups led by COP members were established within the COPs (for example working groups on Ontology, Dataverse, and Metrics) to tackle specific OA implementation issues of significance across Centers.

The review team found that establishing the two COPs has resulted in interaction among the CGIAR communities and enhanced common understanding about the benefits of OA. The COPs have been instrumental in promoting the exchange of best practices and state of the art, as well as the alignment of practice between different Centers. They have provided opportunities for cross-learning and sharing. Together they have built a community within CGIAR where information and data managers can exchange and consolidate their views on best practice and ways to address the main challenges for technical and other aspects of OA/DM Policy implementation and compliance.

The review team identified many areas of good progress, for example regarding metadata and the CG Core Schema. The approach of moving forward through smaller, specialized working groups that focus on particular issues of interest (for example on data prioritization, data quality, and semantics) can be effective as long as the results are communicated to the wider community.

Some capacity development has also occurred within the COPs, with webinars and trainings provided to COP members. As members of the COPs continue to develop and implement the OA/DM Plans, issues and topics for additional capacity development will continue to emerge. In the latest 2017 DMTF meeting, participants identified a series of topics for webinars to build capacity and common understanding; these, and any future training event, should be recorded in an appropriate format and shared through the OA/DM Support Pack.

Overlap on topics and membership between the two communities occurs, as can be expected. While separation between the two COPs has been understandable at the initial stages of implementation, the issues are mostly intermingled and it is important to address them jointly. This was apparent in the 2017 meetings where many working groups and topics overlapped. In the OAWG 2017 meeting after-event assessment completed by participants⁴², many pointed to the need for enhanced cooperation and for involving DMTF more and merging the two COPs when common subject are dealt with.

The team, however, thinks that it is better to enhance collaboration of the COPs, rather than merge them. Activities through specialized working groups are the most effective way to continue to build ownership and move forward with OA/OD efforts in CGIAR. The review team considers that the COPs could become more effective through the following:

⁴¹ The Terms of Reference for the DMTF were not found in the OA/DM support pack or online. Assessment focused on meeting records, interviews, and participation in an annual meeting by review team members.

⁴² <https://www.mentimeter.com/s/b500b743e975198cd7f2b134e5ff947a/1ba0e1f9e3b7>

- A decision-making process or panel should be in place for addressing roadblocks
- Information about the work done and progress in all working groups should be shared within the communities.
- Linkages and joint sessions between the OAWG and DMTF COPs is needed, also through their working groups.

The review team concludes that the COPs (both DMTG and OAWG) and their working groups are worthwhile efforts in need of continued support. The small working groups in particular can allow for complex issues to be tackled at the Center level – which is where the data and information management resides. The annual COP meetings provide opportunities for cross-learning and sharing of knowledge and information.

4.4 Resources

According to the Implementation Guidelines, *“Centers should begin to budget for Open Access and data management costs in future funding proposals, including the forthcoming second call for CRPs.”* During Phase I implementation support funding was not dispersed directly to Centers to assist them with their OA/OD efforts. Instead, the project focused on building and supporting System and cross-Center efforts to address common challenges, such as the OA/DM support pack, COPs, and working groups.

Center staff interviewed voiced their concern that sufficient funds had not been allocated for the implementation of OA/DM at the Center level, although many had expected funds to be made directly available. Some Data and OA managers were concerned about transparency in the use of the Phase I funds. While the 2015 Phase I project financial report indicating that 35 percent of the USD 1.1 million expenditures in that year were spent through Centers, the team learned that funds had not been given directly to Centers, but only for consultancies to support implementation at Centers. Future decision-making on System-level resources, including those for the “Organize” module of the Big Data Platform, should be made more transparent and, when appropriate, inclusive of representatives from the two communities.

According to the OA/OD survey, only one Center indicated having a centralized OA fund to help pay publication fees. Another noted that OA publication costs are split between the Center’s Knowledge Management Unit and research divisions. Two Centers indicated in the survey – as others have indicated in their OA/OD implementation plans – that researchers are being encouraged to incorporate budgets for OA in new project proposals. Efforts to budget for OA should be encouraged. Without prescribing a fixed amount, the team encourages Centers and CRPs to share experiences about costs. Currently, only a few Centers indicated that the OA/OD focal point(s) were being consulted during project planning to ensure resourcing for OA implementation.

One Center estimated that total resource requirement (budget and staff) to implement OA, concluding that it adds up to roughly USD 3000 by publication. This covers OA publication fees, data curation (data formatting, codebook preparation, de-identification of personal information, etc.) and data publishing costs (posting on portals such as Harvard’s Dataverse). The staff from different Centers confirmed that costs for OA and OD are quite distinct and need to be planned and budgeted for separately from other budgeting. Donors, whom the review team contacted, were well aware of the

cost implications and budgeting requirements of OA/OD. Some donors are already including OA costs in their bilateral project budgets, which should be considered a best practice. However, OD requirements (pose continuous, long-term needs for data storage and curation. It is important to note, as was emphasized by one research manager that these kinds of costs accrue well beyond the lifetime of the project that produces these data, which needs to be taken into account at budgeting.

4.5 Conclusions

In the first years of the transition period, CGIAR SMO (previously Consortium Office) operated a Phase I project for assessment, prioritization and coordination of OA. The team considers that the COPs developed within the project have been very successful. The project also resulted in development of an OA/OD Support Pack, which has served as a useful centralized space for sharing information and resources. It is also serving a communication and dissemination purpose to external stakeholders, although that was not initially envisioned. However, the Support Pack should be further developed to a truly useful resource and guidance tool for policy implementation, which requires restructuring and regular curation. The Phase I project was also expected to support resourcing of OA activities in Centers but it was not clear what this role has been. It was difficult for the team to get information on project budget and expenditures and the way the project provided financial resourcing to Centers. For accountability and transparency, more prudent reporting would be needed.

Recommendation 4: The System Management Office should oversee restructuring of the OA/OD Support Pack to enhance its usefulness as the main source of guidance to different kinds of users within CGIAR, and to make it more suitable as a communication and dissemination channel and tool.

Recommendation 5: For resourcing OA implementation, action is needed at Center and System levels, and should include the following:

- Centers should make budget planning for OA as part of normal project formulation for all research and consider earmarking institutional financial resources for peer-reviewed publication which generate costs long after project termination.
- SMB should consider approaches to OA financing, considering that Centers have limited resources beyond project lifetime for guaranteeing OA.
- Central funding, necessary for supporting Policy implementation and facilitating the communities of practice and capacity development, should to be planned in an inclusive manner and transparently reported.

5. Governance and management

5.1 Introduction

As the full implementation of the OA/DM Policy is not expected until the end of 2018, The System-level governance, management and reporting system has not yet been put in place. In considering aspects of governance and management, the Team reviewed the following materials:

- FC/System Council records in relation to Policy development and discussions related to OA and OD
- Implementation Guidelines with reference to reporting, governance or management roles and responsibilities
- Governance and reporting concerning the IA Principles that present a comparable System-wide policy
- Feedback from interviewees (senior management, data managers, and OA focal points)

The role of the System-level bodies is also to promote awareness through rewards, *“establishing policies and procedures for scientists that encourage, support, and reward deposits of research outputs and data in appropriate repositories”* and also offer *“reporting metrics via agreed-upon channels”*.

The System Organization is assigned a coordinating role in *“sharing of tools, seed funding, resources, and advocacy needed by the Centers and acting as a facilitator, champion and advocate for Open Access and Open Data internally and externally”*.

5.2 System Council role

The development and endorsement of the OA/DM Policy emerged soon after the FC endorsement of the IA Principles. FC meeting records show widespread support from members for OA. Discussions around the Policy opened the way for the development and endorsement not only of the OA/DM Policy, but also the scope of the implementation support and its funding over the past 5 years.

The FC meeting records capture the main issues discussed by donors; particularly the importance of OA and linkages with agencies outside CGIAR, and the need for central coordination of OA/OD. Members have also stated the need to contain costs and ensure appropriate budgeting.

As with any CGIAR System-wide policy, the System Council have a central role in endorsing any changes to the formal policy documents (Policy itself and the Implementation Guidelines). Approval of any central core funding for OA/DM is also the door council’s role. Furthermore, funders in their deliberations have identified the key issues that still remain current, such as data quality and prioritization challenges, end use benefits and resource requirements. This indicates that OA issues are considered important in the highest level of governance. With the establishment of the System Organization in 2016, the System-level governance and central management of the OD/DM Policy implementation will require a division of roles between the SMB and the System Council.

5.3 Roles and responsibilities

Regarding roles and responsibilities as defined in the Policy and Implementation Guidelines, CGIAR Centers have primary responsibility and accountability for implementation of the Policy. This includes

establishing Center specific policies (in 2015 ten Centers had a policy for OD/data management) and procedures, managing and curating repositories, and monitoring as well as reporting. Centers are also responsible for ensuring that budgets reflect OA needs, and that all Center agreements and contracts comply with the Policy, thereby making also partners compliant with it.

As per the Implementation Guidelines the System Organization would be responsible for central support for Policy implementation, developing and compiling aggregate data and metrics for Policy compliance across CGIAR, reporting to the System Council, and central communication about OA. Given that CGIAR went through a recent governance transition in 2016, the roles and responsibilities as stated in the Guidelines are no longer clear. The role that the SMO would need to play as a central unit for communications, coordination, and reporting is not yet well defined. For example, the review team observed that Web information about OA in CGIAR is currently limited and considers that there is scope to improve this function.

According to the Implementation Guidelines, the System Organization is expected to act *“as a champion and advocate for Open Access and Open Data internally and externally as CGIAR strives to become a leader in open knowledge for agriculture research”*. Senior management and donor representatives felt that the System Organization is currently not fulfilling this role in a decisive way. The review team emphasizes that a stronger engagement of the System Organization would be necessary, especially in facilitating the community of practice.

Another recent change is that the Big Data Platform, in its Module 1, has been given the mandate to develop tools for harvesting, analysis and visualization of OA information products. The Big Data Platform management and governance structure includes a Steering Committee with representatives from Centers and CRPs, as well as partners. The Steering Committee would be well placed to provide advice on next stages of implementation support, including budget allocation.

5.4 Center-level governance and management

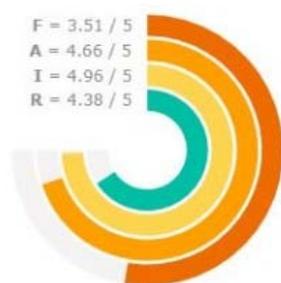
For Center-level implementation, the policy guidelines place complete governance and implementation onus on the Centers. The review team found a good level of buy-in and commitment to implementing the policy among Center senior management. From its observations of progress to date, the team concludes that full embracement of OA culture and practice requires further steps from Center leadership. This would entail defining and advertising strong incentives both at a Center and CRP level and individual data manager level to improve performance in both OA and OD. Incentives could include competitions and awards, of which some Centers have already gained experience. Policy implementation needs champions amongst senior management as well as researchers, to convince research communities in the Centers that OA adds value to Centers' knowledge assets and is necessary for fully realizing CGIAR objectives and mandate. The review team observed that OA leaders are already emerging among managers at Centers and CRPs.

5.5 Reporting

Reporting on OA/OD activities from Centers/CRPs to the System is not yet in place, and will be expected in early 2019. At the initial stages, reporting is expected to be completed by tracking progress in accordance with the Center implementation plans, and that may be sufficient for the initial years. There is a need of clear and harmonized reporting approach for the Policy, within and across Centers.

The policy implementation is at a stage where clear metrics and indicators need to be defined for the progress of the Policy.

Figure 4: Example summary image of FAIR reporting



A set of indicators has been developed by a COP working group with the aim to standardize and measure levels of compliance to the FAIR principles for both publications and data across CGIAR (see examples in Table 1). Standardized tangible measurable elements for each FAIR principle have been defined as the indicators. The indicators and guidelines for reporting against them include detailed definitions of what each Principle translates to in terms of outputs. These would be self-reported by the entity with a score of “1” for yes and 0.5 for partial compliance for each criteria, and results across CGIAR would be visualized (see Figure 4). The team considers that the indicators are very suitable.

MARLO⁴³, currently adopted by eight CRPs, and other similar systems⁴⁴ used by some CRPs/Centers Currently, allow reporting on OA and is “FAIRness” elements. In 2016, MARLO was already being used for planning and reporting on OA against the draft indicators. By using the MARLO, or other reporting systems, Centers and CRPs would be able to automatically generate annual reports summarizing all compliance data and progress. .

Table 1: Example of indicator reporting guidelines

Component Indicator	What it means in general terms...	What is required of the repository	Score
FAIR Principle F2 (Findable): (Meta)data are assigned a globally unique and persistent identifier.	Each resource (dataset or publication) is assigned a globally unique and persistent identifier (PID; e.g. DOI/URI) to enable the finding, citing, and tracking of (meta)data.	Repository has a predictable way to assign a PID to each component of a dataset (e.g. each file or nano-publication), to be able to include these identifiers into the corresponding metadata before submission.	Yes 1 point Partial 0.5 point
FAIR Principle A1 (Accessibility) : Resource is accessible for download or manipulation by humans and is ideally also machine readable.	Resource can be accessed: <ul style="list-style-type: none"> • free of cost to the user • via an open protocol (subject to an authentication/authorization procedure if necessary). • freely downloaded and/or manipulated by humans, • ideally also machine readable via standard protocols (i.e. access and manipulation of the underlying data is automated). 	Resource is archived in a repository that is accessible by humans -- and ideally also machines, via a standardized protocol.	Yes 1 point Partial 0.5 point

The set of indicators, and the online reporting tool, allows are harmonized understanding and reporting for CGIAR, while also promoting transparency. Translating FAIR principles into indicators for planning and reporting is a clear tangible result of the Policy and its implementation efforts, and the review tea, is encouraged by the progress made.

⁴³ MARLO reporting tool is adopted by CCAFS, PIM, WLE, A4NH, FTA, Livestock, Wheat and Maize CRPs, and by Big Data and Breeding Platforms.

⁴⁴ The Monitoring, Evaluation and Learning (MEL) System is used by RTB, ICARDA, CIP and IITA.

To increase efficiency in reporting, OA/DM reporting could be incorporated into the IA reporting, to avoid additional reporting burden at Centers. This would enhance collaboration between the different communities and networks involved in the two System-wide policies.

5.6 Conclusions

The Implementation Guidelines define clearly the roles and responsibilities of the Centers that have primary responsibility and accountability for implementation of the Policy. However, given the recent governance transition, and establishment of the Big Data Platform, the roles and responsibilities at the System-level need to be clarified (see Recommendation1). The central reporting and oversight have not yet been put in place given that CGIAR is still in transition towards full implementation of the Policy.

Tracking compliance with the Policy and reporting will be important for governance of the Policy. There has been progress in developing indicators for reporting on OA, both on publications and on the FAIR principles as they apply to OD. Centers are sharing experiences on workflows in the joint fora. The review team concludes that efficiency gains could be made by integrating reporting on OA and on IA given that for the latter there is a well-established mechanism and that the policies are complementary.

Apart from setting up a monitoring and reporting mechanisms for OA/OD, the System Organization is expected to champion advocate OA for making CGIAR a leader in open knowledge in its field. The System Organization has not yet been very active and decisive for fulfilling this latter role.

Recommendation 6: The System Organization should assume an active role in championing OA across CGIAR through the following ways:

- taking on a coordinating role for promotion and enhancement of implementation the OA/DM Policy across CGIAR;
- inclusion of OA issues regularly in SMB meetings;
- developing reward mechanisms for exemplary OA/OD management;
- supporting System-wide sharing of OA/OD best practices and promotion;
- improving online visibility and information sharing of CGIAR efforts on OA..

Annex: Team Bios



Johannes Keizer has more than 20 years of experience in working with science data. After his PhD in Biology he founded in 1994 “Knowledge Brokers” and contributed to some important Italian Science data projects (“Registro Nazionale Gaucher”, “Archivio Stanislao Cannizzaro”, “Banca Dati sui pesticidi in Italia”). While serving at FAO, he created the Pesticide registry for the Rotterdam convention, and was responsible for AGRIS (“International Information System on Agriculture Science and Technology”) and AGROVOC and brought these traditional documentation tools into the Semantic Web World. He currently works with the GODAN (“Global Open Data in Agriculture and Nutrition”) Secretariat and is an advisor to the H2020 funded project “eROSA”. He is passionate about open science and “going FAIR”. He has published more than 100 papers in scientific journals.



Devika Madalli is a Professor of the Documentation Research and Training Centre, Indian Statistical Institute, India and Adjunct faculty, DISI, University of Trento, Italy. She is a founding member of and served as the co-chair of the Agriculture Data Interest Group [IGAD], Research Data Alliance. Her interest lies in the area of Open Data Repositories and Data Management, Knowledge Organization, Information Systems and Infrastructures, Digital Libraries, Ontologies, Content Management System, multilingual information services and e-learning. She served as a member of a high level evaluation committee of UNFAO’s statistical database and information services, FAOSTAT. She is a member of the Karnataka Evaluation Authority. She contributed to UNESCO’s Global Open Access Portal (GOAP). She is also the Chair of the OpenAccessIndia forum. She is on the advisory board of Universal Decimal Classification. She served as a member of the G8+06 Data infrastructures working group hosted by European Commission. Currently Dr.Madalli is a technical advisory board member of the Research Data Alliance.