FAIR and Open Data Alignment Across Excellence in Agronomy Initiative Related Organizations

Program Officer	Christian Witt
Funders	Bill & Melinda Gates Foundation
Dates	2023 February – 2023 July

Executive summary

Key insights

The Bill & Melinda Gates Foundation believes that "Providing access to underlying data is key in fulfilling the foundation's mission of rapid and free exchange of scientific ideas to move humanity forward by improving and saving lives. Without barriers the scientific community can freely benefit from data and build upon each other's work."[1]

If data is not managed well from the collection phase and throughout the full data life cycle, its value to other users decreases. FAIR compliance is a significant objective for the Excellence in Agronomy (EiA) program, and it is data management that underpins this. Effective and well-constructed data management processes, together with suitable data management infrastructure, are at the core of addressing the issues around developing large agricultural databases, which is necessary in order to advance agricultural science and policy, which is a key objective of EiA. The program has made admirable efforts in planning for data management and data sharing. The initiative, OneCGIAR and the foundation are aligned on a policy level with regard to FAIR data principles, but alignment and implementation of the data management principles that underpin FAIR weaken as one moves away from a core group of donors and grantee partners. Divergence in support of FAIR and open data at national system or center level can represent multiple points of potential failure in data flows, which need to be addressed with the foundation's support.

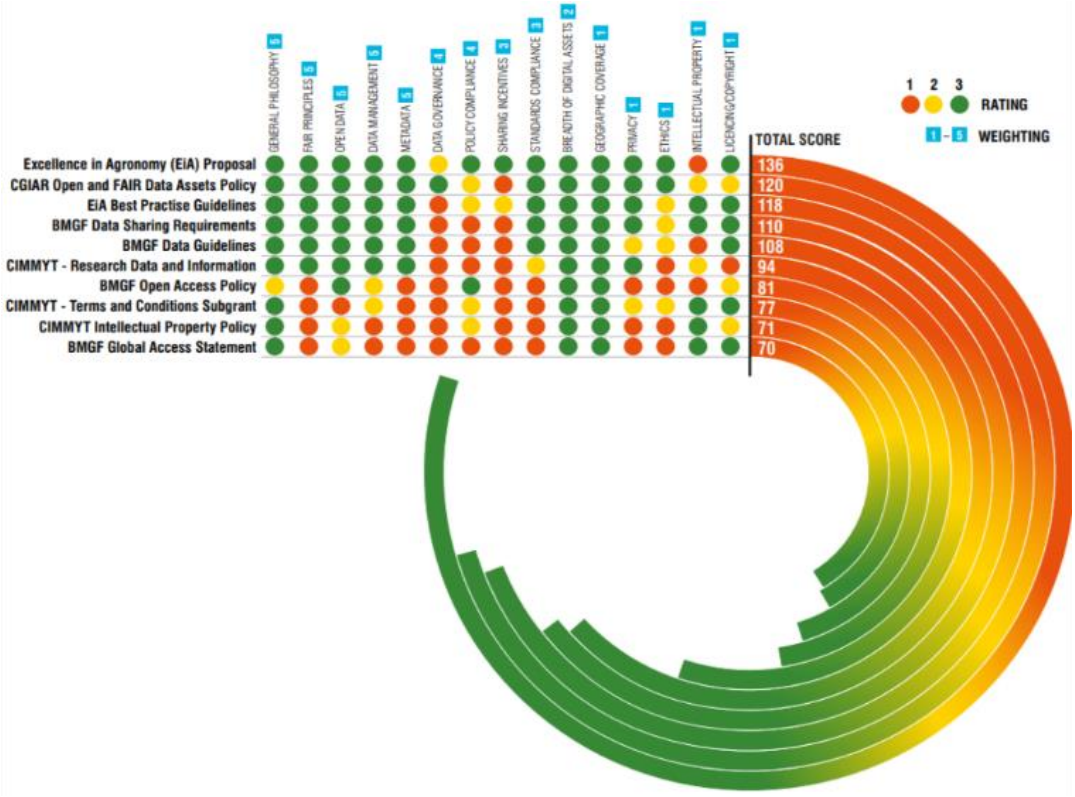
CABI reviewed several policies, guidelines and agreement documents from organizations involved in EiA, belonging to the foundation, OneCGIAR, CIMMYT and EiA. Firstly, CABI assessed the level of alignment within individual documents with 15 criteria consisting of elements underpinning best practice data management principles, weighted according to the influence each criterion has on meeting the FAIR and open data objectives. Secondly, the level of alignment of the promotion of FAIR and open data between the core organizations involved in EiA was assessed. A high-level review of data sharing and data management documents of other EiA donors (the United States Agency for International Development (USAID), the UK

[1] The Bill & Melinda Gates Foundation Open Access Policy: https://openaccess.gatesfoundation.org/how-to-comply/data-sharing-requirements/

Foreign, Commonwealth and Development Office (FCDO), the Norwegian Agency for Development Cooperation (NORAD), Indian Council of Agricultural Research (ICAR) and Irish Aid) was also conducted.

Findings

Good data management and data sharing are a prominent component of EiA. The program incorporates a well-defined work package ("TRANSFORM") that is focused on enabling efficiencies in the programme's data lifecycle, with specific considerations on how to improve data management and data sharing. These efficiencies include clear implementation guidelines, allocation of staff, ownership of data agendas (strategy and implementation) through a Chief Data Officer, and allocation of an appropriate budget to support data management which demonstrates a wide range of good practices. EiA represents a potential exemplar for how other projects might plan for good data management and data sharing, in the CGIAR and beyond. The data sharing mandates of some other EiA donors, notably NORAD and Irish Aid, were found to be weak.



To support EiA and to do FAIR well, in order to see the benefits of data sharing, the foundation should do the following:

1. Include clear language and clauses in grant templates on making data FAIR by default, with non-compliance leading to a percentage of funds being withheld.
2. Stipulate that grantees should not publish a research paper unless datasets are included with the paper or deposited in an acceptable repository, as per the foundation's Open Access Policy.
3. Exclude researchers from obtaining another grant if they do not comply with FAIR data standards and (where possible/ethical) make data open.
4. Initiate an open dialogue with the foundation's grantees on how research centres can better reward researchers for complying with good FAIR and open practices, including how researchers' key performance indicators can better reflect the foundation's data sharing mandate.
5. Engage with the donor community and CGIAR centres on the value of FAIR and responsibly shared open data, seeking to make the foundation a leading voice in this space.
6. Invest efforts in developing communities of practice, facilitating spaces and platforms for knowledge exchange.
7. Co-create data governance frameworks with investment partners to do the following:
8. Align terminology used in the various policy and guidelines documents.
9. Cost in and develop a data management plan in the project planning phase to FAIR-ify data from data collection.
10. Create a data governance policy that addresses ownership and responsibilities.
11. Agree a data sharing policy and data sharing agreements.
12. Promote tools that support the above to ensure compliance with existing data standards.

EiA's Chief Data Officer Medha Devare explained: "*We need more 'stick' to encourage real change of habit because it's been too long that we've just been focussing on 'carrots', including raising awareness of good practice. We have the tools and enablers there, we have the policy, we have exemplars, a data sharing agreement, but what's lacking is the teeth in the policy – and the teeth come from*

our funders."[2] This is where contractual stipulations will really make a difference.

Why FAIR and open data?

As articulated by the CGIAR, "*Open and FAIR data assets improve the speed, efficiency, and efficacy of research; they facilitate interdisciplinary research; assist data aggregation, computation, and the derivation of new insights; and allow the global public to benefit from CGIAR Research. They enable CGIAR to collectively leverage the infrastructure, data pools, and new data science capacities necessary for innovation and for effective and agile responses to global challenges. They facilitate text and data mining and analysis to derive insights, recognizing that these are dependent not only on access to high quality data, but also on that data being well-contextualized (through rich metadata and relevant open materials), interoperable, and reusable*."[3]

Harnessing the value of data, including mainstreaming FAIR practices, is an urgent need for organizations. The opportunity cost of not having FAIR data in research is estimated to be €10.2 billion each year in the European Union alone, with the value of open data estimated at €184 billion in that region. Agriculture is seen to be a "high-potential" sector for open data. Unfortunately, wrangling of poor-quality data "*can take up to 80% of the total effort*", leaving only 20% of the effort for analysis. Collaboration can be hampered if data is not FAIR; understanding wicked problems such as the COVID pandemic or climate change demands a broad range of data in different sectors to be used and mutually intelligible (interoperable).[4] [5]

Collaborations such as the Coalition of the willing (CoW) in Ethiopia continue to work with research institutions to institutionalize activities that will sustain CGIAR, EiA and wider FAIR infrastructure. A CoW member noted: "*Donors could help remove technological and human barriers…efforts should be*

[2] International C. Understanding the enablers and disablers of mainstreaming FAIR – the Case of CGIAR [version 1; not peer reviewed]. *Gates Open Res* 2024, **8**:25 (document) (https://doi.org/10.21955/gatesopenres.1117084.1)

[3] CGIAR Open and FAIR Data Assets Policy: https://cgspace.cgiar.org/bitstream/handle/10568/113623/CGIAR_OFDA_Policy_Approved_16April2021.pdf?sequence=1&isAllowed=y

[4] EC, D-G for Research and Innovation (2019) "Cost-benefit analysis for FAIR research data: cost of not having FAIR research data".

[5] Wellcome Trust Open Access Policy: https://wellcome.org/grant-funding/guidance/open-access-guidance/open-access-policy

connected, not based on the interest of each institution of projects and centers". The foundation has the opportunity to support data communities of practice, and through international collaborations, engagement methods and education for national partners, more funders and grantees can support greater adoption of FAIR principles, leading to greater data sharing.

Table of contents

1 Introduction

This document is one component of CABI's support that aims to facilitate conversations around the alignment of understanding and planning on how the FAIR [6]and open data[7] concepts can be made a priority in the Excellence in Agronomy (EiA) programme.

The desk study reported on here involved reviewing in detail 10 relevant data policies, guidelines and other related documents from the Bill & Melinda Gates Foundation ("the foundation"), OneCGIAR, CIMMYT (as a representative of CGIAR centres), together with documents belonging to EiA. In addition, a high-level review of the data sharing policies of the other donors of EiA (the UK Foreign, Commonwealth and Development Office (FCDO), the United States Agency for International Development (USAID), the Norwegian Agency for Development Cooperation (NORAD), Irish Aid and the Indian Council of Agricultural Research (ICAR)) was conducted.

The primary focus of this study is to find areas of alignment and non-alignment around FAIR and open data concepts in institutional and programmatic policy and strategy documents, to draw an overall picture on whether the different institutions seek to promote in their policies (and if they plan for) good data management and data sharing for their investments.[8]

In this context:

[6] The **FAIR principles** focus on having data that is well-documented, well-structured, and well-managed. This enables the data to be **F**indable**, A**ccessible**, I**nteroperable**,** and **R**eusable (FAIR). Data complying with the FAIR principles can be easily reused, although FAIR itself does not focus on sharing data, simply on making it suitable for sharing and reuse. See the GO FAIR website for further details: https://www.go-fair.org/fair-principles/

[7] **Open data** focusses on the sharing of data under an open licence, with digital assets being available for anybody to use without restriction. Open data has "limited" emphasis on ensuring the data can easily be reused (i.e. FAIR-compliant). In other words, open data compliance *could* be as simple as having a discoverable PDF document available on the internet – a potential re-user of data in this form may need to transcribe the PDF document into a spreadsheet. Having data effectively managed and documented, while relevant, is not seen as the primary focus of open data concepts (World Bank. Open Data Toolkit: https://opendatatoolkit.worldbank.org/en/data/opendatatoolkit/home).

[8] The value and utility of open data increases significantly when this data is FAIR-compliant, so the ideal approach when seeking to achieve improved agricultural research outcomes and to better leverage research digital assets is to comply with both of these principles. This would result in research data being effectively managed and documented (to support findability, accessibility, interoperability, and reusability – FAIR) and available without restriction to any and all potential users (open). There are many real-world examples of where the application of good data management (FAIR) and having data open has provided considerable scientific and economic benefit. Refer to **Annex 1** for examples of these benefits.

- **Alignment** refers to where all parties are in close agreement with the "what, how and why" of FAIR and open data within their policies, based on the 15 criteria for best practice data management and data sharing, as well as overall alignment between the different organizations involved: in short, are they promoting and prioritizing the same thing when it comes to FAIR and open data? (Outlined in Section 2.2).
- **Non-alignment** refers to identified gaps where one or more of the policy frameworks specifies something relating to FAIR and open data that the others do not, or where two or more of the key documents state differing viewpoints on the same (data-related) topic. Non-alignment also refers to where a specific "element" needed to effectively support FAIR and open data principles is missing or is not strongly emphasized.

The identified areas of non-alignment are used to highlight potential risks to, and issues for, EiA activities, for the attention of the Senior Project Officer (as a representative of the foundation) and of CGIAR (as the grantee). The partner landscape for EiA is complex and diverse, as seen in this stakeholder map on Kumu. Therefore, the identification of areas of non-alignment identifies areas where strategic and co-created solutions are needed. In this regard, CABI provides recommendations to strengthen support for the EiA data-related objectives, in Section 5.

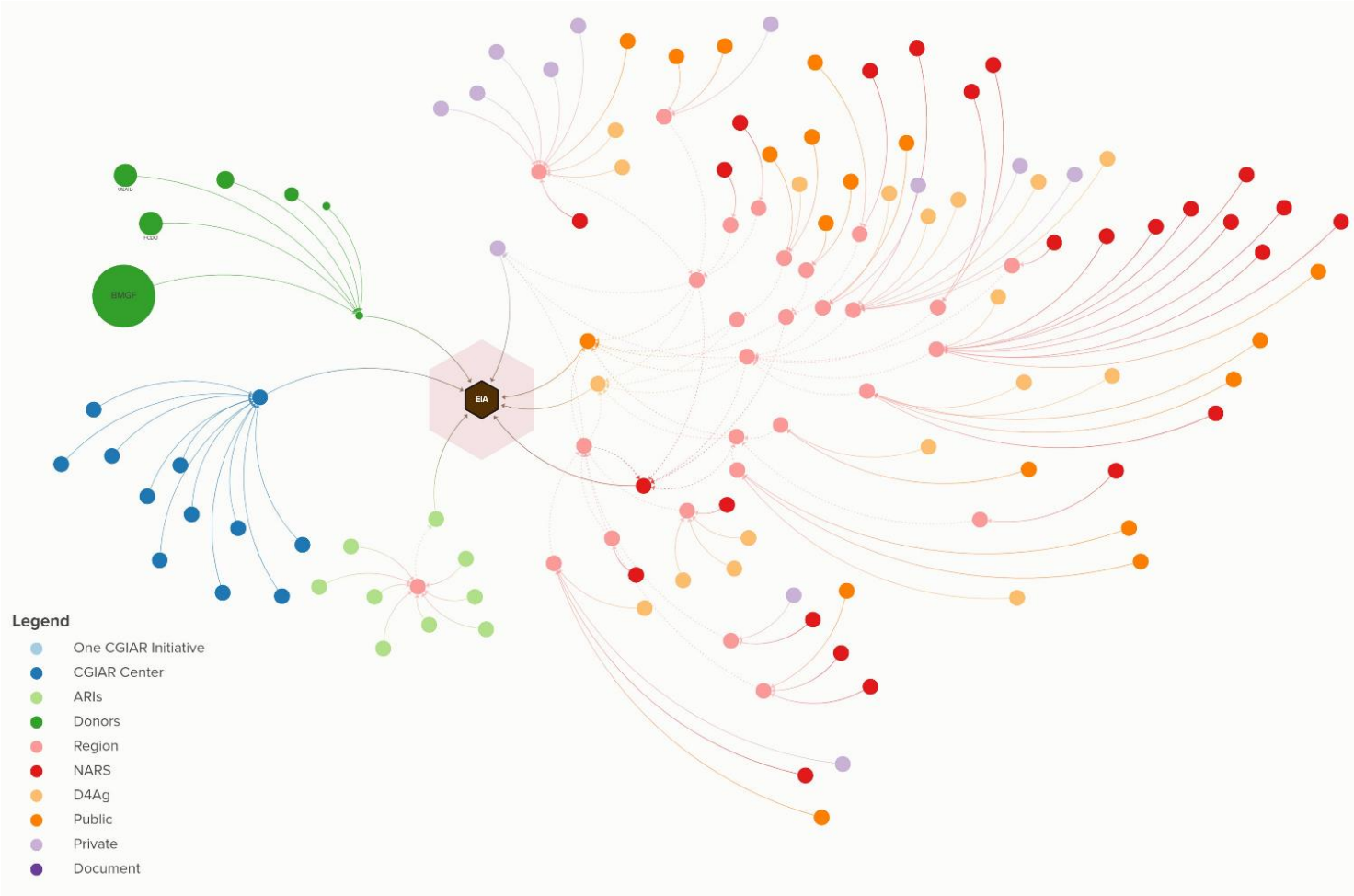


Figure 1.0 EiA partner landscape

Although complex, Figure 1.0 reflects the significant ability and potential resources that EiA has to achieve its aim of delivering an increase in productivity and quality per unit of input (agronomic gain) for millions of smallholder farming households in prioritized farming systems by 2030.

To do this, several of the EiA activities and outcomes focus on the application of "at scale" solutions. For example, some of the EiA data and information-related objectives include the following:

- **Facilitating the delivery of agronomy-at-scale solutions,** including development and technical/user experience validation and the co-creation and deployment of gender- and youth-responsive solutions to smallholder farmers via scaling partners.
- **Enabling the creation of value from big data and advanced analytics** through the assembly and governance of data and tools; the application of existing analytics and solutions for specific use cases; the supply of information on the climate impacts, inclusivity, and sustainability of agronomic solutions; and the strengthening of national agricultural research system capacity.
- **Driving the next generation of agronomy-at-scale innovations** by addressing key knowledge gaps and facilitating innovation in agronomy research through engagement with partners.
- **Nurturing internal efficiencies to enable an agile and demand-driven agronomy research and development community**, through internal organization and external partnerships for prioritization, demand mapping and foresight.

A specific data-related outcome of EiA is:

Outcome 3

1. At least 75% of research and scaling partners use and share common, **open and FAIR (findable, accessible, interoperable, and reusable) data**, tools and analytics to support the co-creation of locally relevant agronomic solutions **integrating** climate-smart, inclusivity and sustainability dimensions, and assess their performance using **standardized** protocols.

The ability to create agronomy at scale and use big (aggregated) data to drive the next generation of agronomy-at-scale innovations depends on the ability of EiA to effectively bring together the digital asset outputs from the

multitude of CGIAR-supported research projects. The aggregation of the digital assets from each individual research programme, and their integration with data from other data collection activities, such as ongoing monitoring to support advanced analytics, will only be achieved effectively if several critical elements are in place. One of the key elements is that data management, FAIR and open data policies of the donor organizations, OneCGIAR and the CGIAR research centres, should be compatible and consistent – this is what this review seeks to assess. While this review has not incorporated reviewing the policies of EiA country partners, we recognize that they are a key component to being able to unlock the benefits of FAIR and open data, and we recommend including such stakeholders in the conversation on how to align and implement FAIR data principles, and sharing methods and tools openly as part of a collaborative effort to achieve a global data pool.

An element of the review that is considered important is that the concept of good data management and making data FAIR should be a primary focus. While the sharing of data is essential to build "big data" repositories, without effectively managed data, the data will be difficult to integrate and of limited value. Additionally, there are more barriers to data sharing, such as national legislation and policies that are complex and will potentially take time to resolve. However, if the data is not effectively managed, then being able to overcome the sharing constraints will provide minimal benefit to meeting EiA's objectives.

Background

The CABI EDA3 team met with key stakeholders in January 2023 to begin the process. Christian Witt (the Project Officer at the foundation), Bernard Vanlauwe (Principal Investigator, EiA) and Medha Devare (Chief Data Officer, EiA) met to discuss CABI's tentative Action Plan to support the Project Officer in efforts to make FAIR and open data a priority in the EiA initiative. At this meeting, the following potential areas of need for additional foundation support were identified by EiA management:

1) A need to move to better-worded legal agreements from the foundation and CGIAR to ensure data compliance by EiA stakeholders (subgrantees, etc.).

2) A need for the foundation (and especially its public face and leadership) to require FAIR and (except for the usual caveats) open data sharing, both in OneCGIAR and centre grants. All data can benefit from being FAIR, and most should be open.

3) A need for the foundation to drive improved and joined-up messaging across the donor community on FAIR data principles, as a starting point for the conversation on open data sharing, especially those donors that support EiA (and OneCGIAR more generally), providing concrete recommendations from the desk review.

4) A need for support for meaningful discussion on the incentives for, and consequences of, data sharing; resourcing to help stakeholders commit to, co-develop and adopt incentives and tools that allow them to meet data standards; and enforced compliance of open and FAIR to support EiA and wider agricultural research and innovation objectives.

2 Methodology

This review has examined several policies, guidelines and agreement documents from organizations involved in EiA (the foundation, OneCGIAR, CIMMYT, EiA) to specifically assess two things:

1) The level of alignment within individual documents with the 15 criteria (the findings on this are outlined in Section 3.1)

Then,

2) The level of alignment of the promotion of FAIR and open data *between* the core organizations involved in EiA (the findings on this are outlined in Section 3.2).

By following this approach, we can see what the different institutions seek to promote in their policies, and whether they plan for good operationalization of FAIR and sharing for their investments.

The methodology used in this desk study analysis is based on the Florida International University's comparative analysis technique. This basic technique is part of Florida International University's Libraries Introduction to Policy Development and Implementation programme.[9] This approach is used to determine whether the relevant policies, guidelines and other related documents from the foundation, OneCGIAR, CIMMYT and EiA align, or do not align, with FAIR and open data concepts.

To provide an empirical outcome from this review, a scoring system based on specific criteria has been created to measure different aspects of each document against other documents. Details of this scoring are provided in Section 2.4 below.

The methodology applied consisted of the following elements:

- Locating and reviewing relevant documents. A list of the reviewed documents is provided below (Section 2.1).

[9] Hammill, Sarah J. (2022) "FIU Libraries: PAD3034 / Introduction to Policy Development & Implementation (GL): Comparative Analysis". Accessed 11 September 2023. https://library.fiu.edu/PAD3034/comparative.

- Defining the analysis criteria and patterns that are relevant in the documents (Section 2.2).
- Providing a weighting value against the analysis criteria relevant to the "impact" each criterion has on meeting FAIR and/or open data principles (Section 2.4).
- Analysing the documents for these patterns and criteria and providing a subjective "score" against each analysis criteria (Section 2.5).

A high-level review of data sharing mandates on the other EiA donor websites (Section 2.6).

A range of factors have influenced this review, including the complexities resulting from the different perspectives and roles of the various organizations. Furthermore, a wide range of additional factors influence these policies. These factors are discussed in more detail in **Annex 2.**

2.1 Core donor organization and programme policy documents reviewed

The following documents have been reviewed in detail against the criteria described in Section 3.

- Excellence in Agronomy (EiA) Proposal.
- EiA Best Practice Guidelines to Support Global Access Implementation.
- Bill and Melinda Gates Foundation Data Sharing Requirements.
- Bill and Melinda Gates Foundation Data Guidelines.
- Bill and Melinda Gates Foundation Global Access Statement.
- Bill and Melinda Gates Foundation Open Access Policy.
- CGIAR Open and FAIR Data Assets Policy.
- CIMMYT Research Data and Information Products Management Policy.
- CIMMYT – Terms and Conditions Subgrant Agreement.
- CIMMYT Intellectual Property Policy.

A detailed analysis of these 10 documents can be found in **Annex 4.**

2.2 Developing the analysis criteria

The analysis criteria have been developed using several approaches. These include a quick review of data management policy structures, including the following:

- Inter-university Consortium for Political and Social Research.[10]
- The UK Digital Curation Centre.[11]
- University of Queensland Research Data Management Policy.[12]

Additionally, the policy documents that were reviewed contain several key elements that contributed to the development of the analysis criteria. Finally, the consultant undertaking this analysis has applied over 40 years of data management experience in the oceanographic, geospatial and statistical data management fields in order to develop the 15 criteria and to determine the weighting given to each criterion.

The criteria applied to the analysis of the reviewed documents are described in detail below.

2.3 Rating criteria

A range of criteria were developed and used in the analysis process. These criteria influence the ability to effectively apply FAIR and open data principles. The criteria cover several of the main elements that are considered to underpin effective data management (ensuring data can be easily and effectively reused) and data sharing activities. Additionally, each criterion has a different level of impact on, or significance to, the effectiveness of data management and the FAIR and open data principles.

Each document was reviewed and evaluated against the following 15 criteria, which are listed in descending order of their significance and impact on ensuring effective data management and data sharing, according to the "weightings" described in Section 2.4 below. The words in bold below are the key words for each analysis element.

[10] ICPSR. "Elements of a Data Management Plan". Accessed 11 September 2023. https://www.icpsr.umich.edu/web/pages/datamanagement/dmp/elements.html.
[11] DCC. "Data Management Plans". Accessed 24 November 2023. https://www.dcc.ac.uk/resources/data-management-plans.
[12] The University of Queensland, Australia. "4.20.06 Research Data Management", Policies and Procedures Library. Accessed 24 November 2023. https://ppl.app.uq.edu.au/content/4.20.06-research-data-management.

- **General philosophy:** Does the document provide a strong message on what it aims to achieve through applying FAIR and open data principles?
- **FAIR core principles:** How much detail is provided about the application of FAIR data principles in the research activity?
- **Open data** concepts: Are open data concepts well described in, and are they required by, the document?
- **Data management** practices: Are best practice data management processes, expectations or conditions referred to in the document?

Metadata: Is metadata referred to and does the document specify the creation of appropriate metadata to effectively describe the digital asset?

Data governance frameworks: Does the document cover data governance concepts that would assist in supporting good organization-level data management practices?

- Requirements to ensure **policy compliance:** Are any policy compliance requirements mentioned in the document?
- **Incentives** to encourage the sharing of digital assets: Does the document provide incentives to encourage researchers to share their data?

Compliance with relevant data management, metadata, data formats, exchange, and any other relevant standards: Are relevant standards mentioned or required?

- **Breadth of digital assets**: What digital assets does the document cover? Just data or are other types of research-developed digital material included?
- **Geographic coverage:** Is the document relevant to global activities or is it limited in some geographic way to regional, national, or local scales of operation?
- **Privacy** requirements: Are digital asset privacy issues or mechanisms mentioned?
- **Ethics** issues: What reference is made to ethics in the context of managing or sharing digital assets?
- **Intellectual property** awareness: Is there any acknowledgement of the possible intellectual property management requirements for research-generated digital assets?

Licensing/copyright controls: Does the document cover the possible use of licensing and/or copyright mechanisms to facilitate data sharing?

Each criterion is given a subjective rating between 1 and 3 (traffic light concept) based on the level of detail and clarity provided in the document. The meaning of these three rating levels is described below:

3 – a strong emphasis on the criterion.
2 – limited emphasis on the criterion.
1 – very limited or no emphasis on the criterion.

It is important to note that the ratings given are subjective since the wording and level of detail in each document is often quite different. Because of this, a more empirical approach (applying “like for like” principles) to the scoring is not possible.

The ratings provided for each of the criteria can be used to assess alignment and non-alignment between the documents. In other words, if one document receives a rating of 3 for metadata and another receives a rating of 1 for metadata, then there is a non-alignment for this criterion between these two documents.

Many of the documents also refer to secondary documents that provide additional detail about the analysis criteria. When this occurs, it is considered that the primary document also contains this information, but with a much lower level of clarity and detail since the reader needs to go to multiple documents rather than a single document to obtain the specific policy requirement.

2.4 Criteria weighting

In addition to the rating on how each document meets the analysis criteria described above, each of the criteria is given a weighting consisting of a number from 1 to 5, with 5 signifying the highest impact. The weighting is based on a subjective view of the “influence” or impact each criterion has in relation to supporting compliance to open and FAIR data principles. For example, metadata is considered more significant than intellectual property in ensuring that data meets FAIR and open principles, and therefore metadata has a higher weighting.

A detailed description of the justification for the weightings of each criterion, and what the criteria summaries, ratings, weightings, and scores mean, is provided in **Annex 3**.

2.5 Analysis score calculations

Three numbers are provided for each of the 15 criteria, based on the subjective analysis of each document. The first number is the rating for the criterion based on the inclusion and perceived "strength" of the statement (s) describing that criterion. The second number is the "weighting" given to each of the 15 criteria, as described above. The final figure is the result of multiplying the rating by the weighting to provide an overall score for each criterion within each document.

The objective of this scoring system is to assist in identifying areas of alignment and non-alignment between these documents. For example, if one document is given a rating of 3 for FAIR and a second document has a rating of 1, then there is some level of non-alignment and potential risk (of data not being directed to be FAIR and open) in the document with the lower score. Each document can then be compared at the individual criterion level, and alignment and non-alignment for each criterion can be identified.

The outcome of this analysis is shown in **Table 1** and the accompanying **Figure 2.0** (Section 3.1).

2.6 Other donor policy documents reviewed

Additionally, a high-level review of data policies belonging to the other EiA donors (FCDO, USAID, NORAD, Irish Aid, ICAR) was also conducted to assess their overall narrative on data management and data sharing, and to compare this to the core EiA group (the foundation, OneCGIAR and its research centres).
The other donor documents are the following:

- FCDO [13]
    - Department For International Development (DFID) Research Open and Enhanced Access Policy[14]
- USAID

[13] UK Government. "Research at FCDO". Accessed 24 November 2023. https://www.gov.uk/government/organisations/foreign-commonwealth-development-office/about/research.
[14]UK Government, Department for International Development. "DFID Research Open and Enhanced Access Policy V1.1", 2013. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/181176/DFIDResearch-Open-and-Enhanced-Access-Policy.pdf.

    - Open Government Plan v4.0 (2016)[15]
    - USAID Scientific Research Policy[16]
- NORAD
- Irish Aid
- ICAR

A summary of the results of the analysis of other donor policies is provided in **Annex 5.**

[15] USAID (2022) *Open Government Plan*. 13 September 2022. https://www.usaid.gov/open/open-government-plan.
[16] USAID (2014) *Scientific Research Policy*. https://pdf.usaid.gov/pdf_docs/PBAAD895.pdf.

3 Analysis findings

3.1 Individual document findings

In the context of the EiA objective of ensuring that agricultural research data is both FAIR- and open-compliant, this desk review focusses on 10 different documents (listed above in **Section 2.1**) from the following programmes and organizations:

- EiA.
- The foundation.
- OneCGIAR.
- CIMMYT.

The review findings are presented using the following categories:

1. What are these organizations trying to achieve, what is their **philosophy and what are their objectives**?
2. What is the **significance of FAIR and open data principles** within the organization's documents?
3. Where do these policies and guidelines **align**?
4. Where is there **non-alignment** of the reviewed documents?

It was found that five documents have a high level of alignment with the criteria, have a medium level of alignment, and three have a low level of alignment (which in some cases is expected as the document is not intended to cover all aspects of data management).

The total level of alignment for each document against all of the analysis criteria is given in the table below, with the documents listed in order from the highest score (most aligned) to the lowest score (least aligned). This table is colour coded, with a high level of alignment shown in green, a medium level of alignment in orange and a low level of alignment or non-alignment in red.

Order of alignment	Document name	Alignment	Total score
1	Excellence in Agronomy (EiA) Proposal	Aligned	136
2	CGIAR - Open and FAIR Data Assets Policy	Aligned	120
3	EiA Best Practice Guidelines	Aligned	118
4	Bill and Melinda Gates Foundation – Data Sharing Requirements	Aligned	110
5	Bill and Melinda Gates Foundation – Data Guidelines	Aligned	108
6	CIMMYT – Research Data and Information Products	Partially aligned	94
7	Bill and Melinda Gates Foundation – Open Access Policy	Partially aligned	81
8	CIMMYT – Terms and Conditions Subgrant	Poorly aligned	77
9	CIMMYT – Intellectual Property Policy	Poorly aligned	71
10	Bill and Melinda Gates Foundation – Global Access Statement	Poorly aligned	70

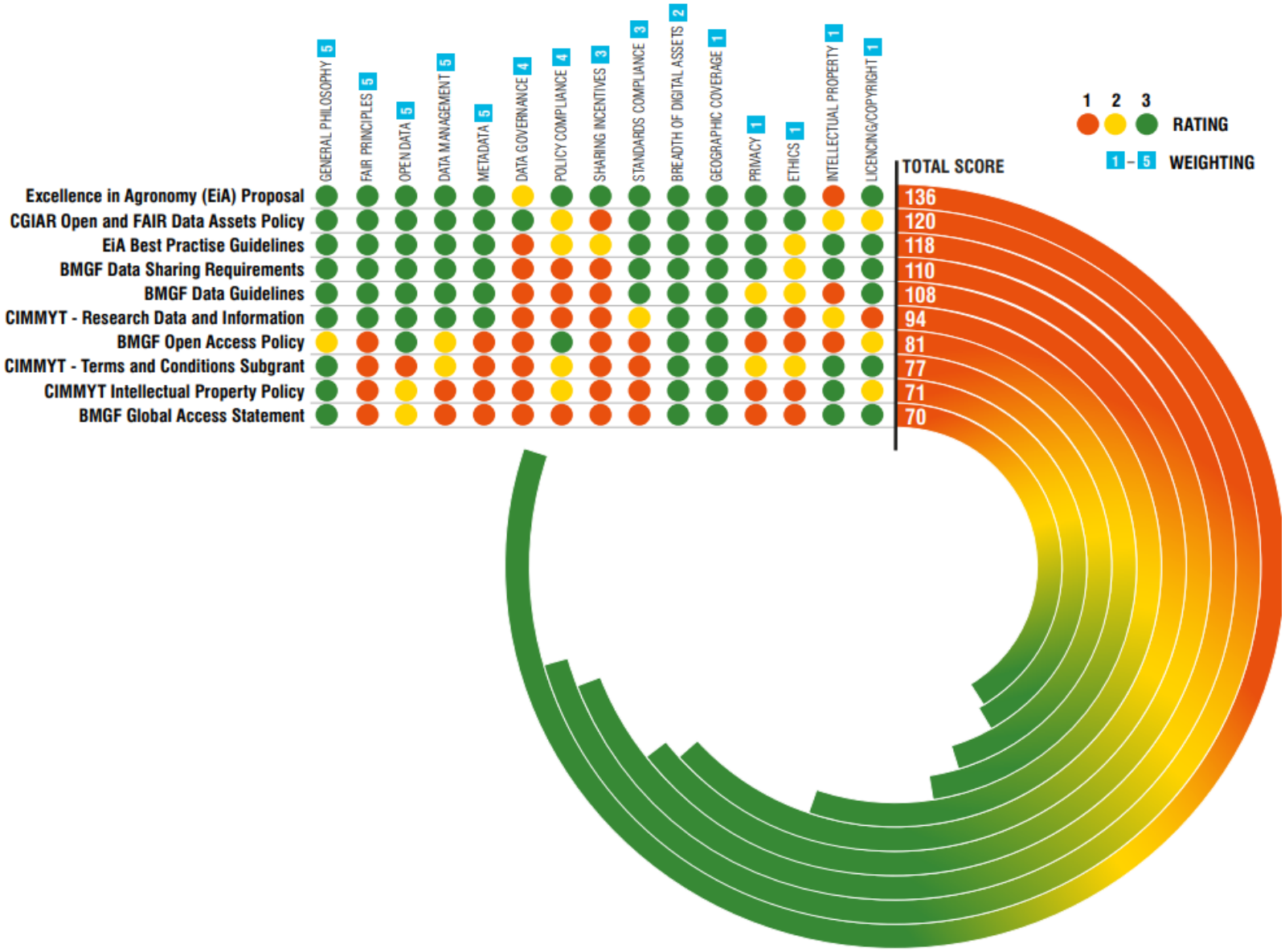
***Table 1** – Document scores against the analysis criteria (maximum score possible = 138)*

The table below shows the number of documents that were considered to be aligned to each of the analysis criteria. The criteria are listed from the highest weighting to the lowest. The table is also colour coded to indicate a high, medium or low level of alignment against each criterion.

Criteria	Weighting	Number of documents aligned	Alignment value
General philosophy	5	9	High
FAIR principles	5	6	Medium
Open data	5	7	High
Data management	5	6	Medium
Metadata	5	6	Medium
Data governance	4	1	Low
Policy compliance	4	2	Low
Sharing incentives	3	1	Low
Standards compliance	3	5	Medium
Breadth of digital assets	2	10	High
Geographic coverage	1	10	High
Privacy	1	5	Low
Ethics	1	2	Low
Intellectual property	1	5	Medium
Licensing/copyright	1	6	Medium

***Table 2** – Alignment of documents with analysis criteria*

A visualization of these findings using a traffic light system is shown in Figure 2.0 below, showing the total scores for each document, measured against the 15 criteria noted in the top row of the infographic.


GENERAL PHILOSOPHY 5
FAIR PRINCIPLES 5
OPEN DATA 5
DATA MANAGEMENT 5
METADATA 5
DATA GOVERNANCE 4
POLICY COMPLIANCE 4
SHARING INCENTIVES 3
STANDARDS COMPLIANCE 3
BREADTH OF DIGITAL ASSETS 2
GEOGRAPHIC COVERAGE 1
PRIVACY 1
ETHICS 1
INTELLECTUAL PROPERTY 1
LICENCING/COPYRIGHT 1
TOTAL SCORE
1 2 3
RATING
1 - 5 WEIGHTING
Excellence in Agronomy (EiA) Proposal 136
CGIAR Open and FAIR Data Assets Policy 120
EiA Best Practise Guidelines 118
BMGF Data Sharing Requirements 110
BMGF Data Guidelines 108
CIMMYT - Research Data and Information 94
BMGF Open Access Policy 81
CIMMYT - Terms and Conditions Subgrant 77
CIMMYT Intellectual Property Policy 71
BMGF Global Access Statement 70

3.2 Overall institutional findings

Table 3 below displays the evaluation scores for each organization. These scores were calculated by selecting the highest score given to each criterion from all the documents of the respective organization. In other words, if a document from an organization had a higher score against a criterion than other documents from the same organization, the higher score was used to indicate alignment with the specific criteria.

		EiA		BMGF		CGIAR		CIMMYT
No.	**Criteria**	**Score**		**Score**		**Score**		**Score**
1	General philosophy	15		15		15		15
2	FAIR principles	15		15		15		5
3	Open data	15		15		15		15
4	Data management	15		15		15		15
5	Metadata	15		15		15		15
6	Data governance	12		4		8		4
7	Policy compliance	12		12		8		8
8	Sharing incentives	9		3		3		3
9	Standards compliance	9		9		9		3
10	Breadth of digital assets	6		6		6		6
11	Geographic coverage	3		3		3		3
12	Privacy	3		3		3		3
13	Ethics	3		2		3		2
14	Intellectual property	3		3		1		3
15	Licensing/ copyright	3		3		1		3
	Total score	**138**		**123**		**120**		**103**

Table 3 – *Total organizational alignment scores*
Note: The maximum possible score against all criteria is **138.**

Based on the findings of the review at the organizational level, it can be concluded that EiA demonstrates full alignment with the analysis criteria when all documents from each organization are considered collectively. The foundation, as a donor organization, is the second most aligned organization, with a slight reduction in alignment observed when moving to the co-ordinating research organization (OneCGIAR) and then to the research centre (CIMMYT).

This can also be seen in Figure 3.0 below:

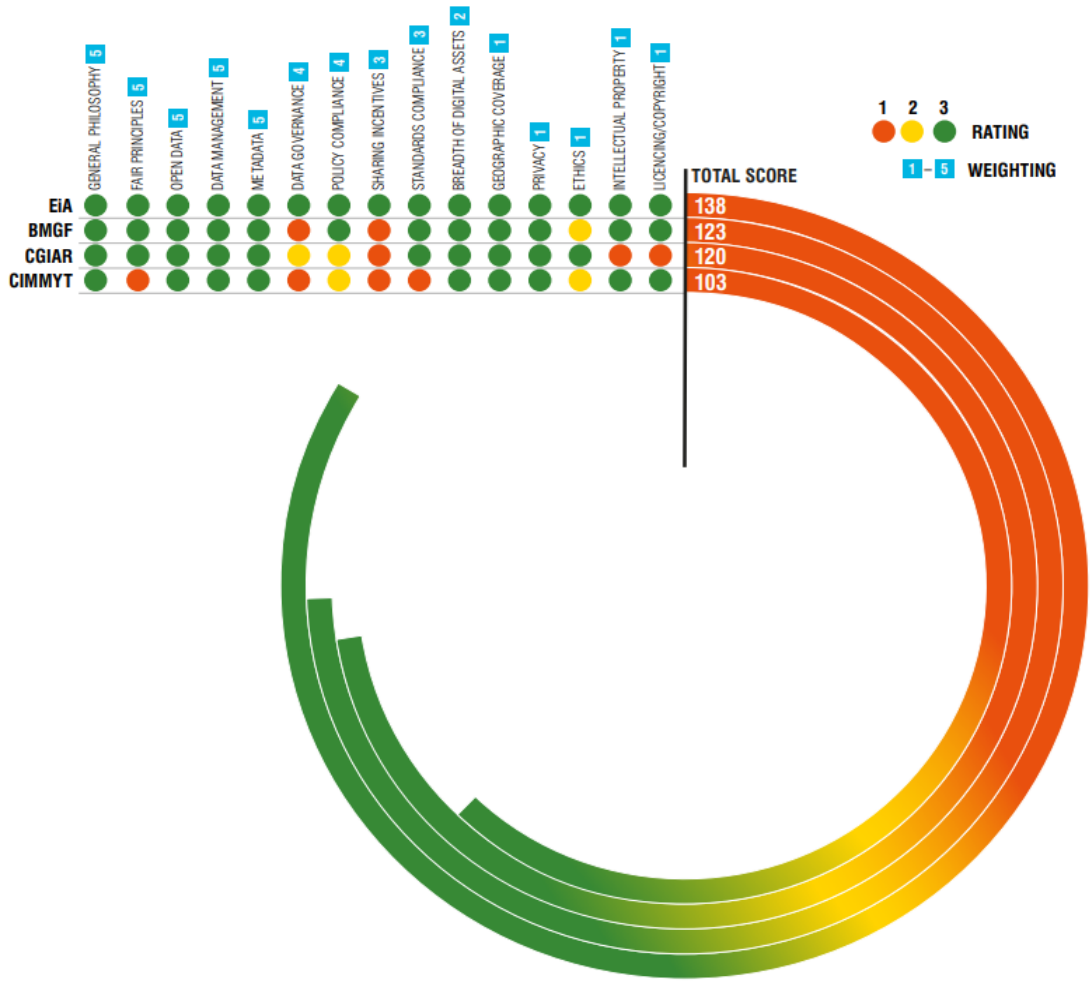

GENERAL PHILOSOPHY 5
FAIR PRINCIPLES 5
OPEN DATA 5
DATA MANAGEMENT 5
METADATA 5
DATA GOVERNANCE 4
POLICY COMPLIANCE 4
SHARING INCENTIVES 3
STANDARDS COMPLIANCE 3
BREADTH OF DIGITAL ASSETS 2
GEOGRAPHIC COVERAGE 1
PRIVACY 1
ETHICS 1
INTELLECTUAL PROPERTY 1
LICENCING/COPYRIGHT 1
TOTAL SCORE
EIA 138
BMGF 123
CGIAR 120
CIMMYT 103
1 2 3
RATING
1 - 5
WEIGHTING

Figure 3.0

A closer examination reveals that EiA and the foundation have a clear understanding of the policies required to ensure effective data management and data sharing in individual research projects, with a strong emphasis on FAIR-compliant and open data outputs. Again, the strength of policies decreases slightly at the co-ordinating research organization level (OneCGIAR) and reduces further again at the research centre level (i.e. CIMMYT).

There could be several reasons for this variation, but one major factor is likely to be the proximity between the research centre and the individual researcher, which gives them a different understanding of the practical constraints and limitations faced by implementing institutions and researchers in fully complying with FAIR and open data principles. Many donor organizations, especially those with a less engaged and less iterative model for grant-making, may approach data management from a more abstract, value-driven and philosophical perspective, due to being more distanced from individual researchers. Consequently, such donors do not make allowances for the real-world challenges faced by researchers who operate within a very particular set of career-relevant incentives.

3.3 Other donor findings

A summary of the findings of the high-level review of data sharing and management documents of the other donors of EiA is presented below. Their overall perspective on data sharing is compared to the findings for the "core" EiA group, to assess whether there is high-level alignment on the importance of complying with the FAIR and open data concepts across all donors and the grantee.

1) **FCDO: DFID Research Open and Enhanced Access Policy – summary**[17]

The DFID Research Open and Enhanced Access Policy is very focused on ensuring that research data is made open and available for all to use. It provides examples of the benefits of open data and makes some references to FAIR principles. It refers to an Implementation Guide that provides more specific details on how to ensure there is open access to research data. This

[17] UK Government, Department for International Development (2013) *DFID Research Open and Enhanced Access Policy V1.1*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/181176/DFIDResearch-Open-and-Enhanced-Access-Policy.pdf

policy can be considered to be aligned with the high-level objectives of the other polices reviewed but it contains little information to support compliance.

2) USAID – summary[18]

The USAID policy documents examined suggest that USAID has a good understanding of the requirements for effective data management and how the application of good practice leads to compliance with FAIR concepts. Where "good quality data" is available to all as open data, the benefits that are generated are much greater than is the case where data is not effectively managed.

It is considered that the current USAID documents are aligned with the EiA expectations regarding FAIR and open data principles, although we understand that access to data from all initiatives historically supported by USAID is not guaranteed. It is not known whether this is because data collected in those initiatives precedes the open mandate, or whether it is because initiatives are non-compliant with it.

3) NORAD – summary[19]

There is a low level of alignment by NORAD with the EiA views on FAIR and open data principles. This is primarily due to the high-level focus of the NORAD material, rather than any specific omission of what is a more technical focus for FAIR and open data.

4) Irish Aid – summary[20]

While it is most likely to be supportive of FAIR and open data, Irish Aid does not provide much detail on the application of these principles in the documents that were reviewed. There is a lack of alignment with the EiA FAIR and open data priorities.

5) ICAR – summary

ICAR has an open access policy which mandates the use of institutional repositories to curate all agricultural knowledge generated in ICAR, but the

[18] USAID (2023) "U.S. Agency for International Development". Accessed 16 November 2023. https://www.usaid.gov/.
[19] NORAD. "Research". Accessed 24 November 2023. https://www.norad.no/en/front/thematic-areas/higher-education-and-research/research/.
[20] Irish Aid. *Policy for International Development - Department of Foreign Affairs*. Accessed 24 November 2023. https://www.irishaid.ie/about-us/policy-for-international-development/.

emphasis is on research (journal) publications.[21] ICAR's institutional approach to research data management is covered in published ICAR guidelines, which emphasize the role of the National Data Sharing and Accessibility Policy for India as being the overarching framework with which ICAR complies.[22] The guidelines, published in 2014, predate widespread acceptance of the FAIR framework, but do indicate the merits of publishing selected data sets as open data, alongside journal publication. Sharing outside of this is not encouraged. "Cooling off" periods are one to three years for publishing data post-research completion and in some ways data sharing is actively discouraged: "*Before publication, there is no obligation to share any preliminary data that have been collected. In fact, sharing at this stage needs to be discouraged because the inferences from such data may not be conclusive while a project is still in progress*". As a result, ICAR can be considered to be partially aligned with FAIR and open data principles, but non-aligned with the incentives and aspirations of the EiA programme, which seeks to utilize and share raw data.

More detail on the review of other donor policies can be found in **Annex 5**.

ICAR. "ICAR Book Process". Accessed 30 November 2023. https://ebook.icar.org.in/index.php/bookprocess.

Department Of Science & Technology. *National Data Sharing and Accessibility Policy*. Accessed 30 November 2023. https://dst.gov.in/national-data-sharing-and-accessibility-policy-0.

4 Discussion

This review of data-related documents and policies belonging to the foundation, OneCGIAR, CIMMYT and EiA, highlights 15 key criteria which should be considered in the development of a FAIR and open data management and data sharing policy. However, the review identifies that **sharing incentives, policy compliance** and **data governance** are the top criteria requiring further attention from the institutions (bar EiA) (as seen in Figures 2.0 and 3.0).

Most of the documents do not describe any incentives for policy compliance, nor do they articulate the consequences of non-compliance. While it is acknowledged that there are currently minimal direct benefits to researchers from effectively managing and making their data open, little is said in the documents about ways to overcome this significant issue through providing incentives or imposing consequences.

Lack of compliance with the various data policies is considered to be predominantly the result of a human behavioural barrier, and approaches to overcome this barrier are addressed in the recommendations.

The study has focussed on the alignment or non-alignment of the policies and related documents. This approach has been determined by the wording and content of the various policies reviewed in the context of the analysis criteria. However, a perspective was gained during the review that is considered important in relation to the EiA-related organizations achieving their policy objectives. This perspective has not been articulated previously in this report and is not addressed in the policy documents. The issue is the subtle and often conflated view of the technical,[23] human behavioural,[24] and cultural[25] elements that hinder compliance with FAIR and open data principles, which results in constraints on the effective management, sharing and reuse of agricultural data.

[23] "Technical" refers to the technical equipment and systems available for a researcher, and also the technical skills of the researchers and data managers involved in the research activity.
[24] "Human behaviour" refers to the perceptions a researcher or a research centre may have about sharing data – e.g. does sharing increase or decrease their "power and influence"?
[25] "Culture" refers to the need to make a decision that is not dependent on technical issues, but rather on the willingness of an organization (or country) to share data. If an organization is unwilling to share data, the decision to reverse this is considered to be purely cultural.

When real-world technical, human behavioural, and cultural issues arise, it can become difficult to conform to the policies. While it is acknowledged that the policies and guidelines are certainly required, to achieve the outcomes of FAIR and open data compliance desired by the donors, research organizations and research centres, there needs to be recognition of these distinct, yet intertwined issues, as summarized in Figure 4.0.

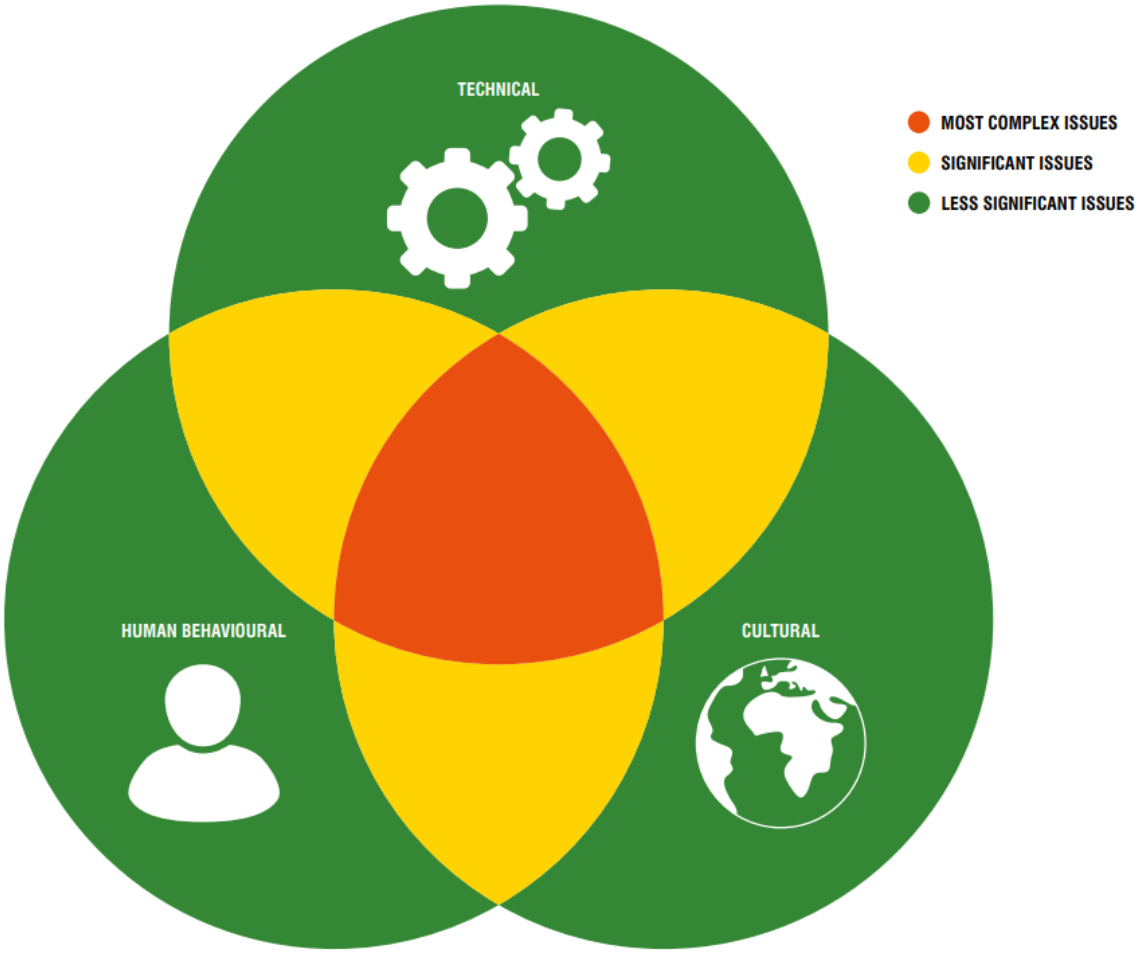


Figure 4.0

4.1 Technical issues

Technical issues in this context are considered to cover many different areas, such as the following:

- The level of data management capabilities, knowledge and experience of researchers and research centres.
    - Significantly different levels of “maturity” in FAIR-focussed data management capabilities will exist among researchers.

    - There are different levels of capability, understanding and infrastructure between the research centres of developed and developing countries and between individual research centres.
- Data and data management standards maturity.
    - Are there relevant data standards for specific agricultural research activities and are they being applied?
    - If they do not exist, are efforts under way to rectify this?
    - If community data standards do exist, how can key influencers in the FAIR and open data ecosystem, like the foundation, encourage the adoption of existing tools to generate standards-compliant interoperable data, such as those developed through EiA and the Big Data platform?[26]

4.2 Human behavioural Issues

The thought processes and perspectives of individual researchers as regards the importance of data management play a major role in the compliance or non-compliance with data policies. Additionally, the views of the research centre, and those of national governments in relation to sharing data across sovereign borders, also contribute to researchers' “desire” to ensure compliance with FAIR and open data principles.

Some of these human behavioural barriers are as follows:

- Many researchers do not see the value in making their data available. They gain little or no benefit from managing data at a FAIR-compliant level or having their data openly available. The research centre may have a similar view of the value of FAIR and open data and not wish to use resources to support these ideals. This is more likely to be the case if the research activity is small-scale and there is a limited need to collaborate internally or with other organizations.
    - When research is conducted by one researcher there is no benefit to them having the data in a state that is suitable for sharing. As more researchers are involved, and when multiple research centres participate in larger research activities, there is a greater imperative for data to be shared, at least between the research partners, necessitating some level of agreement on data management activities.

[26] CGIAR. *CGIAR BIG DATA Platform - CGIAR Platform for Big Data in Agriculture*. Accessed 24 November 2023. https://bigdata.cgiar.org/.

- A “what’s in it for me?” view in relation to data management and data sharing is common among researchers, despite some of these researchers being bound by public goods mandates if they work in government- or donor-funded institutions. There are currently minimal direct benefits to researchers from sharing their data, even if policies say that they should do so.

4.3 Cultural issues

The cultural issues that create barriers to sharing data are many and complex. They include organizational and government perspectives, which can include the following:

- The cost of training and infrastructure development, with potentially a perceived lack of any direct benefit from data sharing, is seen as a waste of resources. The focus is on producing more research, rather than the aggregation of research data to produce more effective research outcomes.
- There is a perception that data management provides no, or very limited, benefit to the host centre or the country involved.
- Some countries fear that their sovereignty will be compromised by sharing data internationally.
- An organization’s monetization of data – the value of data in a digitized world – means data is often treated as a private good, and accessibility is limited so that it can be traded to support profitability, which can be an important part of a sustainable business model. However, this slows down the speed at which innovative solutions can be made and further sharpens the exclusion of marginalized communities that need these solutions. The desire to monetize data obtained from research may run counter to the global public good mandate of some funders.

4.4 Additional observations

One observation that was made during the policy alignment review relates to the lack of broad alignment between all of these policies and EiA’s overall data-related objectives. EiA has some very substantial data-related objectives, with a focus on big data, big data analytics and agronomy-at-scale solutions as key outcomes. However, the reviewed policies are predominantly focused on individual research projects, not on EiA’s big data aspirations.

A research project that fully complies with the various FAIR and open data policies may still not help EiA to achieve its big data-focused outcomes. The reason for this is that to support big data needs, it must be possible for the data collected in the various projects to be aggregated to create large, well-structured data sets that can then be manipulated by sophisticated data analytic tools. Without standardization, it will be difficult to achieve integration, and therefore the potential to create large data sets for the analytical tools. Since not all of the policies provide reference to the application of data standards or data management standards, they are considered to not be aligned to fully support EiA objectives.

While this issue was not part of the review, it is considered that the policy gap is significant enough to raise. More detail on what this gap is, and some thoughts and recommendations on how to resolve it, are provided in **Annex 6.**

5 Recommendations

Taking a people-first approach is necessary when tackling the intertwined technical, behavioural and cultural barriers, recognizing the need to understand the different people involved in agricultural data ecosystems in order to find relevant solutions.[27]

The recommendations provided below consider these intertwined issues and aim to support the alignment of institutional policies and agendas around the FAIR and open data principles to ensure the objectives of EiA can be met.[28]

In this section, CABI provides recommendations for the Program Officer, grantee and other EiA donors, highlighting areas where we may provide potential additional support. However, this is limited, due to the scope of work agreed with the foundation. The recommendations focus on incentives for sharing, policy compliance, data governance and multi-donor alignment.

5.1 Recommendations for the Program Officer/the foundation

There is an opportunity for the foundation to use EiA as an exemplar within the soil and agronomy community for how standardized data can be leveraged in a multi-donor-funded initiative of this level. This needs efforts to be put into achieving alignment with partners, not just by means of a Data Governance Framework (see **Annex 7**) but also by including real changes in the data management and sharing practices of all persons in the EiA data ecosystem, through applying appropriate incentives and compliance methods. If this is not done, there is a real risk that there will be a time cost involved in securing permission to access key assets for EiA, or even that key assets will not be accessible at all. This would significantly impact the programme's decision-support mechanisms and could affect the reusability of the data.

[27] Smith, F., Dodds, L., L'Henaff, P. *et al.* (2018) "Understanding personas in agricultural data ecosystems" [version 1; not peer reviewed]. *Gates Open Res* 2018, **2**:43 (document) (https://doi.org/10.21955/gatesopenres.1114887.1)
[28] CGIAR. "Excellence in Agronomy". Retrieved 11 September 2023, from https://www.cgiar.org/initiative/excellence-in-agronomy/

CABI has a helpful guide[29] that provides suggestions on possible incentives and compliance methods for multi-donor initiatives, like EiA. Some of the following recommendations built upon this guide. Other research evidence that provides examples of where similar recommendations have worked, particularly in the medical sector, are also referenced.

CABI can provide recommendations to the FAIR Data Working Group within the foundation to understand institutional deficits related to FAIR, including which a data governance template for EiA (as well as for other investments) can be provided (see **Annex 7**).

We recommend the foundation does the following:

- Invest in maintenance and improvement of standards to achieve FAIR and to "facilitate more meaningful exchanges of data",[30] following the example of NCBI in the health sector. Additional observations on this are given in **Annex 6**.
- Credit good data management practice and sharing in grant reviews.
- Ensure that research publications also support the ability to cite a researcher's data set.
- Use EiA as a model for how good data practice can be intentionally factored into planning and grant-making, including the importance of data management tools and FAIR data assessments/scoring tools, such as those that have been developed specifically for EiA, CGIAR GARDIAN tools,[31] and possibly the FAIR Implementation Profile template (see **Annex 8**), as developed by GO FAIR.[32]
- Set out in grant contracts where to locate open-source resources and tools (as referenced above) recommended by the foundation to support grantees meet the FAIR data requirements of an investment.
- Include in grant contracts a detailed section on operational costs reserved for achieving FAIR and open data compliance; if it is explicit that there is specific funding for this, this could motivate recipients to change their habits in order to receive this funding. For example,

[29] Musker, R. and Smith, F. (2021) "Incentive systems for research data sharing in funded projects" [version 1; not peer reviewed]. *Gates Open Res* **5**:84, p.9 (document) (https://doi.org/10.21955/gatesopenres.1116783.1)

[30] Gliklich, R.E., Leavy, M.B. "Data Standards". In: Gliklich, R.E., Leavy, M.B., Dreyer, N.A. (editors) *Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide*, 3rd Edition, Addendum 2 [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2019 Oct. Chapter 3. Available from: https://www.ncbi.nlm.nih.gov/books/NBK551886/

[31] CGIAR. *GARDIAN*. Retrieved 11 September 2023, from https://gardian.bigdata.cgiar.org/#/tools

[32] https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/

some research suggests that 5% of research funds should be allocated to data management[33] and that there should be at least one data steward per 30 researchers to uphold best practice implementation of the FAIR data principles.[34]

- Link the final tranche of grant funding to compliance with publishing data sets and other relevant project assets in approved repositories, as defined in a mutually agreed Data Governance Framework for the investment.[35]
- Reinforce the potential socio-economic benefits of aligning on FAIR and open data practices, along with other influential stakeholders in the soil and agronomy field (see footnote for an impact case study that serves as an example).[36]
- Strengthen compliance messaging in the Open Access Policy by making the consequences of non-compliance more explicit.[37]
- Use enforcement (for compliance[38]) as a complementary tool, alongside positive incentives,[39] to ensuring FAIR and open data compliance. Like most people, researchers will ***"follow the money"***: if they realize that they are reducing their ability to secure future funding by not complying with the required data policies, they will change their approach to data management.
- Do not publish a research paper unless data sets are included with the paper or deposited in an acceptable repository – as per the foundation's Open Access Policy.
- Ensure that where a researcher does not comply with FAIR principles, they are excluded from obtaining another grant from that donor.

---

[33] Mons, B. (2020) "Invest 5% of Research Funds in Ensuring Data Are Reusable". *Nature* 578, no. 7796: 491–491. https://doi.org/10.1038/d41586-020-00505-7.

[34] Support and information Wageningen Data Competence Center Contact form. WUR (2021) "Data Stewardship at WUR". https://www.wur.nl/en/value-creation-cooperation/collaborating-with-wur-1/wdcc/research-data-management-wdcc/data-stewardship.htm.

[35] We believe Medha has led a report on data governance for EiA that may be useful for developing this Framework.

[36] ODI (2016) "Open Data's Impact", *NOAA Open Data Portal*. https://odimpact.org/files/case-studies-noaa.pdf

[37] The foundation includes enforcement approaches in its Open Access Policy, stating that "compliance is a requirement of funding" and "compliance will be continuously reviewed", but it is not clear what the consequences of non-compliance are. (*How to Comply*. Gates Open Access Policy. Retrieved 11 September 2023, from https://openaccess.gatesfoundation.org/how-to-comply/)

[38] Federer, L.M., Ya-Ling, L., Joubert, D.J., Welsh, J., and Brandys, B. (2015) "Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific Research Staff", *PLoS ONE* 10 (6): e0129506. https://doi.org/10.1371/journal.pone.0129506.

[39] Leonelli, S., Spichtinger, D. and Prainsack, B. (2015) "Sticks *and* Carrots: Encouraging Open Science at Its Source". *Geo: Geography and Environment* 2, no. 1: 12–16. https://doi.org/10.1002/geo2.2.

5.2 Recommendations for the grantee

- Conduct further research into why their researchers are not managing and sharing their data in ways that align with the OneCGIAR policy, such as through surveys or interviews (see example studies in the footnotes)[40,41]
- Establish training and education programmes for early-career researchers in best practice data management methods, linking good practice to increased chances of promotions/career development
- Invest efforts in developing communities of practice, facilitating spaces and platforms for knowledge exchange and sharing, using Ethiopia's Coalition of the Willing as an example to share with national partners.[42]
- Demonstrate how a researcher's reputation can improve if their quality data is easily available and can be shared.
    - For example, publishing data sets in dedicated peer-reviewed data journals adds to metrics (e.g. citations, h-index) that researchers may be assessed against. An example dedicated journal is *Scientific Data*.[43]
- Turn the implementing guidelines documentation that EiA has already produced into a legally binding document, and model this for future multi-donor projects
- EiA also mentions compliance with the OneCGIAR policy, but there is no mention of what the consequences of non-compliance are. Consider adding explicit caveats into the policy, such as a reporting component within projects, involving providing feedback to donors.
- Gather key stakeholders within the CGIAR leadership, along with EiA donors, to a convening to bring about alignment on the importance of data sharing for ensuring EiA's success.

5.3 Recommendations for all (including other donors):

[40] "Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific Research Staff", *PLOS ONE*. Accessed 24 November 2023. https://doi.org/10.1371/journal.pone.0129506.t016

[41] Dyke, S.O., Hubbard, T.J. (2011) "Developing and implementing an institute-wide data sharing policy". *Genome Med* **3**, 60. https://doi.org/10.1186/gm276

[42] Coalition of the Willing (2020) *Coalition of the Willing for soil and agronomy data access, management and sharing. Data sharing guidelines*. Addis Ababa (Ethiopia): Ethiopian Institute of Agricultural Research (EIAR). 32 p. https://hdl.handle.net/10568/107988

[43] *Nature.* "Scientific Data". Accessed 24 November 2023. https://www.nature.com/sdata/.

- To address the lack of data governance and improve alignment, co-create a Data Governance Framework for the programme (see examples in footnote[44]) to seek agreement from all donors on standardizing data policies. Ensuring that the policies contain all relevant criteria (activities, steps, procedures) necessary for establishing effective data management for research projects that result in compliance with FAIR and open data principles. If the standardization of data management policies also included research organizations and research centres, there would be considerable benefits (**see Annex 1**).[45] Agreement would be required on what elements and processes would be needed to cover all necessary components in a standardized data management policy. For example:
    - Including elements such as those used in the analysis criteria in this review, including metadata, data standards, data governance, intellectual property management, etc.
    - Creating a list of "ideal" data management steps for all stages in the data lifecycle that all organizations should include in their policies.
    - Having a common set of guidelines and policies that address each of the agreed data management steps.
    - Including a list of tools that EiA has developed to (1) enable collection of standards-compliant data; (2) render legacy data standards-compliant; (3) contribute to and co-develop these tools; (4) contribute to and co-enhance the global agronomy database – potentially through a foundation-mediated memorandum of understanding. The latter would allow multiple entities beyond CGIAR to derive value from open and FAIR data and demonstrate its importance and utility.
    - Identifying outlying issues, such as intellectual property, ethics and privacy, and linkages to the private sector, and developing common frameworks to ensure there is a consistent and beneficial approach to them.

[44] Republic of Kenya Ministry of Agriculture, Livestock, Fisheries and Cooperatives (2022) *Data Governance Framework*. https://kilimo.go.ke/wp-content/uploads/2022/04/MoALFC-Data-Governance-Framework-2022.pdf
https://datagovernance.com/wp-content/uploads/2020/07/dgi_data_governance_framework.pdf

[45] See example from ACIAR. "An Assessment of Data Management and FAIR Data Principles across the ACIAR Research Portfolio - Final Report". Accessed 24 November 2023. https://www.aciar.gov.au/publication/slam-2021-156-final-report.

- Agreeing on and aligning around a set of common data management and FAIR-related terminologies and ensuring these are clearly defined and commonly understood.
    - Consistency in the terms used across the various organizations' documents would simplify and clarify the intent of the documents for the researcher and donor recipient.
- In all policy documents, stressing the significance of effective data management and compliance with the FAIR principles to maximize the value of shared data for others.
- To address the lack of incentives, increasing the "value" of research data by:
    - Ensuring research publications used by donor-funded researchers also support the ability to cite a researchers' data set.
    - Demonstrating the benefits of being able to aggregate data into large data sets, with the effect of improving future research outcomes.
    - Demonstrating how big data is becoming a critical element in achieving agronomy-at-scale solutions and in improving agricultural policy development.
    - Addressing the issue around the monetization of data as part of a sustainable business model by considering some data as a "club good",[46] meaning it is considered as being between a public and a private good. The agriculture sector could follow the example of the music industry, which developed "common technical standards for mass consumption, an IP Framework to protect data producers and incentivize sharing, centralized hosting and distribution infrastructure, and cost-sharing arrangements".[47]
- Establishing training and education programmes for early-career researchers in best practice data management methods.
- Ensuring that research proposals contain a section on data management and include a suitable budget (in the past, under the platform for big data, CGIAR has produced templates that can guide this exercise); it has been suggested that 5% is a minimum that might

[46] "They are goods that are non-rivalrous, but excludable. They include things such as satellite TV, private parks, and movie theatres." https://corporatefinanceinstitute.com/resources/economics/club-goods/

[47] Schwartz, H. (2017) "Club goods, intellectual property rights, and profitability in the information economy". *Business and Politics, 19*(2), 191-214. doi:10.1017/bap.2016.11

be allocated to making sure that project assets are made open.[48] The research contract should also have a detailed section on data management expectations and an identified budget for achieving compliance with FAIR and open data principles.

- Engaging in an open dialogue between donors and CGIAR on how research centres can better reward researchers for good practice. How can researcher key performance indicators (KPIs) better reflect donor mandates to share data by complying with FAIR and open principles, not just in research publications, but also its data sets?
- Working with research organizations and research centres to strengthen data management support at the research centre level, using an approach such as that described in a paper[49] Monash University, Australia.
    - A data concierge is described in the paper as "A person (or persons) who provides advice and direction to researchers to help resolve their data management issues."

[48] CGIAR. "Open Access FAQ". Accessed 24 November 2023. https://www.cgiar.org/how-we-work/accountability/open-access/faq/.

[49] Splawa-Neyman, P (2022) "What the dickens is a data concierge? Researcher interviews and data management reviews: misinformation, appreciation and remediation". Conference contribution. https://doi.org/10.6084/m9.figshare.20104631.v1

6 Conclusion

The review of the selected data management and data sharing-related documents suggests that the "parent" organizations all have a broadly similar objective in relation to the digital asset outputs from their agricultural research. This objective is to ensure that **ALL** data assets created during agricultural research activities are shared openly.[50] The concept of data sharing is a prime focus of all of these documents and the reasons used to justify the sharing of data also show similar motivations. The underlying premise and objective of these documents relates to the dependency on data and information to make timely and accurate (agricultural) decisions. The improvement of agricultural output is also seen as highly dependent on, and enhanced by, the sharing of digital assets (predominantly data).

To meet the broad data-related objective described above, each organization has developed a range of documents to ensure that the value of data collected through their research projects is maximized and made available, in many cases as a public good.[51]

In general terms, this "maximization" of the value of data is seen by all organizations to be achieved through the sharing of data and applying open data principles. Additionally, to ensure that the data is easily reusable, the FAIR data principles are also generally referenced by these organizations' policies and procedural documents. A final point, not specifically mentioned in most of the documents reviewed, is that the aggregation of research data into much larger data sets can bring a multitude of additional benefits beyond those generated through simply sharing the data with other researchers. CGIAR articulates this objective by saying:

> *"Open and FAIR data assets improve the speed, efficiency, and efficacy of research; they facilitate interdisciplinary research; assist data aggregation, computation, and the derivation of new insights; and allow the global public to benefit from CGIAR Research. They enable CGIAR to*

[50] Unless there are legitimate reasons for not doing so, such as privacy or commercial confidentiality.

[51] A "public good" is defined by the International Monetary Fund as follows: "*Public goods are those that are available to all ('nonexcludable') and that can be enjoyed over and over again by anyone without diminishing the benefits they deliver to others ('nonrival'). The scope of public goods can be local, national, or global."*

> *collectively leverage the infrastructure, data pools, and new data science capacities necessary for innovation and for effective and agile responses to global challenges. They facilitate text and data mining and analysis to derive insights, recognizing that these are dependent not only on access to high quality data, but also on that data being well-contextualized (through rich metadata and relevant open materials), interoperable, and reusable."*[52]

To conclude, this desk review has focused on the level of alignment or non-alignment of a range of data management policies of donors, research organizations and research centres, in relation to the FAIR and open data principles. In broad terms, the policies reviewed can be considered to be aligned with the FAIR and open data objectives and there is a general level of alignment between the organizations, although with a gradual reduction in alignment when moving from the EiA policy, to the donor's (the foundation's) policy, to the research organization's (OneCGIAR's) policy and then to the research centre's (CIMMYT's) policy. If the recommendations from this review are applied, EiA will be set up for the successful achievement of its objectives and will be in a position to contribute to catalysing the adoption and implementation of the FAIR and open data principles within the wider data ecosystem.

[52] CGIAR (2021) *CGIAR Open and FAIR Data Assets Policy*. para 1.2. https://cgspace.cgiar.org/bitstream/handle/10568/113623/CGIAR_OFDA_Policy_Approved_16April2021.pdf?sequence=1&isAllowed=y

Acknowledgements

This report was prepared for the Bill & Melinda Gates Foundation. The findings and conclusions contained within it are those of the authors and do not necessarily reflect the positions or policies of the Bill & Melinda Gates Foundation.

Lead author: Ben Searle, CABI Consultant.
Co-authors: Martin Parr and Melissa Allan, CABI.
Contributors: Arun Jadhav, CABI, and Medha Devare, CGIAR.
Editors: Mark Palmer and Melissa Allan, CABI.

How to cite this report: Searle, B., Parr, M. and Allan, M. (2023) "FAIR and Open Data Alignment Across Excellence in Agronomy Initiative-Related Organizations".

Annexes

1. Open data to improve research outcomes

Research processes

Research organizations and their donor partners are working to achieve improvements within the global agricultural community, with a specific focus on developing countries. Agricultural-related research is being undertaken with a range of outcome objectives, including climate change mitigation, productivity improvements, and gender equity. The research programmes aim to provide improvements in many areas of agricultural endeavour around the world.

At a conceptual level, research can be considered to follow a basic workflow, consisting of defining a problem, collecting data about that problem, analysing the data, and writing a research paper. The primary focus for the researcher is usually the research paper, and the data collected for the research is often given limited attention. The researcher knows how the data was collected, where it is held, what format it is in, etc. Processing the data to make it reusable for others is an effort that brings the researcher little, if any, direct benefit.

This paradigm of a researcher not focusing on the external value of data applies in most research sectors. However, some sectors have successfully addressed this issue, identifying the considerable benefits of good data management and data sharing and creating standardized aggregated data sets. In some science sectors where this has occurred, new commercial opportunities have been established to leverage the availability of these large integrated data sets, bringing considerable benefits.

Economic benefits of open data

As an example of the benefits of open data, prior to the World Meteorological Organization (WMO) being formed in 1950, individual scientists and the national meteorological organizations of each country predominantly collected, managed and analysed their own data. The quantities of meteorological observations were therefore limited to what each country was able to collect. The establishment of the WMO resulted in the creation of rules, standards and a range of procedures to effectively enable the sharing of an agreed range of meteorological data types. This sharing has resulted in

significantly larger quantities of data being available to each national meteorological agency, resulting in much more accurate predictions and forecasts at the global, national and local scales. The benefits of this are spread across many sectors, including agriculture. The economic benefits from this in the USA alone run into billions of dollars annually and have resulted in the creation of whole new sectors, such as the multi-billion-dollar weather derivatives financial industry.

The Open Data Institute has described the benefits achieved through the USA's National Ocean and Atmospheric Administration's (NOAA's) open data philosophy in the following way:

> *"Since its inception, NOAA has boasted a strong open data culture and is considered a leader in open data, if not the leading open data example among government agencies. When the Obama administration launched* ***data.gov*** *as part of its flagship Open Government Initiative in January 2009, NOAA was cited as the paradigmatic example as to how government agencies can both publish data and make that data accessible for the private sector to use and build a multi-billion-dollar industry.*[53]
>
> *Opening up weather data through the United States National Oceanic and Atmospheric Administration (NOAA) has significantly lowered the economic and human costs of weather-related damage through more accurate forecasts; the development of a multi-billion-dollar weather derivatives financial industry; and the growth of a million-dollar industry of tools and applications derived from NOAA's real- time data. In many ways, the industry built around NOAA's weather data is seen as the paradigmatic example of how the release of open data can yield major economic impacts."*[54]

Some specific economic benefits resulting from NOAA's open data policy[55] are given below:

1. *NOAA real-time data supplies a burgeoning private weather service industry, with well over $700 million in value added annually.*

---

[53] ODI Impact. "United States' NOAA: Opening Up Global Weather Data in Collaboration with Businesses". Accessed 24 November 2023. https://odimpact.org/case-united-states-noaa-opening-up-global-weather-data-in-collaboration-with-businesses.html.

[54] ODI (2016) "Open Data's Impact", *NOAA Open Data Portal*. https://odimpact.org/files/case-studies-noaa.pdf

[55] OECD (2008) "Assessing the Economic & Social Benefits of NOAA Data". https://www.oecd.org/sti/ieconomy/40066192.pdf

2. *The USA's $8–10 billion and growing annual weather derivatives financial industry relies on NOAA's seasonal weather data and records.*

The benefits from sharing meteorological data are considerable, with massive economic, social and environmental impacts. However, to achieve these benefits agreement had to be reached and training, education, work processes, policies, incentives and compliance requirements established. While this was a non-trivial exercise, the demonstrated benefits have far outweighed the effort expended to achieve effective data sharing.

An additional and often unrecognized benefit of developing agreed protocols, standards, rules and procedures is that the commercial sector is able to more easily (and cheaply) develop technology to leverage the standardized data. This results in technology being more accessible than if every organization developed bespoke systems to support its internal data formats, data management, processing and analysis needs. A standardized approach encourages and supports public–private partnerships and further increases access to, and the utilization of, aggregated data sets.

Policy benefits of open agricultural data

The United Nations Food and Agriculture Organization (FAO) operates CountryStat, a web-based source of food and agricultural statistical data. FAO views data as being highly critical in the development of policies and in supporting the achievement of the Sustainable Development Goals. In relation to "Data for Policy",[56] the FAO states:

> *"In today's data-driven world, policymakers rely more than ever on accurate information and analysis to make informed decisions. Yet, missing and poor-quality data can hamper the design of successful policies – thus, jeopardizing the achievement of the SDGs".*

FAO further adds that:

> *"By providing free access to reliable data, the Organization helps to make informed policy decisions and supports civil society, the private sector and policy practitioners to undertake evidence-based research".*

Food and Agriculture Organization of the United Nations. "Data for Policy". Accessed 30 November 2023. https://www.fao.org/policy-support/policy-themes/data-for-policy/en/#c869797.

It is well understood in the agricultural community that access to quality data improves our knowledge and provides a wide range of benefits to the community, especially in developing countries.

2. Factors influencing the review

The documents reviewed, covering policies, guidelines, agreements and other related material, are all broadly aimed at defining requirements for researchers and their organizations in regard to complying with the FAIR principles, and open data and data sharing concepts. They all have a similar high-level philosophy and objectives, aimed at encouraging the sharing of all research-generated digital assets so that others can use the material to increase local, regional and global agricultural understanding and knowledge.

However, the documents have different purposes and multiple and overlapping orientations, drivers, perspectives and motives in relation to data management, data sharing and FAIR and open data principles. For example, some are described as policies or guidelines and some as agreements. The different "orientation" of these documents results in slightly different "biases" regarding what they contain, what they are trying to achieve and how the requirement for achieving data sharing is worded.

The variation in the approach to achieving similar goals adopted by each document also reflects the document owner's role and function. The focus, priorities and interests vary according to their role in the research process: for example, whether they are funding organizations, international research organizations or the regional research centres.

It is recognized that these different roles, varying scales of operation and specific requirements of the organizations result in subtle differences in the focus and content of their policies and other documents.

Factors impacting compliance with data policies

It is also important to appreciate that there will also be different interpretations and understandings of these documents, and different degrees of compliance with them, from the researchers and their host organizations, due to a number of factors. These factors include the following:

- Research and data management capability, infrastructure and maturity levels.
- Data management experience and available expertise.
- Academic background, education and training.
- Organizational-level data management policies, awareness, capability, infrastructure and sophistication.
- Human behavioural issues
- Organizational administrative frameworks.
- Cultural issues, such as national legislation and related policies on handling and sharing data and other digital assets.

From a practical perspective, there will be varying views and perspectives, understandings and interpretations concerning data management in general, and the FAIR and open data principles in particular. Variations also arise between developed and developing country-based scientists and their organizations. Additionally, the motivations and potential incentives in regard to complying with FAIR and open data policies will also differ. The barriers to implementing and supporting FAIR and open data are also likely to differ between organizational and national levels as well.

3. Details of analysis criteria

The analysis criteria and weighting values are set out in the table below, together with a brief description of each criterion and some justification for the weighting value given to it.

Analysis – rating and weighting descriptions

No	Criteria	Description	Weighting	Weighting description
1	General philosophy	Does the document reflect an overall philosophy in relation to data sharing, and to FAIR and open data principles in particular? Does the document aim to achieve open and FAIR data compliance?	**5**	The broad philosophy of FAIR and open data as set out in each document is considered significant in regard to its ability to impact FAIR and/or open data compliance.
2	FAIR core principles	Are these referred to in any detail?	**5**	The level of detail regarding FAIR principles is considered an indication of the understanding of the value of FAIR-compliant data sets.
3	Open data	Is any reference made to open data or open access to digital assets?	**5**	Mentioning open data is a clear recognition of the value of making data available to the broad (global) community without restriction.
4	Data management	Is the concept of data management (or the effective management of digital assets) mentioned in the context of the research projects?	**5**	Without effective data management, the ability to easily reuse a digital asset is reduced.
5	Metadata	Is there any mention of the concept of metadata?	**5**	Metadata is a critical element of effective data management and without it the potential findability and

				usability of data is weakened considerably.
6	Data governance	Is data governance for a research project mentioned or referenced?	**4**	Effective data governance is more likely to lead to effective data management, which is considered a critical element in meeting the FAIR objectives and enhances open data principles.
7	Policy compliance requirements	Does the policy document make any reference to the need to comply with the policy and how strong is the compliance requirement?	**4**	The stronger the compliance requirements are, the more likely it is that the objectives of the policy will be met, so it is considered that this is important in regard to meeting FAIR and/or open data principles.
8	Incentives to encourage the sharing of digital assets	Does the document provide any incentives or refer to incentives for the effective management of digital assets and/or specifically encourage the sharing of research-developed digital assets?	**3**	Given the general understanding that it is often difficult to get researchers to make their data available or to comply with the FAIR or open data principles, the concept of incentives is considered important from a practical perspective. This is the “carrot” approach to improving data access.

9	Compliance with standards	Does the document contain any reference to complying with standards including data formats, metadata structures and other relevant standards?	**3**	While compliance with standards is important and can facilitate interoperability, it is not as critical to reusability as metadata is. If a data set is well described, then even if it is not standards-compliant, its metadata format description etc. will assist its reusability.
10	Breadth of digital assets	Does the policy go beyond just data and include other digital assets, such as documents, software, models, etc.?	**2**	While it is useful to see more, rather than fewer, digital assets included in a policy, their inclusion does not significantly strengthen the desire of the policy to meet FAIR and open objectives
11	Geographic coverage	Does the document aim to influence or manage activities at a particular scale, e.g. locally, nationally, regionally, or ideally globally?	**1**	Geography, while important, does not necessarily reflect the document's strength in complying with FAIR and open data principles, but may confine it to an area that is sub-global.
12	Privacy	Is the privacy of data (personal or perhaps commercial) described in any way?	**1**	While privacy is an important issue, mentions of it in the policy do not add significantly to the policy’s strength in

				relation to supporting FAIR and open data principles.
13	Ethics	What references are made to ethical issues relating to data, such as access to personal or commercial information?	**1**	Like privacy, ethics is important in relation to the reuse of data but compliance with ethical considerations does not greatly strengthen a policy as it relates to FAIR and open data.
14	Intellectual property	How is intellectual property handled in the policy or is there any reference to intellectual property?	**1**	IP is like ethics and privacy references. It can potentially add some value but does not significantly strengthen compliance with open and FAIR principles.
15	Licensing/copyright	What aspects of licensing or copyright are referenced and what forms of licence are mentioned?	**1**	This is considered similar in value to the above-mentioned criteria of ethics, privacy and intellectual property, for similar reasons.

Table 1 *– Ratings and weighting descriptions for each analysis criterion*

What do the criteria summaries, ratings, weightings and scores mean?
The overview of the documents and the summaries of each specific criterion aim to provide some justification and basis for how the evaluation scores were determined. The key element of this analysis is the numerical ratings given to each of the 15 criteria. These can be used to provide a direct comparison between each document (comparing compliance with each criterion) in relation to alignment and non-alignment. In other words, when the ratings for the same criterion differ between documents, there is a level of non-alignment between these documents.

If a criterion has a weighting of 5 and a rating of 3, giving a score of 15, this suggests full compliance with that criterion. If another document gets a rating of 1 for the same criterion, with a weighting of 5, compliance with that criterion is minimal.

When the ratings are multiplied by the weightings, the significance of any variation between documents for that specific criterion can help identify the "issue" or level of risk for the document in terms of each analysis criterion.

NOTE: *It is important to note that the scores given in this analysis are subjective. There are many reasons for this, but a key reason is that the descriptions of the various criteria elements in each of the documents have different wording and a different focus. Additionally, some documents have a lot of content about a relevant criterion, but the details are not very clear, while other documents have short and simple – but very clear and unambiguous – statements supporting a specific criterion. All of the documents reviewed also have different orientations, focuses and purposes, depending on the role of the organization "owning" the document and the specific purpose of that document. For these reasons, an accurate empirical analysis is not considered possible.*

4. Detailed review of documents

Excellence in Agronomy (EiA) Proposal

Overview
The EiA Proposal document is very strongly focussed on ensuring that research-generated digital assets are widely shared. Reference is made to FAIR and open data in various sections of the document. In particular, Section 3.2.3, "Work Package 2 – TRANSFORM", provides a wide-ranging perspective on what is required by researchers to meet the EiA requirements in the context of FAIR and open data. This section presents a very effective graphic showing the pathways through the research activity, and their alignment with the various outputs in Work Package 2 – TRANSFORM. This work package is aimed at "enabling the creation of value from improved data and advanced analytics".

Since the document is considered a "guideline" it will probably have great significance to researchers, who will need to address these requirements in order to receive funding. The guideline provides references to additional detail in the EiA Open and FAIR Data Assets Policy, and refers to various CGIAR data-related materials, such as metadata schemas and licensing conditions.

Compliance with the desk study analysis criteria

Criterion 1 – General data sharing philosophy
Criterion summary: The document is very clear on the need to make all research-generated digital assets available for reuse to support "agronomy-at-scale research". It provides targets in the context of KPIs for various outcomes that will leverage the digital assets created through the research activities. The outcomes of, and the general philosophy of, FAIR and open data principles are strongly emphasized throughout much of the document.

The analysis and scoring are based on the overall focus of the document in relation to data sharing concepts. The overall approach to data sharing is stressed across the whole document, and is specifically referred to in the following sections:

- Section 2.1, Challenge Statement, paragraphs 3 and 4
- Section 2.2, Measurable three-year outcomes, particularly Outcome 1 and Outcome 2.

- Section 2.3, Learning from prior evaluations and Impact Assessments, particularly paragraphs 1 and 3.
- Section 3.1.2, Full Initiative Theory of Change narrative, paragraph 4.
- Section 3.2.3, Sketch of the TRANSFORM Work Package – many sections in 3.2.3 provide detail of FAIR and open data and the sharing of digital assets to underpin the various use cases, such as climate impacts, sustainability etc.
- Section 8.2, Open and FAIR data assets – dot points 1 to 4 and paragraph 2.
    - *Rating* – 3
    - *Weighting* – 5
    - *Score* – 15

Criterion 2 – FAIR core principles (*Findable*, *Accessible*, *Interoperable*, *Reusable*)
Criterion summary: The document very specifically and frequently promotes the application of FAIR principles, with multiple references to the application of FAIR in numerous sections. Section 8.2 is titled "Open and FAIR data assets" and contains strong wording on the need for researchers to adhere to the CGIAR Open and FAIR Data Assets Policy. Reference is also made to various tools and other support for ensuring FAIR compliance.

Analysis and scoring are based on FAIR principles references found in the following:

- Page 5 – Section 2.2, Measurable three-year outcomes, Outcome 2.
- Page 14 – Section 3.1.2, Full initiative ToC narrative, paragraph 4.
- Page 19 – Section 3.2.4, Work Package TRANSFORM, The Science – paragraphs 1, 2, and 3.
- Page 20 – Section 3.2.4, Work Package TRANSFORM, The Theory of Change, paragraphs 2 and 5.
- Page 35 – Section 6.1, Results Framework – End of Initiative Outcome 2 and Output 2.1.
- Page 44 – Section 8.2, Open and FAIR data assets.
- Page 45 – Section 9.1, Human Resources – TRANSFOM Work Package (Research).

    - *Rating – 3*
    - *Weighting – 5*
    - *Score – 15*

Criterion 3 – Open data

Criterion summary: There are multiple references to "open" data and usually these references are associated with FAIR. The document is very strong on the requirement to ensure research data is open.

Analysis and scoring are based on open data references that occur throughout the document, with some of the specific references provided below:

- Page 5 – Section 2.2, Measurable three-year outcomes, Outcome 2.
- Page 14 – Section 3.1.2, Full initiative ToC narrative, paragraph 4.
- Page 19 – Section 3.2.4, Work Package TRANSFORM, The Science, paragraphs 1, 2, and 3.
- Page 20 – Section 3.2.4, Work Package TRANSFORM, The Theory of Change, paragraphs 2 and 5.
- Page 35 – Section 6.1, Results Framework – End of Initiative Outcome 2 and Output 2.1.
- Page 44 – Section 8.2, Open and FAIR data assets.
- Page 45 Section 9.1, Human Resources – TRANSFOM Work Package (Research).

    - *Rating – 3*
    - *Weighting – 5*
    - *Score – 15*

Criterion 4 – Data management

Criterion summary: There are numerous references to data management in general and the use of effective data management and advancing data management practices in particular. Data management is seen as a key objective of Work Package 2 – TRANSFORM. This work package aims to answer the question whether "improved data management (can) facilitate the development and scaling of solutions for agronomic gain". Additionally, there is recognition that "improved data science capacity and culture is a core driver for CGIAR and NARS to realize the benefits of agronomy R&D at scale". A further reference to the significance of data management is found in Section 9, "Human Resources", where reference is made to the need for data management and data standards skills to assist with moving research activity towards open and FAIR outcomes.

Analysis and scoring are based on data management-related references found in the following parts of the document:

- Page 6 – Section 2.3, Learning from prior evaluations and Impact Assessments, paragraph 1.
- Page 19 – Section 3.2.4, Work Package TRANSFORM, paragraphs 2 and 3.
- Page 20 – paragraph 5.
- Page 35 – Results Framework – TRANSFORM Work Package Outcome 2.
- Page 45 – TRANSFORM Work Package (under section on research)

    - *Rating – 3*
    - *Weighting – 5*
    - *Score – 15*

Criterion 5 – Metadata

Criterion summary: There is only limited reference to metadata. However, there is a compliance requirement for researchers to "adhere to the terms of the CGIAR Open and FAIR Data Assets Policy", with "rich metadata conforming to the CGIAR Core Schema". It is considered that while it is not widely mentioned in the document, the strength of the single reference should result in a high score for this criterion.

Analysis and scoring are based on metadata references found in the following:

- Page 44 – Section 8.2, Open and FAIR data assets – Point 1.

    - *Rating – 3*
    - *Weighting – 5*
    - *Score – 15*

Criterion 6 – Data governance

Criterion summary: The document identifies the value of data governance and suggests that End of Outcome 2 requires standardized, open and FAIR data and analytics, supported by effective data governance. However, there is no specific requirement to implement data governance frameworks, other than a recognition of their value in the data management chain.

Analysis and scoring are based on data governance references found in the following:

- Page 14 – Section 3.1.2, Full initiative ToC narrative, paragraph 4.

- *Rating – 3*
- *Weighting – 4*
- *Score – 12*

Criterion 7 – Policy compliance requirements
Criterion summary: Section 8 covers compliance required by the researcher in some detail and makes reference to various other policy documents, such as the CGIAR Research Ethics Code, CGIAR's Ethics Framework, and the Framework for Gender, Diversity and Inclusion in CGIAR's workplaces.

Analysis and scoring are based on references to policy compliance found in the following:

- Page 44 – Section 8, Policy compliance, and oversight

  - Rating – 3
  - Weighting – 4
  - Score – 12

Criterion 8 – Incentives to encourage data sharing.
Criterion summary: Incentives are referred to frequently in the document as a mechanism to encourage data sharing and also to encourage the acceptance of agronomic advice. The concept of evaluating the role and appropriateness of incentives is also discussed. Incentives are seen as important within the document.

Analysis and scoring are based on incentives references found in the following:

- Page 6 – Section 2.3, Learning from prior evaluations and Impact Assessments, paragraph 3.
- Page 16 – Section 3.2.2, Work Package DELIVER, The Science, paragraph 4 – (this paragraph is focussed on the role of incentives in accepting agronomic advice).
- Page 19 – Section 3.2.4, Work Package TRANSFORM, The Science, paragraph 1.
- Page 23 – Section 3.2.6, Work Package INNOVATE, The Theory of Change, paragraph 4.
- Page 28 – Section 5.2, Poverty reduction, livelihoods and jobs, paragraph 2 – Research questions.
- Page 29 – Section 5.3, Gender equality, youth and social inclusion, paragraph 2 – Research questions.

- Page 39 – Section 6.2, MELIA plan, paragraph 3.
- Page 40 – Section 6.3, Planned Monitoring, Evaluation, Learning and Impact Assessment (MELIA) studies and activities – Table items 1 and 3.
- Page 43 – Section 7.3, Risk assessment – Risk no. 4.

    - Rating – 3
    - Weighting – 3
    - Score – 9

Criterion 9 – Compliance with standards (including data formats, metadata structures and other relevant standards)
Criterion summary: There are frequent references to standards throughout the document. Standards are referenced in the context of data set compliance, semantics, data collection processes, ethics and scientific processes covering quality, safety, privacy, risk, and financial management. The need to apply appropriate standards is considered to be expressed very strongly in the document.

Analysis and scoring are based on references to compliance with standards found in the following:

- Page 19 – Section 3.2.4, Work Package TRANSFORM, The Science, paragraph 2.
- Page 20 – Section 3.2.4, Work Package TRANSFORM, The Theory of Change, paragraphs 2, 3, 4 and 5.
- Page 44 - Section 8.1, Research governance, paragraph 1.
- Page 45 – Section 9.1, Initiative team, under TRANSFORM Research.

    - *Rating – 3*
    - *Weighting – 3*
    - *Score – 9*

Criterion 10 – Breadth of digital assets (i.e. does the policy go beyond just data and also include other digital assets?)
Criterion summary: While the document does not appear to specifically define digital assets, other than referring to data, it does refer to data assets in the context of the OFDA Policy's open and FAIR requirements, and it also refers to software as an asset.

Analysis and scoring are based on the types of digital assets referred to in the document, with references found in the following:

- Page 44, Section 8.2, Open and FAIR data assets – bullet points 1 to 4 and paragraph 2.

    - *Rating – 3*
    - *Weighting – 2*
    - *Score – 6*

Criterion 11 – Geographic coverage

Criterion summary: In Section 2, in the proposal structure summary table, the geographic scope is stated as being "global". There are many other references to "global" throughout the document, although Section 2.4, Priority-setting (para. 4), indicates that while EiA's scope is global the initial priorities will be directed to specific regions. The four work packages all state "global" in the work package geographic scope.

Analysis and scoring are based on geographic coverage references found in the following:

The document contains multiple references to the geographic coverage being global, and these include the following:

- Page 3 – Section 2, Proposal Structure – Summary Table – geographic scope.
- Page 16 – Section 3.2.2, Work Package 1 – DELIVERY – geographic scope.
- Page 19 – Section 3.2.4, Work Package 2 - TRANSFORM - geographic scope.
- Page 22 – Section 3.2.6, Work Package 3 – INNOVATE – geographic scope.
- Page 25 – Section 3.2.8, Work Package 4 – ORGANIZE – geographic scope.

- *Rating – 3*
- *Weighting – 1*
- *Score – 3*

Criterion 12 – Privacy

Criterion summary: While privacy is only mentioned a few times in the document it is in reference to several relevant privacy concerns and the

context is one of requiring the researchers' compliance. The document refers to appropriate standards and ethics in relation to privacy.

Analysis and scoring are based on privacy references found in the following:

- Page 44 – Section 8.1, Research governance, paragraph 1.
- Page 44 – Section 8.2, Open and FAIR data assets, paragraph 3.

    - Rating – 3
    - Weighting – 1
    - Score – 3

Criterion 13 – Ethics

Criterion summary: Ethics is referred to in the context of research governance and the document requires researchers to comply with CGIAR's Research Ethics Code, the Ethics Framework and the Framework for Gender, Diversity, and Inclusion in CGIAR's workplaces. It is considered that ethics is strongly encouraged in the document, although detail on what this entails is provided by the CGIAR documents referred to above.

Analysis and scoring are based on ethics references found in the following:

- Page 44 – Section 8.1, Research governance, paragraph 1.
- Page 44 – Section 8.2, Open and FAIR data assets, paragraph 2, item 4.

    - Rating – 3
    - Weighting – 1
    - Score – 3

Criterion 14 – Intellectual property

Criterion summary: Intellectual property is not mentioned in the document. This is likely due to the requirement to make all digital assets open. There is also no reference to potential commercialization of digital assets, for the same reason. With no views on commercialization there is therefore no need to "manage intellectual property", other than through appropriate open licences such as Creative Commons and General Public Licence.

    - Rating – 1
    - Weighting – 1
    - Score – 1

Criterion 15 – Licensing/copyright

Criterion summary: Licensing is referred to in the context of open digital assets and there is a requirement that it be covered by Creative Commons and General Public Licences, as appropriate. As above, this is due to the requirement of ensuring that all digital assets are open.

- Rating – 3
- Weighting – 1
- Score – 3

Total document score: 136

CGIAR Open and FAIR Data Assets Policy

Overview
The CGIAR Open and FAIR Data Assets Policy provides a very clear view of CGIAR's expectations and requirements in regard to considering all digital assets as international public goods, encouraging their widespread dissemination and reuse. Open access is considered the approach to be adopted to meet this requirement and the application of FAIR principles is seen as the way to support and underpin open access of quality, usable data. The policy does acknowledge that sharing digital assets may not always be possible but says that data should always be compliant with FAIR principles. This is a strong message about effective data management. The policy identifies that with the goal of sharing all digital assets in a useful form, the first step is to have the data effectively managed. It recognizes the need to have in place the application of FAIR principles, regardless of whether the data will be shared.

The policy describes why open access and FAIR is important, giving a perspective on the considerable benefits that arise from complying with both of these principles.

Compliance with the desk study analysis criteria

Criterion 1 – General data sharing philosophy
Criterion summary: The policy has a very strong focus on the overall philosophy of making data and other digital assets available for reuse and uses the concept of open data as the vehicle for achieving this objective. It sees the results of its research and development activities as international public goods. Considerable justification and reasoning are provided to

support the value of this approach as CGIAR explains that it aims to leverage these research outcomes, infrastructure and data science initiatives to support innovation and to provide effective and agile responses to global challenges.

The policy clearly recognizes the importance of data management and sharing as a key pillar in achieving its objectives.

Analysis and scoring are based on the overall focus of the document in relation to FAIR and open data concepts, and references are found in the following:

- Page 3 – Section 1, CGIAR's Commitment, subsections 1.1, 1.2 and 1.3.
- Page 3 – Section 2, Purpose, paragraph 1.
- Page 4 – Section 4, Policy, most of subsection 4.1 and on Page 6 – subsection 4.2.

    - *Rating* – 3
    - *Weighting* – 5
    - *Score* – 15

Criterion 2 – FAIR core principles (*Findable*, *Accessible*, *Interoperable*, *Reusable)*
Criterion summary: The policy has FAIR in its title and states that data assets "shall be managed in compliance with FAIR Data Principles". It provides considerable detail about findability, accessibility, interoperability and reusability. There is a very strong focus on the FAIR principles, with a requirement for data assets to "always (be) FAIR".

Analysis and scoring are based on FAIR principles references found in the following:

- Page 3 – Section 1, CGIAR's Commitment, subsections 1.1 and 1.2.
- Page 3 – Section 2, Purpose.
- Page 4 – Section 4, Policy, subsection 4.1 is entirely about FAIR principles.
- Page 6 – Section 5, Implementation, subsections 5.1, 5.2 (b and c).
- Page 8 – Annex 2, subsections I, ii, iii, and iv.

    - Rating – 3
    - Weighting – 5

- Score – 15

Criterion 3 – Open data

Criterion summary: The policy includes the term "open" in its title and defines the concept of open in an annex. There is a very strong focus on making CGIAR's research outputs as open as possible. For justifiable reasons, the policy does not "mandate" open in the same manner as it supports FAIR. There is obvious recognition that for many legitimate reasons not all data can be shared, but it does need to be well-managed (FAIR-compliant), regardless of whether it can be shared.

Analysis and scoring are based on open data references that occur throughout the document, with some specific references provided in the following parts of the document:

- Page 1 – Policy title.
- Page 3 – Section 1, CGIAR's Commitment, subsections 1.1 and 1.2.
- Page 3 – Section 2, Purpose.
- Page 4 – Section 4, Policy, subsection 4.1 (a) ii.
- Page 5 – Section 4, Policy, subsection 4.1 (b) i, ii, and iii.
- Page 5 – Section 4, Policy, subsection 4.1 (d) ii.
- Page 6 – Section 5, Implementation, subsections 5.1, 5.2 b and c.
- Page 8 – Annex 2, subsections i, ii, iii, iv, v and vi.

  - Rating - 3
  - Weighting – 5
  - Score – 15

Criterion 4 – Data management

Criterion summary: The policy is focussed on effective data management, specifically within the context of making data FAIR-compliant. The policy also addresses the need to have data management plans.

Data management is referred to in the context of the Global Access Commitments Agreement with partners. However, there is no specific mention of the need for effective data management to facilitate reuse within the Global Access Statement.

Analysis and scoring are based on data management references found in the following:

- Page 3 – Footnote 1.

- Page 4 – Section 4, Policy, subsection 4.1 a(v).
- Page 5 – Section 4, Policy, subsection 4.1 c (ii).
- Page 6 – Section 5, Implementation, subsection 5.2 (c).
- Page 7 – Annex 1, Relevant Defined Terms – Section vi – Data management.

    - Rating – 3
    - Weighting – 5
    - Score – 15

Criterion 5 – Metadata

Criterion summary: Metadata is referred to frequently within the policy and many of the references are in the context of describing the FAIR principles. Metadata is seen as a key element in the policy, to the point where it is considered one of the specific data assets covered by the policy. Reference is also made to having metadata describing data assets conforming to the CGIAR Core metadata schema.

Analysis and scoring are based on metadata references found in the following:

- Page 3 – Section 1, CGIAR's Commitment, Subsection 1.2.
- Page 4 – Section 3, Scope, subsection 3.2.
- Page 4 – Section 4, Policy, subsection 4.1 (a) I, ii, iii and iv.
- Page 5 – Section 4, Policy, subsection 4.1 (b) I, ii and v.
- Page 5 – Section 4, Policy, subsection 4.1I) i.
- Page 5 – Section 4, Policy, subsection 4.1 (d) iii.
- Page 7 – Annex 1, subsection xi.
- Page 8 – Annex 2, subsection vii.

    - *Rating – 3*
    - *Weighting – 5*
    - *Score – 15*

Criterion 6 – Data governance

Criterion summary: Data governance is not specifically mentioned; however, there are references to FAIR, and the "advanced maturity" levels discussed in relation to FAIR would suggest that data governance may be required in order to implement an effective project outcome and to achieve FAIR maturity.

Analysis and scoring are based on data governance references found in the following:

- There are no specific references to data governance, although this can be inferred.

    - *Rating – 2*
    - *Weighting – 4*
    - *Score – 8*

Criterion 7 – Policy compliance requirements
Criterion summary: The policy refers to the need for data assets to comply with the FAIR and open data principles. However, there is no other statement on compliance or what consequences there are for non-compliance.

Analysis and scoring are based on references to policy compliance found in the following:

- Section 3, Policy, subsection 7.

    - Rating – 2
    - Weighting – 4
    - Score – 8

Criterion 8 – Incentives to encourage data sharing
Criterion summary: The policy does not refer to any incentives to facilitate and encourage data sharing. Given that CGIAR's focus is on its research outputs being considered international public goods, this perhaps implies that data sharing incentives are not required.

Analysis and scoring are based on incentives references found in the following:

- There are no references to incentives to share data in the policy.

    - Rating – 1
    - Weighting – 3
    - Score – 3

Criterion 9 – Compliance with standards (including data formats, metadata structures and other relevant standards)
Criterion summary: The policy makes references to metadata standards and to interoperability standards, such as ontologies and controlled vocabularies.

While data formats are not referred to specifically in the context of standards, the focus on "accepted" interoperability-based standards would imply that the use of international (or national) data format standards is considered by the policy as important in order to facilitate interoperability.

Analysis and scoring are based on references to compliance with standards found in the following:

- Page 5 – Section 4, Policy, subsection 4.1 (c) ii.

    - *Rating* – 3
    - *Weighting* – 3
    - *Score* – 9

Criterion 10 – Breadth of digital assets (i.e. does the policy go beyond just data and also include other digital assets?)
Criterion summary: The policy refers to an extensive range of digital assets that are covered by the policy, including (but not limited to) journal articles, reports, training content, data and databases, analysis tools, GIS outputs and computer software, models, mobile applications, and web-based services.

Analysis and scoring are based on the types of digital assets referred to in the document and these are found in the following:

- Page 7 – Annex 1, Relevant Defined Terms – Section i – Data assets.

    - *Rating* – 3
    - *Weighting* – 2
    - *Score* – 6

Criterion 11 – Geographic coverage
Criterion summary: The policy refers to the "global public" and responding to "global challenges". While it does not specifically state that its policy has global coverage, it strongly implies this, and this is reinforced by the concept of the digital assets being an international public good. The policy does identify key stakeholders who will benefit from the application of the policy, namely "the poor, especially smallholder farmers in developing countries".

Analysis and scoring are based on geographic coverage references found in the following:

- Page 3 – Section 1, CGIAR's Commitment, sections 1.1 and 1.2.
- Page 7 – Annex 1, Relevant Defined Terms – xii – Open Access.

- *Rating* – 3
- *Weighting* – 1
- *Score* – 3

Criterion 12 – Privacy
Criterion summary: The requirement to manage data assets responsibly is described, with mention of both privacy and ethical approaches to data management and accessibility.

Analysis and scoring are based on privacy references found in the following:

- Page 3, CGIAR's Commitment, subsection 1.3.
- Page 6 – Section 4, Policy, subsection 4.2.

  - Rating – 3
  - Weighting – 1
  - Score – 3

Criterion 13 – Ethics
Criterion summary: Reference is made to the CGIAR Research Ethics Code and also to the fact that CGIAR is committed to managing data assets responsibly in terms of both privacy and ethical approaches.

Analysis and scoring are based on ethics references found in the following:

- There are several mentions of the CGIAR Research Ethics Code throughout the policy document:
  - Page 3 – Section 1, CGIAR's Commitment, subsection 1.3.
  - Page 6 – Section 4, Policy, subsection 4.2.

  - Rating – 3
  - Weighting – 1
  - Score – 3

Criterion 14 – Intellectual property
Criterion summary: There is no specific reference to intellectual property, but the policy does refer to the CGIAR Principles on the Management of Intellectual Assets, and this is in the context of limited exclusivity agreements.

Analysis and scoring are based on intellectual property references found in the following:

- There is no specific reference to intellectual property in the policy, other than the reference to the CGIAR Principles on the Management of Intellectual Assets.

    - Rating – 1
    - Weighting – 1
    - Score – 1

Criterion 15 – Licensing/copyright
Criterion summary: Reference is made to licensing in relation to third-party software that "may" be made available under appropriate licensing arrangements.

Analysis and scoring are based on references to licences and copyright found in the following:

- Page 8 – Annex 2, subsection vi.

    - Rating – 1
    - Weighting – 1
    - Score – 1

Total score: 120

EiA Best Practice Guidelines to Support Global Access Implementation

Overview
The EiA Best Practice Guidelines to Support Global Access Implementation is a very comprehensive document covering open data (and based on the FAIR principles). It effectively covers many of the aspects that support, and have an impact on, data sharing. The general philosophy of this document is to make **all** data (digital assets) available, supporting global access. It also explains in detail the range of conditions and exceptions for not releasing data. This is a realistic approach given various countries' legislation and other conditions that exist around national data, as well as acknowledging commercial interests. Additional detail is provided, such as templates and examples to assist research centres and researchers to apply the guidelines on how to effectively share data and how to proceed with engagement with the private sector when appropriate. The document takes into consideration

centre-level policies and best practice in relation to engaging with the private sector.

The document refers to a range of relevant FAIR and open data elements and provides links to more detailed information that provides additional details on these, such as metadata and CGIAR-relevant standards.

Compliance with the desk study analysis criteria

Criterion 1 – General data sharing philosophy
Criterion summary: The document is very clear on making "all knowledge and information from EiA funded projects broadly disseminated" and accessible at affordable prices to those most in need. Many aspects of the document are closely aligned with the Bill & Melinda Gates Foundation Open Access Policy and the CGIAR Open and FAIR Data Assets Policy.

- The overall approach to data sharing is stressed across the whole document, and specifically in the following sections:
  Section 1, paragraphs 1, 2, 3 and 4.
  Part A: 1.

  - Rating – 3
  - Weighting – 5
  - Score – 15

Criterion 2 – FAIR core principles (*Findable*, *Accessible*, *Interoperable*, *Reusable)*
Criterion summary: The document broadly promotes the FAIR concept, and Annex 4A in particular describes a FAIR Compliance Model adopted from the Netherlands Institute for Permanent Access to Digital Resources (DANS) metrics.[57,58]

Analysis and scoring are based on FAIR principles references found in the following:

- Footnote 18.
- Section A1, paragraph 2.

[57] Doorns, P. (2017) "Data Archiving and Networked Services, How FAIR Am I?". Plan-Europe Fhttps://planeurope.files.wordpress.com/2015/03/doorn-fair-interoperability-poznan-plan-e-april-2017.pdf
[58] AIMS. "Put FAIR Principles into Practice and Enjoy Your Data!" Accessed 24 November 2023. https://aims.fao.org/news/put-fair-principles-practice-and-enjoy-your-data.

- Annex A4, including footnotes 1, 2 and 3 in this annex.

    - Rating – 3
    - Weighting – 5
    - Score – 15

Criterion 3 – Open data

Criterion summary: The document provides its own description of open data, focussing on open access to digital assets by depositing these assets in relevant open repositories.

Analysis and scoring are based on open data references are found in the following:

- Section 1, Purpose, paragraphs 2 and 3.
- Part A.1, included in the section heading and in paragraphs 1, 2 and 3, plus paragraph 1, dot points 1 and 4.
- Section A2, included in section heading and paragraphs 1 and 2.

    - Rating – 3
    - Weighting – 5
    - Score – 15

Criterion 4 – Data management

Criterion summary: Data management is referred to in the context of relevant CGIAR policies, rather than specifically in the EiA Implementation Guidelines themselves. Given the CGIAR references, it is considered that data management is seen to be essential.

Analysis and scoring are based on data management references found in the following:

Section 1, Purpose, paragraph 3 and footnotes 4 and 6.
There are other references to data management and CGIAR policies in other sections of the EiA Implementation Guidelines.

- Rating – 3
- Weighting – 5
- Score – 15

Criterion 5 – Metadata

Criterion summary: Details of the CGIAR Core Metadata Schema and relevant links, as well as details of data dictionaries, including ontologies and controlled vocabularies, are provided in many parts of the document to support interoperability.

Analysis and scoring are based on metadata references found in the following:

- Section A1, footnote 10 – CGIAR Open Access and Data Management Policy
  Section A1, dot point 3 contains numerous references to metadata, including metadata schemas.
- Footnote 26 refers to CGIAR's Core Metadata Schema and the Dublin Core Metadata Schema.
- Annex A2 refers to the CGIAR Metadata Schema Reference Guide and provides details on this guide.

    - Rating – 3
    - Weighting – 5
    - Score – 15

Criterion 6 – Data governance

Criterion summary: Governance is referred to in the document but only in the context of organizational governance, not specifically data governance.

Analysis and scoring are based on data governance references found in the following:

- No references were found.

    - Rating – 1
    - Weighting – 4
    - Score – 4

Criterion 7 – Policy compliance requirements

Criterion summary: There is no comment on the need for researchers to comply with these guidelines other than being aware of and complying with the relevant CGIAR obligations, in particular the access and benefit sharing obligations arising from the Nagoya Protocol of the Convention on Biological Diversity. As explained above, it can be considered that the objective of EiA is to support research outcomes as international public goods and so the document possibly does not need to make a specific reference to compliance.

Analysis and scoring are based on references to policy compliance found in the following:

Part B, Public–Private Partnerships, under the section heading "Complying with the CGIAR Intellectual Assets Principle".

- A.3, dot points 1 and 6.

  - Rating – 1
  - Weighting – 4
  - Score – 4

Criterion 8 – Incentives to encourage data sharing
Criterion summary: Incentives are referred to in the context of partnerships with the private sector but there are no specific incentives to encourage researchers to share data. However, EiA's overall philosophy is to make the data accessible and globally available, so it could be considered that mention of "incentives" is not required and there should simply be compliance with the data sharing requirement.

Analysis and scoring are based on incentive references found in the following:

- B.4, first paragraph, but in the context of developing partnerships with the private sector.

  - Rating – 2
  - Weighting – 3
  - Score – 6

Criterion 9 – Compliance with standards (including data formats, metadata structures and other relevant standards)
Criterion summary: There are several references to relevant CGIAR standards, including metadata, ethics, due diligence, open data and interoperability standards, throughout the document.

Analysis and scoring are based on references to compliance with relevant standards found in the following:

- Part A.1, dot points 2 and 5.
- Annex A2, the first paragraph on the CGIAR Core metadata reference guide.

- Rating – 3
- Weighting – 3
- Score – 9

Criterion 10 – Breadth of digital assets (i.e. does the policy go beyond just data and also include other digital assets?)
Criterion summary: The coverage includes data and information products as defined by the CGIAR Open Access and Data Management Policy, and covers a broad range of digital materials, beyond simply data. It includes such digital items as papers, audio, images, computer software and web services.

Analysis and scoring are based on the types of digital assets referred to in the document and references to these are found in the following:

- Section A.1, paragraph 1, including the first dot point.
- Reference is made to what constitutes data and information (digital assets) in the CGIAR Open Access and Data Management Policy.
  Footnotes 9 and 10.

  - Rating – 3
  - Weighting – 2
  - Score – 6

Criterion 11 – Geographic coverage
Criterion summary: Global sharing is in the title of the document and the global coverage required by the guidelines is referred to many times throughout the document.

Analysis and scoring are based on geographic coverage references found in the following:

The document contains multiple references to the geographic coverage being global, including in the title of the document.

- Rating – 3
- Weighting – 1
- Score – 3

Criterion 12 – Privacy

Criterion summary: The concept of privacy and issues around privacy and confidentiality, such as personally identifiable information, is well-described, and there are links to further detailed information.

Analysis and scoring are based on privacy references found in the following:

- Part A. A1, the first dot point.
- Footnote 11.
- Paragraph 3.
- A2, paragraph 1.
- A2, dot point 5.
- A3, dot point 5.
- B2, Generic Risks and Challenges, dot point 5.

    - Rating – 3
    - Weighting – 1
    - Score – 3

Criterion 13 – Ethics

Criterion summary: Ethics is referred to in the context of centres' research responsibilities, with only limited emphasis on the ethics around the sharing of data.

Analysis and scoring are based on ethics references found in the following:

- Section1, paragraph 3.
- Footnote 30.
- Annex A6, Part c, General Terms and Conditions, paragraph 6.

    - Rating – 2
    - Rating – 1
    - Score – 2

Criterion 14 – Intellectual property

Criterion summary: Intellectual property is included in some detail in the document in relation to contractual rights over data, and especially in the context of dealing with the private sector.

Analysis and scoring are based on intellectual property references found in the following:

- Section 1, Purpose, Footnote 2.
- Section A1, first dot point

- Section A2, first paragraph.
- Section A3, fourth dot point.
- Annex A6, Part C, paragraphs 3 and 6.
  Annex B2, Intellectual Property Rights Section (covers a number of intellectual property aspects).

    - Rating – 3
    - Weighting – 1
    - Score – 3

Criterion 15 – Licensing/copyright
Criterion summary: There is considerable detail in this document concerning open licensing with CC-BY for publications and MIT or GNU Public Licences for software.

Analysis and scoring are based on licensing and copyright references found in:

- Part A, A1, fourth dot point.
- Footnote 24.
- Part B, B2 (Generic Risks and Challenges), dot point 2.
- Part B, B4, in the paragraph describing dot point 3.
- Annex A6, Part B, paragraph 5ii.

    - Rating – 3
    - Weighting – 1
    - Score – 3

Total score: 118

Bill & Melinda Gates Foundation – Data Sharing Requirements

Overview
This document focusses the various approaches and mechanisms that should be followed to comply with the foundation's data sharing requirements. It aims to apply best practice approaches to making data sharing possible, including strong references to FAIR and open data principles. As a stand-alone document it provides details on the approaches to be used to ensure that research-generated data is available for reuse and that sufficient structures and metadata etc. are in place to ensure its usefulness to others.

This document provides instructions on what is expected of researchers in the context of ensuring that others can use their data.

NOTE: There is no formal numbering of the sections in this document. For this study, numbers have been allocated based on each consecutive section and its subsections.

Compliance with the desk study analysis criteria

Criterion 1 – General data sharing philosophy
Criterion summary: The document is focussed on the outcome of having research-generated data available for secondary users. It promotes mechanisms to ensure that the data is both FAIR-compliant and unencumbered by constraints on reuse, except where some specific restrictions are legitimately applied, such as privacy issues.

Analysis and scoring are based on the overall focus of the document in relation to data sharing concepts and references are found in the following:

- Most sections of the document provide information on ensuring that research-generated digital assets are sharable.

    - Rating – 3
    - Weighting – 5
    - Score – 15

Criterion 2 – FAIR core principles (*Findable*, *Accessible*, *Interoperable*, *Reusable)*
Criterion summary: The FAIR data principles are described in some detail and there are links to additional information about FAIR.

Analysis and scoring are based on FAIR principles references found in the following:

- Section 7, What is FAIR data? Paragraphs 1 and 2.
- Section 8, How do FAIR and Open Data differ? Paragraphs 1 and 3.
- Section 9, How can grantees meet FAIR Data Principles – first paragraph.
- Section 10, What resources are available to help make data FAIR, paragraph 1 and dot points 1, 2, 3 and 5.

    - Rating – 3

- Weighting – 5
- Score – 15

Criterion 3 – Open data

Criterion summary: The document is aimed at defining the requirements for ensuring that research data is open, and so it is strongly focussed on "all" data being available through open mechanisms, with limited restrictions applying only in specific and justifiable cases.

Analysis and scoring are based on open data references that occur throughout the document, with some of the specific references provided below:

- Section 1, What are the requirements for Open Access Policy source data? Paragraphs 1, 3 and 4.
- Section 5, How should grantees make data accessible and open? Dot point 1.
- Section 6, Where should grantees deposit data? Paragraph 2 and paragraph 4 (dot point 2).
- Section 8, How do FAIR and Open Data differ? Paragraphs 1 and 3.
- Section 9, How can grantees meet FAIR Data Principles? Subsection 1 – Findable, paragraph 1.
- Section 9, How can grantees meet FAIR Data Principles, subsection 2 – Accessible, paragraph 1, and Grantee Actions, dot point 1.
- Section 9, How can grantees meet FAIR Data Principles, subsection 3 – Interoperability, paragraph 1.
- Section 10, What resources are available to help make data FAIR, dot point 1.

  - Rating – 3
  - Weighting – 5
  - Score – 15

Criterion 4 – Data management

Criterion summary: Data management is not specifically discussed but there is reference to a number of data management activities, such as data collection, analysis and storage. Additionally, FAIR data management, and a description of what is required to ensure data is FAIR-compliant, including details of metadata, also indicate the view that appropriate data management is required.

The document also refers to the need to store data in approved repositories, which also implies that a suitable level of data management must be applied to the data.

Analysis and scoring are based on data management references found in the following:

- Section 1, What are the requirements for Open Access Policy source data? Paragraph 2.
- Section 7, What is FAIR data? Paragraphs 1 and 2.
- Section 9, How can grantees meet FAIR Data Principles, all paragraphs.

    - Rating – 3
    - Weighting – 5
    - Score – 15

Criterion 5 – Metadata
Criterion summary – Metadata is referred to several times, especially in relation to the FAIR principles.

Analysis and scoring are based on metadata references found in the following:

- Section 2, What is underlying data? Paragraph 1.
- Section 7, What is FAIR data? Paragraph 2.
- Section 9, How can grantees meet FAIR Data Principles, subsection 1 – Findable, paragraph 1, and Grantee Actions.
- Section 9, How can grantees meet FAIR Data Principles, subsection 4 – Reusable, paragraph 1.
- Section 10, What resources are available to help make data FAIR, dot point 4.

    - Rating – 3
    - Weighting – 5
    - Score – 15

Criterion 6 – Data governance
Criterion summary: Data governance is not mentioned in the document.

Analysis and scoring are based on data governance references found in the following:

- There are no specific references to data governance.

    - Rating – 1
    - Weighting – 4
    - Score – 4

Criterion 7 – Policy compliance requirements
Criteria summary: There is no reference to compliance in the document. As with incentives, it is assumed that this is implied. Additionally, reference is made to a range of activities that are addressed at the individual grant and contract stage, and this may include compliance requirements.

Analysis and scoring are based on references to policy compliance found in the following:

- There is no specific reference to compliance.

    - Rating – 1
    - Weighting – 4
    - Score – 4

Criterion 8 – Incentives to encourage data sharing.
Criterion summary: There is no mention of any incentives to encourage the sharing of any form of research output. The assumption is made that all data must be shared and perhaps therefore no specific incentive to ensure this occurs is required.

Analysis and scoring are based on incentive references found in the following:

- There are no references to incentives to share data or research outputs in the policy.

    - Rating – 1
    - Weighting – 3
    - Score – 3

Criterion 9 – Compliance with standards (including data formats, metadata structures and other relevant standards)
Criterion summary: The document refers to various standards, such as disciplinary-specific standards, vocabularies, ISO trustworthy digital repository standards and metadata standards.

Analysis and scoring are based on references to compliance with standards found in the following:

- Section 1, What are the requirements for Open Access Policy source data? Paragraph 2.
- Section 5, How should grantees make data accessible and open? Dot point 5.
- Section 6, Where should grantees deposit data? Paragraph 1.
- Section 9, How can grantees meet FAIR Data Principles? Subsection 1 – Findable, paragraph 1.
- Section 9, How can grantees meet FAIR Data Principles? Subsection 3 – Interoperability, paragraph 1, and Grantees Actions, dot point 2.
- Section 10, What resources are available to help make data FAIR? Dot point 4.

    - Rating – 3
    - Weighting – 3
    - Score – 9

Criterion 10 – Breadth of digital assets (i.e. does the policy go beyond just data and also include other digital assets?)
Criterion summary: The guidelines do not provide a description of what is considered data, although they do refer to "all primary data, associated metadata, and any additional relevant data". They also point the reader to the Gates Open Research website for information on what data needs to be included. While the references are to "data" it could be implied that this refers to all relevant digital assets.

Analysis and scoring are based on the types of digital assets referred to in the document and references to these are found in the following:

- Section 1, What are the requirements for Open Access Policy source data? Paragraphs 1, 2 and 3.
- Section 2, What is underlying data? Paragraphs 1 and 2.

    - Rating – 3
    - Weighting – 2
    - Score – 6

Criterion 11 – Geographic coverage

Criterion summary: The document does not specifically refer to its geographic coverage, but it does suggest the use of a globally unique persistent identifier. There are no geographic restrictions identified in the document and so it is considered that the geographic extent is global.

Analysis and scoring are based on geographic coverage references found in the following:

- Section 9, How can grantees meet FAIR Data Principles – Findable, paragraph 1.

    - Rating – 3
    - Weighting – 1
    - Score – 3

Criterion 12 – Privacy

Criterion summary: The issue of privacy is not described specifically in the document other than through a note which states that it is not required to share data that is "ethically unsound or legally encumbered". It could be considered that privacy is recognized but no detail is provided.

Analysis and scoring are based on privacy references found in the following:

- Section 2, What is underlying data? Note.

    - Rating – 1
    - Weighting – 1
    - Score – 1

Criterion 13 – Ethics

Criterion summary: There is no specific reference to ethics, other than the reference made in the note in Section 2 that mentions data that is "ethically unsound".

Analysis and scoring are based on ethics references found in the following:

- Section 2, What is underlying data? Note.
    - Rating – 1
    - Weighting – 1
    - Score – 1

Criterion 14 – Intellectual property

Criterion summary: The document does not refer to intellectual property. This is possibly because, apart from some exceptions, all data is considered open.

Analysis and scoring are based on intellectual property-related references found in the following:

- There is no reference to intellectual property.

    - Rating – 1
    - Weighting – 1
    - Score – 1

Criterion 15 – Licensing/copyright

Criterion summary: Reference is made to Creative Commons and Open Source Initiative (OSI) licences but the document focusses on licences with little or no real restriction on the use of the data. Reference is also made to the use of open repositories.

Analysis and scoring are based on references to licences found in the following:

- Section 5, How should grantees make data accessible and open? Dot point 3.
- Section 9, How can grantees meet FAIR Data Principles? Various references to open licences in the context of making data openly published.

    - Rating – 3
    - Weighting – 1
    - Score – 3

Total document score: 110

<u>Bill & Melinda Gates Foundation – Data Guidelines</u>

Overview

The Bill & Melinda Gates Foundation Data Guidelines[59] provide a very detailed and comprehensive perspective on what the foundation seeks in terms of ensuring that data and other relevant digital assets (e.g. software) created during funded research activities are managed effectively. The aim of this is

[59] Gates Open Research. *Data Guidelines*. https://gatesopenresearch.org/for-authors/data-guidelines

to ensure that the data assets are in a suitable state that enables them to be reused or re-analysed with minimal difficulty. The guidelines support both the foundation's Open Data Policy and the FAIR data principles, with a number of links to detailed information about both open data and FAIR concepts, relevant policies and open repositories. The document also sets out a four-step process for preparing data so that it is sharable.

Compliance with the desk study analysis criteria

Criterion 1 – General data sharing philosophy
Criterion summary: As mentioned above, the Bill & Melinda Gates Foundation Data Guidelines provide a wide range of detail around general data management and data sharing under the FAIR and open data concepts. The guidelines also provide useful references to additional information to support the effective management and sharing of digital assets.

- Most sections of the document provide information on ensuring that research-generated digital assets are sharable.

    - Rating – 3
    - Weighting – 5
    - Score – 15

Criterion 2 – FAIR core principles (*Findable*, *Accessible*, *Interoperable*, *Reusable)*
Criterion summary: The FAIR data principles are described in some detail and there are links to additional information about FAIR.

Analysis and scoring are based on FAIR principles references found in the following:

- Section 1.2 FAIR Data Principles.

    - Rating – 3
    - Weighting – 5
    - Score – 15

Criterion 3 – Open data
Criterion summary: The guidelines provide strong references to the concept of open data, with the first sentence in the background section beginning "Gates Open Research…". There is a complete section on open data and a

recognition that not all data can be considered open, as described in the section on Exceptions.

Analysis and scoring are based on open data references that occur throughout the document, with some of the specific references provided below:

- Section 1, Background, paragraph 1.
- Section 1.2, Open Data Policy.
- Section 1.2, FAIR Data Principles.
- Section 2.2, Select a Repository.

    - Rating – 3
    - Weighting – 5
    - Score – 15

Criterion 4 – Data management

Criterion summary: Data management is not specifically discussed but there is reference to a "Data Management Plan" associated with the research activity. Additionally, effective data management could be implied by the references to FAIR, metadata and standards. The guidelines also refer to the need to store the data in approved repositories, also implying that a suitable level of data management must be applied to the data.

Analysis and scoring are based on data management references found in the following:

- Section 1, Background, paragraph 1.
- Section 1.2, FAIR Data Principles.
- Section 2.1, Prepare Your Data for Sharing, paragraph 1.
- Section 2.2, Select a Repository.

    - Rating – 2
    - Weighting – 5
    - Score – 10

Criterion 5 – Metadata

Criterion summary: Metadata is referred to a number of times, especially in relation to the FAIR principles.

Analysis and scoring are based on metadata references found in the following:

- Section 1.2, FAIR Data Principles, subsections on Findable, Interoperable and Reusable.

    - Rating – 3
    - Weighting – 5
    - Score – 15

Criterion 6 – Data governance
Criterion summary: Data governance is not mentioned in the document.

Analysis and scoring are based on data governance references found in the following:

- There are no references to data governance.

    - Rating – 1
    - Weighting – 4
    - Score – 4

Criterion 7 – Policy compliance requirements
Criteria summary: There is no reference to compliance with the guidelines.

Analysis and scoring are based on references to compliance found in the following:

- There is no reference to compliance.

    - Rating – 1
    - Weighting – 4
    - Score – 4

Criterion 8 – Incentives to encourage data sharing
Criterion summary: There is no mention of any incentives to encourage the sharing of any form of research output.

Analysis and scoring are based on incentive references found in the following:

- There are no references to incentives to share data or research outputs in the guidelines.

    - Rating – 1
    - Weighting – 3

- Score – 3

Criterion 9 – Compliance with standards (including data formats, metadata structures and other relevant standards)
Criterion summary: The guidelines recommend applying standards relevant to the specific topic of the research activity, if they exist. Reference is made to where topic-specific standards can be found.

Analysis and scoring are based on references to compliance with standards found in the following:

- Section 2.1, Prepare Your Data for Sharing.

  - Rating – 3
  - Weighting – 3
  - Score – 9

Criterion 10 – Breadth of digital assets (i.e. does the policy go beyond just data and also include other digital assets?)
Criterion summary: The guidelines focus specifically on data, although there is reference to the foundation's Data Policy, which does discuss data assets beyond simply data.

- Most paragraphs in the guidelines refer to data.
- Section 1, Background, paragraph 1 refers to the Data Policy.

  - Rating – 3
  - Weighting – 2
  - Score – 6

Criterion 11 – Geographic coverage
Criterion summary: The document does not specifically refer to its geographic coverage, but it does suggest the use of a globally unique persistent identifier. There are no geographic restrictions identified where the guidelines are not applicable, and so it is considered that the geographic extent is global.

Analysis and scoring are based on geographic coverage references found in the following:

- Section 1.2, FAIR Data Principles, Findable subsection, paragraph 1.

  - Rating – 3

- Weighting – 1
- Score – 3

Criterion 12 – Privacy
Criterion summary: The issue of privacy is described in the context of ensuring that any data relating to human participants is de-identified in accordance with the Safe Harbour method. Additionally, privacy considerations are referred to in regard to making software source code available.

Analysis and scoring are based on privacy references found in the following:

- Section 2.1, Prepare Your Data for Sharing, paragraph 2.
- Section 2.1.2, Software source code, paragraph 2.

  - Rating – 3
  - Weighting – 1
  - Score – 3

Criterion 13 – Ethics
Criterion summary: Ethics is referred to in the context of exceptions to releasing data in several sections of the guidelines.

Analysis and scoring are based on ethics references found in the following:

- Section 1, Background, paragraph 1.
- Section1.1, Open Data Policy, paragraph titled “Exceptions”.
- Section1.2, FAIR Data Principles, paragraph titled “Accessibility”.
- Section 2.1.2, Software source code, paragraph 2.

  - Rating – 2
  - Weighting – 1
  - Score – 2

Criterion 14 – Intellectual property
Criterion summary: The document does not refer to intellectual property, perhaps on the assumption that apart from some exceptions all data is considered open.

Analysis and scoring are based on intellectual property references found in the following:

- There are no references to intellectual property.

- Rating – 1
- Weighting – 1
- Score – 1

Criterion 15 – Licensing/copyright
Criterion summary: Reference is made to Open Content Licences to support dissemination of data deposited in a recognized open repository.

Analysis and scoring are based on references to licences and copyright found in the following:

- Section 2.2, Select a Repository, paragraph 1.

  - Rating – 3
  - Weighting – 1
  - Score – 3

Total document score: 108

CIMMYT Research Data and Information Products Management Policy

Overview
The CIMMYT Research Data and Information Products Policy[60] states that it is aligned with the CGIAR Principles on the Management of Intellectual Assets and the CGIAR Open Access and Data Management Policy. As a consequence, the objective of the policy is similar to that of the other CGIAR and foundation documents reviewed, in that they all aim to support target beneficiaries and the global public at large. The policy proposes (with some qualifications) that the created digital assets be made open access and/or managed as international public goods. In general terms, it provides a slightly higher level of detail on data management than some of the other reviewed documents.

The policy refers to the management of all research data and other information products produced or acquired by CIMMYT to enable its reuse to support target beneficiaries. There is a focus on how data generated through CIMMYT research should be handled, from collection through to analysis and archiving. CIMMYT also clearly outlines the data that is excluded from this policy, specifically non-research data such as project budget data, personnel

[60] CIMMYT (2019) *Research Data and Information Products Management Policy*. https://www.cimmyt.org/content/uploads/2019/10/CIMMYT-Research-Data-and-Information-Product-Management-policy-2019-10.pdf

and facilities data, etc. Additionally, CIMMYT acknowledges that some data (such as draft, poor-quality or incomplete material) will not be of value to others and therefore does not need to comply with its policy. The policy indicates that CIMMYT will apply judgement about what data and digital assets will be made open access.

Compliance with the desk study analysis criteria

Criterion 1 – General data sharing philosophy
Criterion summary: Given the acknowledged alignment with the relevant CGIAR polices on open access and FAIR data, it is considered that the philosophy of the document is strongly centred on ensuring open access to digital assets resulting from research activities. It also emphasizes FAIR-based data management principles. The policy provides details on data management obligations and provides a relatively high level of detail concerning actual data management processes. The policy aims to promote a correct approach to this activity.

Analysis and scoring are based on the overall focus of the document in relation to data sharing concepts, and references are found in the following:

- General Objective section, paragraphs 1 and 2.
- Scope section, paragraph 3.
- Policy – Principles section, subsections 1.1, 1.2 and 1.3.
- Policy – Practices section, subsections 2.1, 2.1.1, 2.2, 2.3, 2.3.1.1, and 2.4

    - *Rating* – 3
    - *Weighting* – 5
    - *Score* – 15

Criterion 2 – FAIR core principles (*Findable*, *Accessible*, *Interoperable*, *Reusable)*
Criterion summary: The policy refers to the FAIR principles as a mechanism to enhance innovation, impact and uptake. There is minimal detail on the FAIR concepts, other than the FAIR acronym being expanded in full.

Analysis and scoring are based on references to the FAIR principles found in the following:

- Policy – Principles section, subsection 1.2.

- Rating – 1
- Weighting – 5
- Score – 5

Criterion 3 – Open data

Criterion summary: The policy states that it is aligned with the CGIAR Open Access and Data Management Policy, which would suggest that "open" is an element of this policy. Reference is also made to ensuring that the research outputs are "provided" openly and are considered as international public goods. While open data principles are not strongly pushed, there is sufficient material in the policy regarding open data to see that the policy is aligned with this principle.

Analysis and scoring are based on open data references that occur throughout the document, with some of the specific references provided below:

- General Objective section.
- Policy – Principles section, subsection 1.2.
- Policy – Practices section, subsection 2.3, 2.3.1 and 2.4.

  - Rating – 3
  - Weighting – 5
  - Score – 15

Criterion 4 – Data management

Criterion summary: The policy states that it is aligned with the CGIAR Principles on the Management of Intellectual Assets and the CGIAR Open Access and Data Management Policy. It also provides a perspective on the full process of what it calls "correct" data management, from capture to archival.

Analysis and scoring are based on data management references found in the following:

- General Objective section, paragraphs 1 and 2.
- Section 2, Practices, subsections 2.1 and 2.2.

  - Rating – 3
  - Weighting – 5
  - Score – 15

Criterion 5 – Metadata

Criterion summary: Approved (standards-based) metadata is referred to in the policy as a data asset and is aimed at ensuring the digital assets are "properly documented".

Analysis and scoring are based on metadata references found in the following:

- Scope section, paragraph 3, second sub-dot point.
- Section 2, Practices, subsections 2.2 and 2.3.2.

    - *Rating – 3*
    - *Weighting – 5*
    - *Score – 15*

Criterion 6 – Data governance

Criterion summary: Data governance is not mentioned in the policy. The focus of the policy is slightly more on research outcomes, rather than how the data is treated, which is where data governance would come in.

Analysis and scoring are based on data governance references found in the following:

- There are no specific references to data governance.

    - *Rating – 1*
    - *Weighting – 4*
    - *Score – 4*

Criterion 7 – Policy compliance requirements

Criterion summary: While compliance is mentioned in the policy, it is in the context of legal frameworks that may impact the sharing of data, not specifically researchers' compliance with the CIMMYT Policy.

Analysis and scoring are based on references to policy compliance found in the following:

- Policy – Practices section, subsection 2.3.2.

    - Rating – 1
    - Weighting – 4
    - Score – 4

Criterion 8 – Incentives to encourage data sharing

Criterion summary: The policy does not refer to any incentives to facilitate and encourage data sharing. Given that CIMMYT follows the CGIAR policies and indicates that the digital assets resulting from research activities are considered international public goods, it may be considered that incentives are not required.

Analysis and scoring are based on incentives references found in the following:

- There are no references to incentives to share data in the policy.

    - Rating – 1
    - Weighting – 3
    - Score – 3

Criterion 9 – Compliance with standards (including data formats, metadata structures and other relevant standards)
Criterion summary: While standards are referred to in the policy, it is generally in the context of the policy itself being a "standard", rather than specific agreed international standards for data formats etc. The exception to this is the requirement to comply with CIMMYT-approved metadata, which could be taken as a "standard" of sorts.

Analysis and scoring are based on references to compliance with standards found in the following:

- Scope section, paragraphs 1 and 3.
- Section 2, Practices, subsection 2.1

    - *Rating* – 1
    - *Weighting* – 3
    - *Score* - 3

Criterion 10 – Breadth of digital assets (i.e. does the policy go beyond just data and also include other digital assets?)
Criterion summary: The policy refers to an extensive range of digital assets, including reports, papers, other documents, metadata, analysed data, software, and many other items. It is quite extensive in its coverage of digital assets.

Analysis and scoring are based on the types of digital assets referred to in the document and references to these are found in the following:

- Scope section, paragraph 3 – all dot points and sub-dot points.

    - *Rating* – 3
    - *Weighting* – 2
    - *Score* – 6

Criterion 11 – Geographic coverage

Criterion summary: The policy does not specifically refer to its geographic coverage other than by saying that research data and information products can "render benefits to CIMMYT's targeted beneficiaries and the global public at large". Global coverage is also reinforced by the concept of the digital assets being an international public good. The policy does identify key stakeholders who will benefit from its application, namely "the poor, especially smallholder farmers in developing countries".

Analysis and scoring are based on geographic coverage references found in the following:

- Policy – Principles section.
- Policy – Practices, subsection 2.3.
- Policy – Practices, subsection 2.4.

    - *Rating* – 3
    - *Weighting* – 1
    - *Score* – 3

Criterion 12 – Privacy

Criterion summary: The policy talks about privacy and the need to "safe-guard the privacy and data confidentiality of private, confidential or restricted information". There is acknowledgement of the issue of privacy within the policy.

Analysis and scoring are based on privacy references found in the following:

- Policy – Principles section, subsection 1.3.
- Policy – Practices section, subsection 2.3.2.

    - Rating – 3
    - Weighting – 1
    - Score – 3

Criterion 13 – Ethics

Criterion summary: There is no specific reference to ethics in relation to data assets, although with references to protecting data in the context of privacy, data confidentiality, and release restrictions arising from legislation or internal policies, it could be considered that ethics has some visibility in this policy.

Analysis and scoring are based on ethics references found in the following:

- No specific reference is made to ethical use of digital assets.

    - Rating – 1
    - Weighting – 1
    - Score – 1

Criterion 14 – Intellectual property

Criterion summary: There is no specific reference to intellectual property, although the policy does refer to the CGIAR Principles on the Management of Intellectual Assets.

Analysis and scoring are based on references to intellectual property found in the following:

- General Objective section, paragraph 2.

    - Rating – 1
    - Weighting – 1
    - Score – 1

Criterion 15 – Licensing/copyright

Criterion summary: The policy contains no references to licensing or copyright. This may be due to the premise that all digital assets should be considered open and international public goods.

Analysis and scoring are based on references to licences and copyright found in the following:

- The policy has no direct references to licences or copyright.

    - Rating – 1
    - Weighting – 1
    - Score – 1

Total document score: 94.

Bill and Melinda Gates Foundation Open Access Policy

Overview
The Bill and Melinda Gates Foundation Open Access Policy[61] is brief but provides a very clear and unambiguous approach to ensuring that data and other digital assets created with foundation funding are made open. The policy uses terminology such as "*unrestricted access and reuse of all peer-reviewed publish research*" when the foundation has provided any funding. This is regardless of whether there may be other funding entities involved.

This is a very strong message and while there is little detail in the policy on how the digital assets can be made open it is clearly the intent of the foundation to leverage research outputs through their release into the public domain.

There are some references to FAIR, with a view that recipients are "encouraged" to follow FAIR principles. Because of this "high level" approach to open data and other digital assets it is potentially possible for researchers to apply limited (poor) management practices within a research programme, as long as the results are published. The policy seems to imply good data management practices but without any significant mention of this component other than "encouraging" FAIR compliance. The focus of the policy is simply on making research results available with unrestricted access, with limited emphasis on the potential usefulness of the resulting digital assets.

Compliance with the desk study analysis criteria
General philosophy: The document provides no real focus on the process that should be used to manage data effectively in a research project, other than a reference to encouraging researchers to apply FAIR principles. While the "open" is strongly referenced, there is limited detail on ensuring the digital assets are effectively managed.

Analysis and scoring are based on the overall focus of the document in relation to data sharing concepts and references are found in the following:

- Section 1, Our Policy.
- Section 2, Policy, subsections 1, 3 and 5.

[61] Bill & Melinda Gates Foundation (date) *The Bill & Melinda Gates Foundation's Open Access Policy*. Accessed 24 November 2023. https://openaccess.gatesfoundation.org/.

- *Rating* – 2
- *Weighting* – 5
- *Score* – 10

Criterion 2 – FAIR core principles (*Findable*, *Accessible*, *Interoperable*, *Reusable)*
Criterion summary: While FAIR principles are mentioned, there is no specific detail concerning the level of FAIR compliance that is required within research projects. Researchers are simply "encouraged" to use FAIR.

Analysis and scoring are based on FAIR principles references found in the following:

- Section 3, Policy, subsection 3, paragraph 2.

  - Rating – 1
  - Weighting – 5
  - Score – 5

Criterion 3 – Open data
Criterion summary: The document is titled "Open Access Policy" and is strongly focussed on open (unrestricted) access and reuse of all peer-reviewed published research; this covers the "underlying" data sets and related digital assets.

Analysis and scoring are based on open data references that occur throughout the document, with some of the references provided below:

- Section 1, Our Policy.
- Section 3, Policy, preamble, and subsections 1, 2, 3, 4, 5, 6 and 7.

  - Rating – 3
  - Weighting – 5
  - Score – 15

Criterion 4 – Data management
Criterion summary: Data management is not specifically mentioned in the policy but it could be considered to be implied by the reference made to "encouraging" the application of the FAIR principles.

Analysis and scoring are based on data management-related references found in the following:

- Section 3, Policy, subsection 3, paragraph 2.

    - *Rating – 1*
    - *Weighting – 5*
    - *Score – 5*

Criterion 5 – Metadata

Criterion summary: Metadata is mentioned in the context of a "Data Availability Statement", which the policy states must be provided together with any article submitted for publication. Further reference to metadata is made in the requirement that the foundation is identified in metadata as providing the research funding.

Analysis and scoring are based on metadata references found in the following:

- Section 3, Policy, subsection 3, paragraph 1.
- Section 3, Policy, subsection 5, paragraph 1.

    - *Rating – 2*
    - *Weighting – 5*
    - *Score – 5*

Criterion 6 – Data governance

Criterion summary: There is no reference to data governance in the policy. The policy is more focussed on the mechanisms enabling unrestricted access to digital assets, such as licensing and depositing material in appropriate open repositories. As with the data management criterion, the only reference to data governance is the "encouragement" to apply the FAIR principles.

Analysis and scoring are based on data governance references found in the following:

- The are no references to data governance.

    - *Rating – 1*
    - *Weighting – 4*
    - *Score – 4*

Criterion 7 – Policy compliance requirements

Criterion summary: The document describes the requirement to comply with the policy in strong terms. Compliance is a condition of the funding, with continuous reviews of compliance to be undertaken.

Analysis and scoring are based on references to policy compliance found in the following:

- Section 3, Policy, subsection 7.

    - Rating – 3
    - Weighting – 4
    - Score – 12

Criterion 8 – Incentives to encourage data sharing
Criterion summary: The policy does not describe any incentives to promote data sharing.

Analysis and scoring are based on incentives references found in the following:

- There are no references to incentives to share data in the policy.

    - Rating – 1
    - Weighting – 3
    - Score – 3

Criterion 9 – Compliance with standards (including data formats, metadata structures and other relevant standards)
Criterion summary: There is no reference to relevant data and digital asset standards, structures, or protocols. The only reference in this area is in relation to ensuring that to ensure open access, relevant licensing methods are used, specifically Creative Commons.

Analysis and scoring are based on references to compliance with standards found in the following:

- There are no references to digital asset-related standards.

    - *Rating* – 1
    - *Weighting* – 3
    - *Score* – 3

Criterion 10 – Breadth of digital assets (i.e. does the policy go beyond just data and also include other digital assets?)
Criterion summary: The policy refers to "all peer-reviewed published research consisting of manuscripts and data", and also refers to metadata, software and any other relevant (digital asset) material. While not strongly worded, it does imply that it applies to all digital assets resulting from the research activity.

Analysis and scoring are based on the types of digital assets referred to in the document, and references to these are found in the following:

- Section 1, Our Policy.
- Section 2, Scope.
- Section 3, Policy, subsection 3, paragraphs 1 and 2.
- Section 3, Policy, subsection 5, paragraph 1.

    - *Rating* – 3
    - *Weighting* – 2
    - *Score* – 6

Criterion 11 – Geographic coverage
Criterion summary: While it is not specifically stated, it could be implied that the geographic coverage of the policy for every foundation-funded research activity includes everything from local to potentially global scales, with the outcomes of these research activities needing to be available for reuse at the global scale. No geographic extent or limitation is set out.

Analysis and scoring are based on geographic coverage references found in the following:

- No specific geographic references are made.

    - *Rating* – 3
    - *Weighting* – 1
    - *Score* – 3

Criterion 12 – Privacy
Criterion summary: There is no reference to privacy around data or other digital assets in the policy.

Analysis and scoring are based on privacy references found in the following:

- There are no references to privacy in the document.

- Rating – 1
- Weighting – 1
- Score – 1

Criterion 13 – Ethics
Criterion summary: As with privacy above, there is no reference to ethics in relation to the research or any research outcomes.

Analysis and scoring are based on ethics references found in the following:

- There are no references to ethics in the policy.

  - Rating – 1
  - Weighting – 1
  - Score – 1

Criterion 14 – Intellectual property
Criterion summary: As with ethics and privacy, discussed above, there is no mention of intellectual property in the policy (other than to copyright for research articles).

Analysis and scoring are based on intellectual property references found in the following:

- There are no references to intellectual property in the policy.

  - Rating – 1
  - Weighting – 1
  - Score – 1

Criterion 15 – Licensing/copyright
Criterion summary: Reference is made to licensing and copyright and the use of Creative Commons or equivalent licences.

Analysis and scoring are based on licensing and copyright references found in the following:

- Section 3, Policy, subsections 1 and 2.

  - Rating – 2
  - Weighting – 1
  - Score – 2

Total document score: 81

CIMMYT – Terms and Conditions Subgrant Agreement

Overview
This document is not a policy, but rather a legal document that covers the agreement between CIMMYT and a subgrantee who is receiving research funding. It provides a legal perspective on meeting CIMMYT's research project requirements, including the availability of data and other digital assets for reuse, the management of intellectual property, and the storage of digital assets.

The agreement does not describe any specific approach to data management (such as applying FAIR principles) but does set out the terms and conditions for managing intellectual property and the storage and dissemination of relevant project data and information products. The digital assets are described as international public goods, implying that there is benefit to more than one country – with regional and global public goods also being considered international public goods.[62] Two main sections of the document provide some detail on data sharing requirements: Section 7, Management of Intellectual Property; and Section 8, Storage and Dissemination of Project Data and Information Products. Most of the rest of the document covers other legal aspects of the subgrant agreement and does not significantly influence CIMMYT's data management and sharing objectives.

Compliance with the desk study analysis criteria

Criterion 1 – General data sharing philosophy
Criterion summary: The overall philosophy of the document is that intellectual property resulting from research activity should be managed as international public goods. Its emphasis is on intellectual property, with the management, storage and dissemination of this intellectual property well-documented from a predominantly legal agreement perspective. A range of data management and dissemination references are included in the document.

[62] Yale Center for the Study of Globalization (no date) "Global Public Goods: What are they and why many are in short supply". https://ycsg.yale.edu/sites/default/files/files/Global-Public-Goods-expl.pdf

Analysis and scoring are based on the overall focus of the document in relation to data sharing concepts, and references are found in the following:

- Section 7, Management of Intellectual Property, subsections 7.1, 7.2, 7.3 and 7.6.
- Section 8, Storage and Dissemination of Project Data and Information Products, subsections 8.1, 8.2, 8.3, 8.4, 8.5 and 8.6.

    - *Rating* – 3
    - *Weighting* – 5
    - *Score* – 15

Criterion 2 – FAIR core principles *(Findable*, *Accessible*, *Interoperable*, *Reusable)*
Criterion summary: The agreement makes no specific reference to FAIR principles although it does refer to a number of documents from CGIAR and other CIMMYT documents that do make some references to FAIR.

Analysis and scoring are based on FAIR principles references found in the following:

- There is no specific reference to FAIR.

    - Rating – 1
    - Weighting – 5
    - Score – 5

Criterion 3 – Open data
Criterion summary: The agreement refers to CGIAR and other CIMMYT documents in relation to open data in terms of access, storage and dissemination.

Analysis and scoring are based on open data references that occur throughout the document, with some of the specific references provided below:

- Section 7, Management of Intellectual Property, subsection 7.1 b.
- Section 8, Storage and Dissemination of Project Data and Information Products, subsection 8.1.

    - Rating – 1
    - Weighting – 5
    - Score – 5

Criterion 4 – Data management

Criterion summary: The agreement requires the subgrantee to comply with a range of intellectual property management policies, including CGIAR policy on Principles on the Management of Intellectual Assets and the Open Access and Management Policy. Other CIMMYT and CGIAR internal policies are also referred to. Data management itself is not specifically mentioned (other than in the context of a data management plan).

Analysis and scoring are based on data management references found in the following:

- Section 7, Management of Intellectual Property, subsection 7.1.

    - Rating – 2
    - Weighting – 5
    - Score – 10

Criterion 5 – Metadata

Criterion summary: Metadata is only referred to as a component of project data and information products in the definitions section of the agreement. There is no reference to metadata in the context of data management and FAIR principles.

Analysis and scoring are based on metadata references found in the following:

- Annex 1, Section 1, Definitions.

    - *Rating – 1*
    - *Weighting – 5*
    - *Score – 5*

Criterion 6 – Data governance

Criterion summary: Data governance is not mentioned in the agreement document. The focus of the document is more on how to ensure the sharing of the research-developed digital assets, rather than how they are created and managed.

Analysis and scoring are based on data governance references found in the following:

- There are no references to data governance.

- *Rating – 1*
- *Weighting – 4*
- *Score – 4*

Criterion 7 – Policy compliance requirements
Criterion summary: There is a requirement for researchers to comply with the relevant CGIAR and CIMMYT policies. However, there is no mention of what non-compliance might mean.

Analysis and scoring are based on references to compliance found in the following:

- Policy – Practices section, subsection 2.3.2.

  - Rating – 2
  - Weighting – 4
  - Score – 8

Criterion 8 – Incentives to encourage data sharing.
Criterion summary: The agreement does not refer to any incentives to facilitate and encourage data sharing, other than a reference to complying with both CGIAR and CIMMYT policies.

Analysis and scoring are based on incentives references found in the following:

- There are no specific references to incentives to share data in the agreement.

  - Rating – 1
  - Weighting – 3
  - Score – 3

Criterion 9 – Compliance with standards (including data formats, metadata structures and other relevant standards)
Criterion summary: Standards are mentioned in the agreement document, but not in the context of data formats and related standards. Standards are referred to in relation to social, integrity and international accepted practices concepts.

Analysis and scoring are based on references to compliance with standards found in the following:

- Section 11, Continuing Representation and Undertakings, subsection 11.1, parts a, b and h.

    - *Rating* – 1
    - *Weighting* – 3
    - *Score* – 3

Criterion 10 – Breadth of digital assets (i.e. does the policy go beyond just data and also include other digital assets?)
Criterion summary: The policy defines an extensive range of material covered under the definitions either of intellectual property or of project data and information products. This encompasses many different potential project outputs, including data, metadata, analysed data, books, audio files, videos, etc.

Analysis and scoring are based on the types of digital assets referred to in the document, and references to these are found in the following:

- Annex 1, Section 1, Definitions – Intellectual Property.
- Annex 1, Section 1, Definitions – Project Data and Information Products.

    - *Rating* – 3
    - *Weighting* – 2
    - *Score* – 6

Criterion 11 – Geographic coverage
Criterion summary: The policy does not specifically refer to its geographic coverage, other than by saying that research data and information products are considered international public goods, implying availability at a global scale.

Analysis and scoring are based on geographic coverage references found in the following:

- Section 7, Management of Intellectual Property, subsection 7.1.

    - *Rating* – 3
    - *Weighting* – 1
    - *Score* – 3

Criterion 12 – Privacy
Criterion summary: The agreement talks about the need for a discussion of any restrictions on data sharing, such as privacy issues, before the project commences.

Analysis and scoring are based on privacy references found in the following:

- Section 7, Management of Intellectual Property, subsection 8.4.

    - Rating – 2
    - Weighting – 1
    - Score – 2

Criterion 13 – Ethics
Criterion summary: Ethics is referred to in the context of CIMMYT policies such as the Code of Conduct, Ethics in Research.

Analysis and scoring are based on ethics references found in the following:

- Section 3.7, second paragraph.

    - Rating – 2
    - Weighting – 1
    - Score – 1

Criterion 14 – Intellectual property
Criterion summary: The management of intellectual property is covered to a considerable degree by Section 7, which is devoted to the management of intellectual property. The aim is to meet CIMMYT's objective of making research outcomes and resulting digital assets available for reuse.

Analysis and scoring are based on intellectual property-related references found in the following:

- Section 7, Management of Intellectual Property, most subsections.
- Section 8, Storage and Dissemination of Project Data and Information Products, subsection 8.4.

    - Rating – 3
    - Weighting – 1
    - Score – 3

Criterion 15 – Licensing/copyright

Criterion summary: As with intellectual property, licences are discussed, and are required when necessary to ensure that the research digital assets are available externally. The licences are required to provide a "non-exclusive, worldwide, royalty free, irrevocable license to use (digital assets), further develop, disseminate and sub license any IP generated in the research project". Copyright is mentioned in the definitions section, under intellectual property.

Analysis and scoring are based on references to licences and copyright found in the following:

- Section 1, Definitions – Background Intellectual Property.
- Section 7, Management of Intellectual Property, subsections 7.2.1, 7.2.2 and 7.3.

    - Rating – 3
    - Weighting – 1
    - Score – 3

Total document score: 77.

<u>CIMMYT Intellectual Property Policy</u>

Overview

The CIMMYT Intellectual Property Policy[63] focusses on intellectual property created through CIMMYT research processes, and also covers the intellectual property of organizations (mostly commercial entities) who work with CIMMYT and bring contributions that include their own intellectual property. The objective of the policy is to ensure that intellectual property is catered for when required, and that it is supported to achieve CIMMYT's mission, which is stated to be as follows:

> *"CIMMYT acts as a catalyst and leader in a global maize and wheat innovation network that serves the poor in developing countries. Drawing on strong science and effective partnerships, we create, share, and use knowledge and technology to increase food security, improve the productivity and profitability of farming systems, and sustain natural resources."*

[63] CIMMYT (2009) *Intellectual Property Policy*. https://www.cimmyt.org/content/uploads/2019/04/CIMMYT-Intellectual-Property-policy-2009-04.pdf

A focus of the policy is on ensuring that CIMMYT research products are available as international public goods and are appropriately protected as necessary to fulfil the CIMMYT mission. The policy also recognizes the need to work with the private sector so that CIMMYT can avail itself of the best technologies to achieve the maximum impact from its research outputs.

The policy does not provide any views on data management in general, or the FAIR principles in particular. The concept of the research outputs being an international public good indicates support for the open digital asset concept, although open digital assets are not mentioned specifically in the policy. The only reference to "open" relates to the Open Content Licence to support reuse and distribution while protecting the assets from commercialization.

Compliance with the desk study analysis criteria

Criterion 1 – General data sharing philosophy
Criterion summary: As explained above, the policy focusses on intellectual property. However, the motivation behind the intellectual property policy is to ensure that maximum benefit is gained from CIMMYT research outputs. Intellectual property is seen as a mechanism to assist CIMMYT in meeting its mission, and CIMMYT applies intellectual property when it believes there will be a benefit in doing so. The policy also acknowledges that with private sector involvement there may be commercial intellectual property issues that need support to ensure maximum benefit is realized.

- Most sections of the document provide a view of the research outcomes being widely distributed and considered as international public goods.

    - Rating – 3
    - Weighting – 5
    - Score – 15

Criterion 2 – FAIR core principles *(Findable*, *Accessible*, *Interoperable*, *Reusable)*
Criterion summary: The FAIR data principles are not mentioned in the policy.

Analysis and scoring are based on FAIR principles references found in the following:

- No mention is made of the FAIR data principles.

    - Rating – 1
    - Weighting – 5
    - Score – 5

Criterion 3 – Open data

Criterion summary: The only reference to "open" is in the context of Open Content Licences. However, the mission of CIMMYT is to make its research outputs widely available (as international public goods), so the inference is that the policy does support open data.

Analysis and scoring are based on open data references that occur throughout the document, with some of the specific references provided below:

- Section 1,Preamble, paragraph 3.
- Section 5, Principles and Operating Policies, subsections 12 and 13.

    - Rating – 2
    - Weighting – 5
    - Score – 10

Criterion 4 – Data management

Criterion summary: As described in the overview above, the policy focusses on intellectual property. Consequently, there are no references to data management.

Analysis and scoring are based on data management references found in the following:

- No references are made to data management.

    - Rating – 1
    - Weighting – 5
    - Score – 5

Criterion 5 – Metadata

Criterion summary: There is no mention of metadata in the policy document.

Analysis and scoring are based on metadata references found in the following:

- There are no references to metadata

    - Rating – 1
    - Weighting – 5
    - Score – 5

Criterion 6 – Data governance
Criterion summary – Data governance is not mentioned in the policy.

Analysis and scoring are based on data governance references found in the following:

- There are no references to data governance.

    - Rating – 1
    - Weighting – 4
    - Score – 4

Criterion 7 – Policy compliance requirements
Criterion summary: There is no reference to compliance with the policy, other than stating that all those participating in a project will "abide by" the intellectual property policy.

Analysis and scoring are based on references to policy compliance found in the following:

- Section 5, Principles and Operating Policies, subsections 8, 14 and 15.
- Section 7, Implementation, Review and Revision, subsection 2.

    - Rating – 1
    - Weighting – 4
    - Score – 1

Criterion 8 – Incentives to encourage data sharing
Criterion summary: There is no mention of any incentives to encourage the sharing of any form of research output.

Analysis and scoring are based on incentives references found in the following:

- There are no references to incentives to share data or research outputs in the policy.

    - Rating – 1
    - Weighting – 3
    - Score – 1

Criterion 9 – Compliance with standards (including data formats, metadata structures and other relevant standards)
Criterion summary: There are no references to standards in the policy.

Analysis and scoring are based on references to compliance with standards found in the following:

- There are no references to standards.

    - Rating – 1
    - Weighting – 3
    - Score – 3

Criterion 10 – Breadth of digital assets (i.e. does the policy go beyond just data and also include other digital assets?)
Criterion summary: The policy refers to all aspects of research outputs. It could be considered that everything produced through a CIMMYT research activity is covered by this policy and so therefore the breadth of digital assets covered by the policy could be considered extensive.

Analysis and scoring are based on the types of digital assets referred to in the document, and references to these are found in the following:

- Section 1, Preamble, paragraph 3.
- Section 5, Principles and Operating Policies, subsection 5.

    - Rating – 3
    - Weighting – 2
    - Score – 6

Criterion 11 – Geographic coverage
Criterion summary: The policy does not specifically refer to its geographic coverage, other than by saying that research outcomes are international public goods. It also mentions “serving the global poor”, which provides a second indication of the global coverage of its activities.

Analysis and scoring are based on geographic coverage references found in the following:

- Section 1, Preamble, paragraphs 3 and 4.
- Section 4, Genetic Resources, paragraph 1 (refers to “farmers across the world”).

    - Rating – 3
    - Weighting – 1
    - Score – 3

Criterion 12 – Privacy
Criterion summary: The issue of privacy is not raised in the policy document.

Analysis and scoring are based on privacy references found in the following:

- No reference is made to privacy.

    - Rating – 1
    - Weighting – 1
    - Score – 1

Criterion 13 – Ethics
Criterion summary: Ethics is referred to in the context of genetic resources.

Analysis and scoring are based on ethics references found in the following:

- No specific reference is made to ethical use of digital assets.

    - Rating – 1
    - Weighting – 1
    - Score – 1

Criterion 14 – Intellectual property
Criterion summary: The policy is entirely about intellectual property.

Analysis and scoring are based on intellectual property-related references found in the following:

- The policy is entirely focussed on intellectual property.

    - Rating – 3
    - Weighting – 1

- Score – 3

Criterion 15 – Licensing/copyright
Criterion summary: Reference is made to Open Content Licences to support dissemination.

Analysis and scoring are based on references to licences and copyright found in the following:

- Section 5, Principles and Operating Policies, subsection 13.

  - Rating – 3
  - Weighting – 1
  - Score – 3

Total document score: 71.

Bill and Melinda Gates Foundation Global Access Statement

Overview
The Bill and Melinda Gates Foundation Global Access Statement[64] is focussed on ensuring that the foundation's funded research outcomes are made accessible to the global community. Consequently, it mostly sets out the obligations and requirements to ensure maximum accessibility, with the emphasis on what is required, and limited comment on how this can best be achieved. Some of the elements considered significant to this review, such as data and digital asset sharing, are described, but in limited detail and with no reference to FAIR principles.

The statement sets out a range of obligations and approaches relating to achieving the foundation's accessibility objectives. The statement is realistic in its understanding of how accessibility may work given the potential breadth of research outcomes (not just data and publications), and it includes provisions for licences and intellectual property management components. These elements are required in many circumstances, but especially when development and possible manufacturing activity is required to ensure maximum reuse benefits and access to research outcomes. The concepts of the value of publication of research results and data sharing feature prominently in the statement.

[64] Bill & Melinda Gates Foundation (no date) *Global Access Statement*. Accessed 24 November 2023. https://www.gatesfoundation.org/about/policies-and-resources/global-access-statement.

This document has a specific purpose and provides details on how research outcomes should be leveraged and made widely available. It provides the frameworks for how many of the administrative requirements for this key foundation driver can be met, but without any details on the specific FAIR or other data management activities needed to ensure that the digital products are in fact reusable. It also forms one of a series of documents focussing on slightly different aspects of open and FAIR data concepts.

Compliance with the desk study analysis criteria

Criterion 1 – General data sharing philosophy
Criterion summary: The document makes considerable comment on making digital assets resulting from foundation-funded research activities widely accessible through sharing. It is strongly focussed on the concepts of open and sharing, and the mechanisms to support reuse.

Analysis and scoring are based on the overall focus of the document in relation to data sharing concepts, and references are found in the following:

- Section 3, Components of a Global Access Strategy, first paragraph and dot points 1 and 4.

    - *Rating* – 3
    - *Weighting* – 5
    - *Score* – 15

Criterion 2 – FAIR core principles *(Findable*, *Accessible*, *Interoperable*, *Reusable)*
Criterion summary: The document does not refer to FAIR data principles.

Analysis and scoring are based on FAIR principles references found in the following:

- There are no references to FAIR data principles.

    - Rating – 1
    - Weighting – 5
    - Score – 5

Criterion 3 – Open data

Criterion summary: The Global Access Statement refers to making the digital assets created through the foundation's funded research activities available globally and disseminated broadly. While this is the primary focus of the document, there is no specific mention of the open data concept.

Analysis and scoring are based on open data references that occur throughout the document, with some of the references provided below:

- There is no direct mention of open data concepts in the document, although it could be broadly interpreted as having a generally "open" philosophy.

    - Rating – 1
    - Weighting – 5
    - Score – 10

Criterion 4 – Data management

Criterion summary: Data management is referred to in the context of Global Access Commitments Agreement with partners. However, there is no specific mention of the need for effective data management to facilitate reuse of digital assets within the Global Access Statement.

Analysis and scoring are based on data management references found in the following:

- Section 2, Global Access Requirements, dot point 1.

    - Rating – 1
    - Weighting – 5
    - Score – 5

Criterion 5 – Metadata

Criterion summary: Metadata is not referred to in the document.

Analysis and scoring are based on metadata references found in the following:

- There are no references to metadata.

    - *Rating – 1*
    - *Weighting – 5*
    - *Score – 5*

Criterion 6 – Data governance
Criterion summary: There is no reference to data governance in the document. The statement is focussed on ensuring that research outputs are made widely available.

Analysis and scoring are based on data governance references found in the following:

- There are no references to data governance.

    - *Rating – 1*
    - *Weighting – 4*
    - *Score – 4*

Criterion 7 – Policy compliance requirements
Criterion summary: While compliance is not discussed specifically, there is an implication that all activities and all agreements with the grantee and project collaborators will establish collaboration and sharing agreements to meet the foundation's Global Access Requirements. Reference is also made to the foundation possibly conducting due diligence to determine the capabilities of prospective funding recipient organizations to ensure that global access requirements can be met. There are no indications of what might occur if there is any non-compliance.

Analysis and scoring are based on references to policy compliance found in the following:

- Section 3, Policy, subsection 7.

    - Rating – 2
    - Weighting – 4
    - Score – 8

Criterion 8 – Incentives to encourage data sharing.
Criterion summary: The statement does not describe any incentives to promote data sharing.

Analysis and scoring are based on incentive references found in the following:

- There are no references to incentives to share data in the policy.

    - Rating – 1

- Weighting – 3
- Score – 3

Criterion 9 – Compliance with standards (including data formats, metadata structures and other relevant standards)
Criterion summary: There are no references to relevant digital asset standards, structures or protocols.

Analysis and scoring are based on references to compliance with standards found in the following:

- There are no references to digital asset-related standards.

  - *Rating* – 1
  - *Weighting* – 3
  - *Score* – 3

Criterion 10 – Breadth of digital assets (i.e. does the policy go beyond just data and also include other digital assets?)
Criterion summary: The statement refers to a wide range of digital assets, including data, products, services, processes, technologies materials, software and other innovations to be made globally available.

Analysis and scoring are based on the types of digital assets referred to in the document, and references to these are found in the following:

- Section 1, Global Access Policy.
- Section 3, Components of a Global Access Policy, dot points 1 and 4.

  - *Rating* – 3
  - *Weighting* – 2
  - *Score* – 6

Criterion 11 – Geographic coverage
Criterion summary: The title of the statement makes clear that it is intended to cover activities across the globe. “Global access” is referred to in many of the document’s sections.

Analysis and scoring are based on geographic coverage references found in the following:

- The concept of global access is referred to in most paragraphs of

the document.

- *Rating* – 3
- *Weighting* – 1
- *Score* – 3

Criterion 12 – Privacy

Criterion summary: There is no reference to privacy around data or other digital assets in the statement.

Analysis and scoring are based on privacy references found in the following:

- There are no references to privacy in the document.

  - Rating – 1
  - Weighting – 1
  - Score – 1

Criterion 13 – Ethics

Criterion summary: As with privacy, discussed above, there is no reference to ethics in relation to the research or any research outcomes.

Analysis and scoring are based on ethics references found in the following:

- There are no references to ethics in the policy.

  - Rating – 1
  - Weighting – 1
  - Score – 1

Criterion 14 – Intellectual property

Criterion summary: Reference is made to intellectual property in a number of areas of the document. There is also reference to the foundation's Global Access and Intellectual Property Portal. Most intellectual property references relate to the view that intellectual property rights are often critical to achieving global access for the research outcomes. The focus on intellectual property is in relation to ensuring the broadest access can be provided, without interfering with this access and availability. Additionally, the statement aims to ensure that any intellectual property rights of third parties are explained in the research plan, again with a focus on not restricting accessibility of research-generated digital assets.

Analysis and scoring are based on intellectual property references found in the following:

- Section 1, Global Access Statement, paragraph 2; and included in Related Links.
- Section 2, Global Access Requirements, paragraph 1 and dot points 1 and 2.
- Section 3, Components of a Global Access Strategy, dot points 1, 2, 3 and 6 (sub-dot points 1 and 2).

    - Rating – 3
    - Weighting – 1
    - Score – 3

Criterion 15 – Licensing/copyright
Criterion summary: Reference is made to licensing and copyright in the context of certain situations where research outcomes need some form of protection to ensure global access. Additionally, licences are mentioned to ensure that any background intellectual property is appropriately covered in regard to undertaking the research activity.

Analysis and scoring are based on references to licences and copyright found in the following:

- Section 3, Components of a Global Access Strategy, dot point 6 (sub-dot point 1).

    - Rating – 3
    - Weighting – 1
    - Score – 3

Total document score: 70.

5. Review of other EiA donor policies

The policies of several other organizations with links to CGIAR and EiA have been examined to determine areas of FAIR and open data alignment and non-alignment.

The organizations and documents are the following:

- FCDO
    - DFID Research Open and Enhanced Access Policy
- USAID

    - Open Government Plan v4.0 (2016)
    - USAID Scientific Research Policy
- NORAD
- Irish Aid
- ICAR

DFID Research Open and Enhanced Access Policy

The UK DFID policy on open and enhanced access focuses very strongly on the principles of their funded research-generated digital products being open and available to everybody.

It provides useful examples of the benefits that arise from having open access, citing several reports discussing these benefits. Some examples point to the increased effectiveness and cost effectiveness of research as a result of open access. Others indicate that data residing in data repositories has an impact on "wider society and the economy, through the development of new tools and methodologies, new policies and regulatory controls and new products and services".[65]

There are some references to aspects of the technical requirements of FAIR and open data, such as the use of appropriate licences like Creative Commons Licences. Additionally, under the policy, research-generated data must be placed in an open access repository and the researcher (or institution) must retain that data for a minimum of five years, and make it available on request for free.

When looked at in more detail, the policy provides minimal content on how FAIR and open data principles will be achieved. For example, while there is reference to "quality research outputs", including data, it does not provide any indication of how "quality" will be achieved. It also refers to metadata, which indicates an acknowledgement of the more technical requirements, but it does not refer to what that metadata should contain, and it does not refer to any metadata standard.

The policy does refer to an "Implementation Guide" and this provides additional details and descriptions regarding how to achieve open access.

[65] Research Information Network (2011) "Data centres: their use, value and impact". London.

The policy has a broad objective, which is to make the research-generated digital outputs as widely available as possible, with no or minimum restrictions. However, it does not provide any views on research projects themselves, in terms of effective and "best practice" data management processes. There is no real emphasis on ensuring that the digital outputs are reusable. There would appear to be an implication that the data will be managed effectively during research projects, rather than stating some basic requirements for ensuring this: for example, requiring the use of relevant standards for data management, such as formats and metadata structures to simplify discovery and reuse.

In the context of this review of alignment and non-alignment of policy documents, the DFID policy can be considered to align at a high level. This means that its objectives align with the other reviewed policies. The intention of these policies is to make research-created digital outputs open and accessible.

USAID

USAID supports the USA's principles of open government – transparency, participation and collaboration. USAID sees itself as a "data driven" organization and considers data generated through its activities as being "precious international development capital"; it has developed several detailed documents that describe a range of open data and data management-related activities and requirements. Together, these documents provide a very comprehensive description of the need for effective data management through the entire data lifecycle, which enhances the usefulness of the data and supports their requirement of making the data open.

The following two USAID documents have been examined:

- Open Government Plan v4.0 (2016).[66]
- USAID Scientific Research Policy.[67]

Open Government Plan v4.0 (2016)

The plan is aimed at supporting the USA's open data agenda and provides various examples of some compelling benefits of making data open. It refers to studies on identifying and distilling best practices for managing data

[66] USAID (no date) *Open Government Plan*. 13 September 2022. https://www.usaid.gov/open/open-government-plan.
[67] USAID (2015) *Scientific Research Policy*. 12 January 2022. https://www.usaid.gov/policy/scientific-research.

across the entire data lifecycle – this is a first in any of the policy documents reviewed for this study. It acknowledges the criticality of managing data effectively at all stages of the data cycle. To further ensure the usability of the data, the plan discusses the concept of workshops where data providers work with data users to improve the value of research data outcomes.

USAID has also developed its own public-facing data repository, the Development Data Library, which provides further indication that USAID understands the value of providing appropriate infrastructure, together with the necessary policies, to ensure maximum utility of the data generated through their funding.

The plan refers to the Global Innovation Exchange, which has developed large databases of global development innovations. While it does not examine these innovations, it is highly likely that many of them will rely on access to high-quality open data sources.

To promote the use of their data and to gain additional benefits, USAID runs competitions to leverage innovation ideas to solve problems. Examples include the following:

- The Bureau for Food Security and the Global Development Lab is partnering around a Precision Agriculture Prize to encourage the development of services or tools that can integrate multiple levels of information and localized data sources into an easy-to-understand and action-oriented farmer interface.
- USAID hosted a first-of-its-kind Open Data Hack-a-thon focusing on crime and violence in Latin America and the Caribbean. The event attracted over 100 participants from approximately 50 organizations worldwide. Participants across four different cities formed teams based on skill sets and interests to create eight unique projects in less than 48 hours. The projects illustrated how open data can help understand crime in the region, informing more responsive programme design.

While the second of these examples of the use of open data is not agricultural, it does demonstrate how activities such as hackathons can achieve results with data that the data owners had not thought of. This increases the value of the data and demonstrates the value of having FAIR and open data principles in place. The aggregation and integration of data from multiple sources can provide similar benefits to agriculture, and similar

outcomes can be achieved through such initiatives. However, the key to achieving such outcomes is the requirement that data be FAIR and open compliant.

The USAID Open Government Plan is very comprehensive, covering a wide range of activities, programmes and concepts to enhance the availability of open data and so enable the data to be leveraged bringing significant benefit. The plan looks at end-to-end data management and the need to ensure that all data is open.

One perceived area that could be strengthened is the need to provide more support for data aggregation. While a data repository provides benefits, there will be a multiplier effect if the data in the repository can be easily integrated to create large (big data) data sets.

USAID Scientific Research Policy
This policy is aimed at all USAID-funded research activities, not just agricultural projects; it follows and complements the USAID Open Government Plan discussed above. It identifies some of the more pragmatic needs for research-generated digital assets that have not generally been identified in the other reviewed policies.

For example:

- Within the context of open access to data and publications, the policy refers to Executive Order 13642, issued on 9 May 2013, which calls for making open and machine-readable the new default for government information. The policy suggests that the benefits of this are as follows: “By making USAID-funded data available through user-friendly platforms in machine-readable formats, host countries, scientists, and communities can propel research forward in solving complex development problems.”

The policy provides a description of what comprises a “well written research plan”, and this includes the following:

- A methodology section that includes study design, population, sample size and statistical power, subject selection, and data collection and measurement methods, and possible limitations.
- Strategies for data management and dissemination to the public.

- Project management, personnel roles and responsibilities, and data handling.

The USAID documents provide considerable detail on effective data management (a prerequisite for FAIR) and strongly support the open data concept, which is a US Government requirement. These documents are considered to align with the EiA FAIR and open data objectives.

NORAD

The NORAD Statement focusses on the leverage of knowledge to "improve the situation" across several sectors in developing and middle-income countries. NORAD aims to support these countries in developing and using research-based knowledge to support policy development.

NORAD is one of the few organizations that talks about standardization[68] and the role standardization can play to help free up resources for the technical co-operation of knowledge programmes. However, no reference was found on their website to effective data management (FAIR) or the value of having research data open.

The NORAD website information is aimed at a very high level and apart from the concept of "standardization" does not go into any detail regarding how knowledge is generated, or the underlying data that is used to create that knowledge.

The NORAD material reviewed is not considered to align with EiA objectives, although the reference to standardization is considered an element that is missing in many of the other policies.

Irish Aid

Irish Aid's policy is titled "A Better World" and is aimed at meeting Ireland's foreign policy objectives for a world that is more equal, peaceful and sustainable.

The strategic policy documents that cover support for agriculture by Irish Aid focus on the high-level approach to solving agricultural production-related issues and how the research can best be put to effective use. Assisting small

[68] Norad (2020) "The Knowledge Bank in Brief". Accessed 24 November 2023. https://norad.no/en/toolspublications/publications/2020/the-knowledge-bank-in-brief/.

farmers to grow more nutritious food for themselves and their families is a priority for Irish Aid.

The emphasis described on the website is on leveraging the overall research outcomes, rather than leveraging any value from the data generated through research. There is no detail on research-generated data, data management, or FAIR and open data principles, and as a consequence the Irish Aid material reviewed is considered to be non-aligned to the EiA data objectives.

ICAR

There does not appear to be any relevant information on the concepts of FAIR and open data on the ICAR website. Reference is made to ICAR achieving an award for open data, but their policies and mandates are not easily found. However, through previous interaction with ICAR, CABI has obtained such documents.

ICAR's institutional approach to research data management is covered in published ICAR guidelines, which emphasize the role of the National Data Sharing and Accessibility Policy for India as being the overarching framework with which ICAR complies.[69] The guidelines, published in 2014, pre-date widespread acceptance of the FAIR framework, but do indicate the merits of publishing selected data sets as open data alongside journal publication. Sharing outside of this is not encouraged; "cooling off" periods are one to three years for publishing of data post-research completion and in some ways data sharing is actively discouraged.

Department Of Science & Technology (no date). *National Data Sharing and Accessibility Policy*. Accessed 30 November 2023. https://dst.gov.in/national-data-sharing-and-accessibility-policy-0.

6. Standardization of EiA data management

Introduction

The review looked at the alignment or non-alignment of a range of policy documents from different organizations participating in EiA. The policies aim to ensure that the data resulting from EiA-related research is both FAIR and open. However, do these policies align with EiA's data expectations and objectives?

From a data management perspective, the focus of these policies **DOES NOT** align with some of the EiA-defined activities and outcomes.

EiA's data-related outcomes cover the following:

- **Facilitating the delivery of agronomy-at-scale solutions,** including development and technical/user experience validation and the co-creation and deployment of gender- and youth-responsive solutions to smallholder farmers via scaling partners.
- **Enabling the creation of value from big data and advanced analytics** through the assembly and governance of data and tools; the application of existing analytics and solutions for specific use cases; the supply of information on climate impacts, and the inclusivity and sustainability of agronomic solutions; and national agricultural research system capacity strengthening.
- **Driving the next generation of agronomy-at-scale innovations** by addressing key knowledge gaps and facilitating innovation in agronomy research through engagement with partners.

EiA Outcome 3 also talks about locally relevant agronomic solutions ***integrating*** climate-smart, inclusivity and sustainability dimensions and assessing their performance using standardized protocols.

The key words that suggest a non-alignment of the existing policies in the EiA objectives are:

- big data
- advanced analytics
- agronomy-at-scale solutions
- integrating

Focus of current policies

The focus of the policies reviewed is on ensuring that each research project creates data that is FAIR-compliant and open. This is a very desirable

objective, but it does result in a gap between the objectives of these policies and EiA's data objectives. The current policies target individual research projects that support the creation of one-off data sets which are often specific and unique to a single research project. If the policies are followed, these one-off data sets will be stored in a data repository together with many others, and available to any potential user.

Why is this a problem?
The open data repositories will have multiple data sets from the various research projects and in most cases different data sets will be different in structure, even when the same types of data are involved. The ability to aggregate similar data types and integrate these with other types of data will be very limited. Different data formats are only one of several areas where there is a potential lack of compatibility of these data sets. The existing policies do not prescribe the use of standards for data outputs.

With the current arrangements (and policies) a researcher wishing to create a large and integrated data set must access the many ad-hoc data sets from a repository. They then need to spend a large amount of time and effort to integrate these different data sets, which can potentially be a long, complex, expensive, time-consuming and sometimes impossible task. Additionally, if a researcher goes to all this effort, their integrated "big data" set is likely to have its own unique format and structure and will not be easily integrated with other big data sets developed in a similar manner.

Another issue created by having numerous "individual" data sets is that there is little incentive to build complex data analytical tools. It is not cost effective, nor likely to achieve any significant results, to create one-off analytical tools to support data sets of only a few hundred measurements. Once there are tens of thousands, or even larger numbers, of measurements in a single form there is real benefit in investing in analytical tools to provide more useful and valid outputs.

The missing link
What is missing from the current policies is the requirement for the standardization of data and data management activates for research data outputs. It is acknowledged that few standards exist across the agriculture research landscape but to achieve the EiA data-related objectives, the complete research data cycle needs to be supported by suitable standards.

Standardized data should be submitted to a "data centre", where it can be easily aggregated, and access provided to these aggregated data sets for researchers working on "at scale agricultural solutions". These aggregated and integrated data sets will then enable the application of existing or specifically developed analytical tools that work with known data structures. This will enable EiA to achieve several of its objectives, including Outcome 3, which focusses on integration and standardized protocols.

Achieving this standardization objective will be difficult and will also require considerable negotiation and discussion. It is also potentially expensive, but the results will outweigh any of these costs. Standards cannot be created quickly, but each step forward in creating appropriate standards will result in improved benefits. Moving towards a standards-based data environment can only be achieved with support from the donors, the research organizations and research centres collaborating to make standards happen.

Such direction needs to come from the highest level within the EiA/CGIAR and donor community. For example, the CGIAR System Council, a high-level group that provides organizational oversight for CGIAR activities, including EiA, may be the appropriate level for moving towards standardization. As a "peak" body, the System Council can determine policy and direction on CGIAR and EiA activities. This body could establish a Data Management Subcommittee to oversee and drive the development of agricultural data standards.

Examples of establishing effective data management committees:
The creation of data management sub-committees within organizations is not a new approach and this concept has been used very successfully in other domains. Examples of similar peak bodies that have established successful data management and standardization sub-committees include the following:

- UNESCO's Intergovernmental Oceanographic Commission (IOC)
  - Committee on International Oceanographic Data and Information Exchange (IODE)
- World Meteorological Organization (WMO)
  - Global Climate Observing System
  - Global Data-processing and Forecasting System
  - Agricultural Meteorological Programme
  - WMO Technical Regulations – an international framework for standardization and interoperability, consisting of standard and

recommended practices and procedures adopted by the World Meteorological Congress for universal application by all Members.

- Open Geospatial Consortium (OGC)
  - A consortium of experts committed to improving access to geospatial or location information. OGC connects people, communities and technology to solve global challenges and address everyday needs.
  - It represents over 500 businesses, government agencies, research organizations, and universities united with a desire to make location information FAIR .
  - OGC standards are the result of an international consensus-based process of technical content-development, followed by intense review and formal vote.

As identified previously, there is a gap between the existing policies of organizations supporting EiA and the objectives of EiA in terms of its data-related outputs and expectations. This gap, in simple terms, relates to the lack of significance given to standardization across the full data cycle covering both FAIR and open data principles. The lack of standardization in the current policies is seen as a significant barrier to achieving many of EiA's data-related objectives.

It is acknowledged that moving down a standards path has a cost, which may result in a reduction in the number of research projects that can be funded. However, a standards-based approach will result in data that can more easily be reused and leveraged to create much greater benefit than is currently being achieved with the one-off approach to research data sets.

Moving towards an "at scale" objective requires a significant change in the project and project governance methodology.

The application of appropriate "whole of cycle" data management standards that support and underpin each individual research projects will result in the ability to aggregate and integrate research data more easily. Aggregate data to enable big data analysis will be greatly simplified, becoming easier, less time-consuming and cheaper, and providing greater benefits.

Recommendations:

- Develop a proposal for the CGIAR System Council to establish an Agricultural Data Management Standing Committee to develop appropriate agricultural data-related standards.
- Invite CGIAR System Council Members from both the research and donor areas to participate in the standards development process through supporting membership of the Committee from their own suitably qualified and experienced staff.

7. Data Governance Framework

Research for CABI by Leigh Dodds

A data governance framework describes how **data governance** will be implemented within an organization, project or collaboration.

It consists of a number of elements, which include, at a minimum, the following:

- A high-level **strategy or principles** that describe the vision for how data governance will be carried out and the benefits of doing it well.
- A clearly defined set of **roles and responsibilities** that describe who is responsible for delivering on specific data governance and management activities.
- A set of **policies** that set out the requirements for the responsible access to, use of, and sharing of data. These cover areas such as data security, privacy and sharing.
- **Process** documentation that describes how key aspects of those policies will be implemented, e.g. handling requests to share data, carrying out data privacy impact assessments, anonymization of data, standards for creating and managing metadata, etc

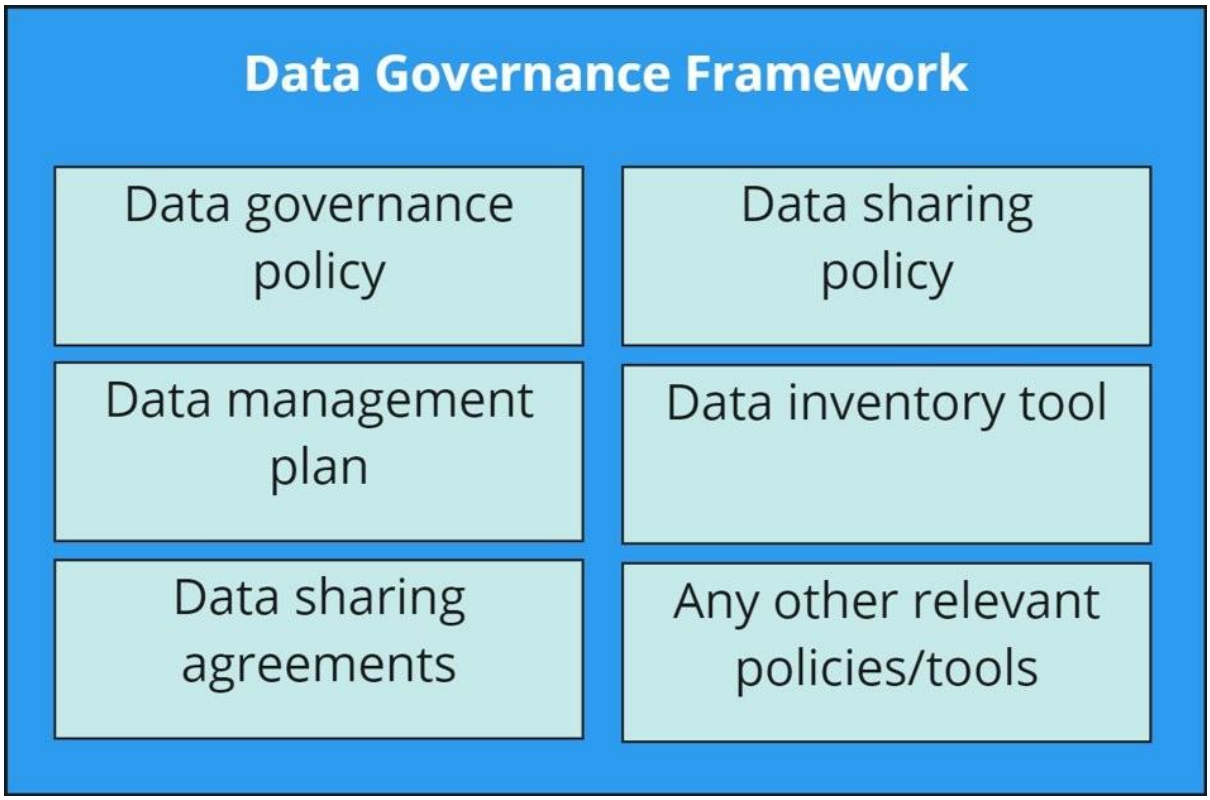
The framework likely consists of a number of separate documents that outline each of these elements in more detail.

A data governance framework might also be supported by the following:

- a set of **metrics or KPIs** that monitor how effectively the framework is being delivered. E.g., as a result of carrying out a data maturity assessment of an organisation or using FAIR assessment tools to review individual datasets or outputs

- a **communication or implementation** plan that will help to embed the framework into the culture of the organisation
- a **training** plan that will ensure that the necessary skills are in place to help deliver on the strategy
- specific **technology**, tools or platforms that will support the implementation of the framework

The image below indicates the key components to be included in a Data Governance Framework:



The blank template below is for a data governance policy specific to a project:

Data Governance Policy: Template

Please refer to the associated Data Governance Policy supporting guidance document to aid the completion of this template.

1. Data Governance Policy Attributes

Version Number	Action / Update	Approved by	Approval date	Effective date	Next review date

2. Definitions

Terms/Acronym	Definition

3. Statement of Purpose

4. Scope

5. Applicability

Entities affected by the policy

Who should read the policy

6. "INSERT PROJECT NAME" Data Governance Framework Principles

High-Level Principles

"INSERT PROJECT NAME" FAIR Aligning Principles

7. Roles and Responsibilities

8. Communication and Implementation Plan

Communication

Implementation

9. Related Funding Requirements, Policy, and Legislation

Funding Requirements

In-Country Policy/Legislative Environment

Existing Institutional Policies

Ethical Implications

8. FAIR Implementation Profile (FIP) template[70]

Community description	
Name of Community	*Leiden Bookseller*
Description of Community	
Supporting Links	
Research Domain	*e.g. Environmental Sciences*
Data Steward	*e.g. ORCID #*
Date of FIP creation	

FAIR principle	Question	FAIR enabling resource types	Your answers
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier type	*e.g. PURL, DOI*
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier type	
F2	Which metadata schemas do you use for findability?	Metadata schema	
F3	What is the technology that links the persistent identifiers of your data to the metadata	Metadata-Data linking mechanism	
F4	In which search engines are your metadata records indexed?	Search engines	
F4	In which search engines are your datasets indexed?	Search engines	
A1.1	Which standardized communication protocol do you use for metadata records?	Communication protocol	
A1.1	Which standardized communication protocol do you use for datasets?	Communication protocol	
A1.2	Which authentication & authorisation technique do you use for metadata records?	Authentication & authorisation technique	
A1.2	Which authentication & authorisation technique do you use for datasets?	Authentication & authorisation technique	
A2	Which metadata longevity plan do you use?	Metadata longevity	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for	Knowledge representation language	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for	Knowledge representation language	
I2	Which structured vocabularies do you use to annotate your metadata records?	Structured vocabularies	
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabularies	
I3	Which models, schema(s) do you use for your metadata records?	Metadata schema	
I3	Which models, schema(s) do you use for your datasets?	Data schema	
R1.1	Which usage license do you use for your metadata records?	Data usage license	
R1.1	Which usage license do you use for your datasets?	Data usage license	
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records?	Provenance model	
R1.2	Which metadata schemas do you use for describing the provenance of your datasets?	Provenance model	

[70] The FIP mini-questionnaire leads users through the creation of their own FAIR Implementation Profile. See: https://bit.ly/yourFIP