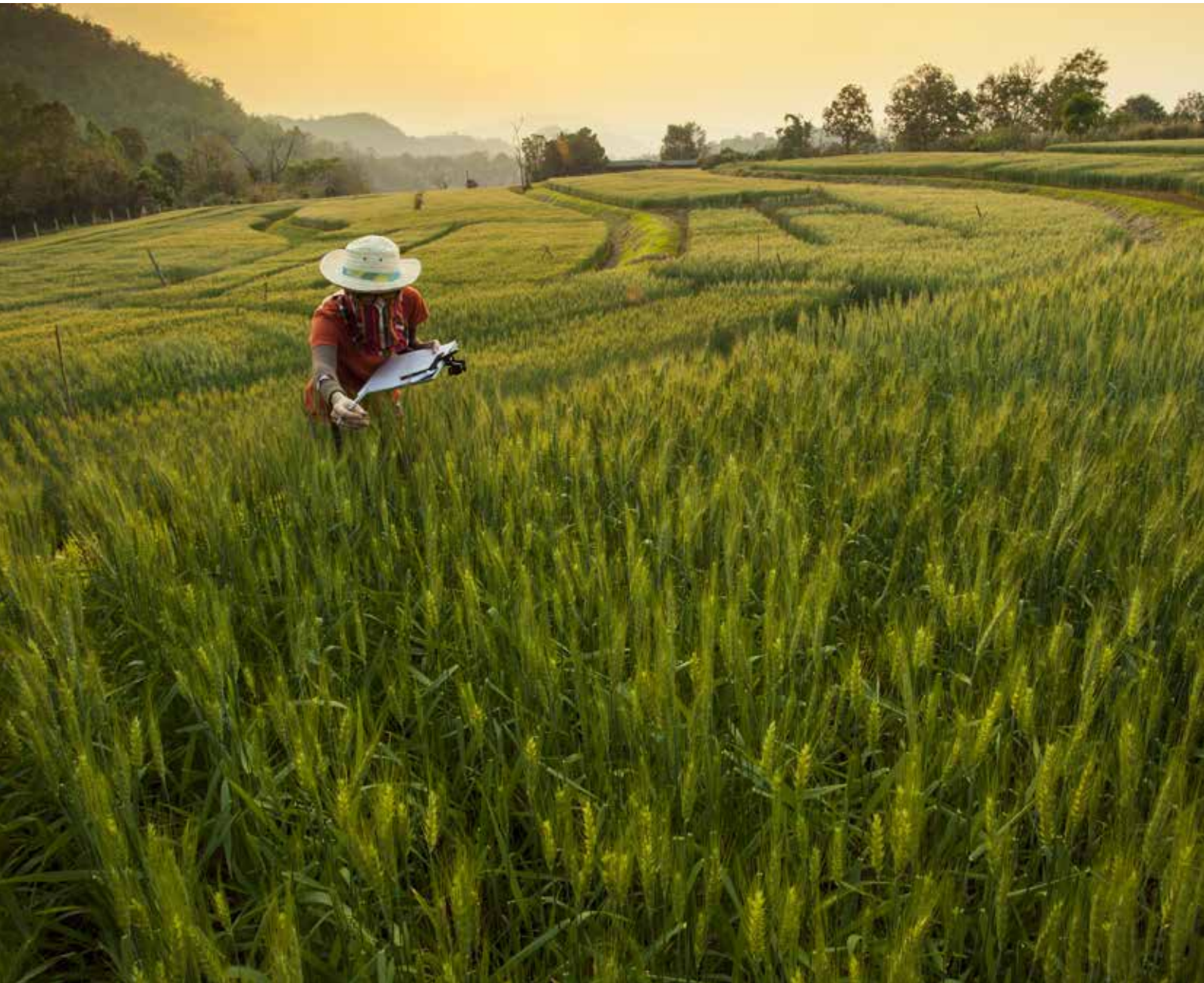
GODAN
Global Open Data
for Agriculture & Nutrition

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



ACKNOWLEDGEMENTS

AUTHORS: Fiona Smith (ODI), Jamie Fawcett (ODI), Ruthie Musker (GODAN)

ACKNOWLEDGEMENTS: Jeni Tennison (ODI), Martin Parr (GODAN), Peter Wells (ODI), Sebastian Mhatre (DFID), Stanley Wood (BMGF), Zachary Bacquet (USAID), Benjamin Vigreaux (DFID), Anna Brenes (USAID)

EDITING: Paul Nagle

DESIGN: Anja Pircher Design (anjapircher.com)

August 2017

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Open data is a powerful tool for promoting more effective decision making, fostering innovation, and driving organisational change through greater transparency[1]. Researchers in the agriculture sector are increasingly encouraged to open data from their field research and activities in order to address the complex challenges around food security and sustainability.

But as funding institutions introduce their own open data policies, researchers are now asked to respond to multiple requests and requirements. This research was commissioned by a group of donors – DFID (the UK Department for International Development), BMGF (the Bill and Melinda Gates Foundation), and USAID (the US Agency for International Development) – in order to understand the opportunities for donors to make open data implementation more efficient and streamlined for their implementing research partners. The hope is to identify patterns of good practice which donors, including those outside agriculture, can build upon and contribute to through further dialogue.

The Open Data Institute and GODAN have conducted a review of policy and data quality in five jointly funded agriculture programmes in order to identify the opportunities for agriculture donors to align their approaches (see Section 3). This was supplemented by a series of interviews and surveys with stake-holders from donor and research partner organisations to gain an impression of how policies are being implemented, including the challenges associated with further adoption of open data (Sections 4 and 5, respectively).

From this multi-faceted review of policy and practice, we found several opportunities where donors of agriculture research programmes can align. We call upon donors to:

1. **Join a global funder dialogue with other donors, researchers, and research institutions**
2. **Support and adopt common policy principles**
3. **Share approaches towards dealing with ethical considerations**
4. **Promote good open data practice among those receiving funding by regularly monitoring compliance and articulating clear expectations regarding budget allocations to ensure open data**
5. **Increase engagement and introduce practical projects to promote data reuse and innovation**
6. **Collect data use stories to demonstrate value and impacts of research data**
7. **Support the capacity of implementing research partners to improve data availability, accessibility, discoverability and quality**
8. **Adopt shared guidelines, tools and templates aimed at reducing the time and cost of policy compliance**
9. **Incentivise researchers to publish by rewarding good quality data production**
10. **Sustainably resource data publication and management**

Furthermore, we call on all funders of agriculture research to join a global funder dialogue with donors and research institutions with the aim of advancing a shared set of principles and sharing good organisational practices, to underpin more harmonized open data implementation.

1 http://www.godan.info/sites/default/files/old/2015/04/ODI-GODAN-paper-27-05-20152.pdf

INTRODUCTION

Food security and sustainable agriculture are among the most pressing global issues of today. Faced with a growing global population likely to reach 9.7 billion by 2050, increasing uncertainty, climate change, land and soil degradation, and food waste, the agriculture sector is under immense pressure to come up with new solutions to complex challenges.

Open access to research and open publication of research data are vital resources to help achieve food and nutrition security. The publication of timely and accurate data can promote more effective decision making, foster innovation and drive organisational change through greater transparency[2] with benefits for farmers, researchers, extension experts, policy makers, governments, and other private sector and civil society stakeholders.

Donors spend billions of dollars on agricultural research every year, in the hopes of improving food and nutrition security. There is wide recognition that publicly funded research and data produced from development programmes are a public good. This means that public and charitable funders of agriculture programmes have a duty to make data collected in the course of implementation available to researchers, policy makers and citizens, subject to appropriate privacy safeguards and ethical considerations.

In response, many public funders and research institutes, including the Wellcome Trust[3] and CGIAR[4], have introduced their own policies and operational guidance for implementing open data across the research programmes they support. As multi-donor research collaborations within agriculture expand, there is a need to consider how to harmonise approaches between donors to maximise impact, while reducing the burden of compliance for research grantees.

Making open data *truly* useful for agriculture and nutrition requires a shared agenda across data producers and funders of agriculture research, alongside actions to build capacity for the use and reuse of data.

Several donors, including the United Kingdom Department for International Development (DFID), the United States Agency for International Development (USAID) and the Bill and Melinda Gates Foundation (BMGF), are responding to this coordination challenge. In partnership with the Global Open Data for Agriculture and Nutrition (GODAN)[5] Initiative and the Open Data Institute (ODI), they are driving efforts to develop shared principles towards open data within the agriculture programmes they support. Harmonising approaches can lead to benefits in terms of policy coherence, streamlined processes, cost efficiencies, and increased confidence and capacity of researchers.

However, there are a number of technical, legal, political, and ethical barriers to overcome before open data becomes the norm. Previous research has highlighted specific challenges relating to data sharing and cooperation within the agriculture sector. These include debates over data ownership[6], responsible use of data[7], equity considerations, and building incentives and trust within a complex ecosystem of actors[8].

Nonetheless, the agriculture sector can take inspiration from other research fields such as genetics, physics, and public health. Data sharing in genomics, for instance, has led to increased efficiency, while also accelerating research discoveries benefiting the public interest. Sector-wide initiatives such as the joint statement by funders of health research[9], convened by the Wellcome Trust, illustrate what this coordinated approach could look like. We hope that this research can provide lessons not only to the agriculture sector, but to other research communities exploring data sharing to address complex systems such as energy, human rights, and conflict and fragility.

2 http://www.godan.info/sites/default/files/old/2015/04/ODI-GODAN-paper-27-05-20152.pdf
3 https://wellcome.ac.uk/funding/managing-grant/policy-data-software-materials-management-and-sharing
4 http://library.cgiar.org/bitstream/handle/10947/4488/Open%20Access%20Data%20Management%20Policy.pdf?sequence=1
5 The Global Open Data for Agriculture and Nutrition initiative http://www.godan.info/
6 http://www.godan.info/documents/ownership-open-data-governance-options-agriculture-and-nutrition-0
7 http://www.godan.info/documents/responsible-data-agriculture
8 http://www.godan.info/documents/global-data-ecosystem-agriculture-and-food
9 https://wellcome.ac.uk/what-we-do/our-work/sharing-research-data-improve-public-health-full-joint-statement-funders-health

BREAKOUT BOX: OPEN DATA VS OPEN ACCESS

"Open access" and "open data" are sometimes used interchangeably, but refer to slightly different ways of releasing and using data and information.

The Open Data Institute defines "**open data**" as "data anyone can access, use or share." The highest quality of open data is data that is accessible to those who need it, machine-readable, accurate, continuously updated, has a unique identifier, can be linked to other data sources, and has an open license allowing reuse of the data in any way as long as the original source is credited. Data can still be considered open even if not all of the above criteria are met owing to necessary privacy or security restrictions.

The main difference between open access and open data policies is the level of mandate required. The open access movement was originally motivated by a desire to open up content that would otherwise be behind paywalls, which meant knowledge wasn't generally available.

Open access usually represents the lowest 'tier' of open data, and is most typically understood in the context of freely available research publications in downloadable format (e.g., PDF). While openly licensed and legally reusable, it might not be documented in an open standard machine-readable format. For data to be interoperable it must be machine-readable, which open access documents usually are not.

Although open access originally referred to the publication of research papers, research funders are now starting to mandate publication of underlying research data as open data. This is motivated by a desire for reproducibility, maximum return on investment, and credit for those researchers who focus on data collection and curation

No matter what the definition, open data and open access must be released responsibly[10,11] with clear ownership and licensing[12].

10 See http://www.godan.info/sites/default/files/old/2015/04/ODI-GODAN-paper-27-05-20152.pdf
11 See http://www.godan.info/documents/responsible-data-agriculture
12 See for example http://www.godan.info/documents/ownership-open-data-governance-options-agriculture-and-nutrition-0

BACKGROUND TO THE RESEARCH

In September 2016 at the GODAN Summit in New York, a group of donors including DFID, USAID and BMGF initiated a discussion about developing common principles to support researchers implementing open data. This research report was commissioned in March 2017 to understand the opportunities, gaps and challenges involved in harmonising donor approaches and provide recommendations.

Research methodology

In this research we are particularly interested in what funders can do to support and encourage their implementing research partners to publish open data. Our focus was on data produced in the course of the agriculture programmes they fund, as opposed to data produced by the donors themselves (e.g., spend data).

The research questions we sought to answer through this research report were:

1. What are **common principles and examples of best practice** across the range of represented donor open data policies? What are the **main areas in which the donor policies differ**, and where does this become a problem for compliance?
2. What are the main **lessons the programmes have learnt** through implementing those policies, including main challenges and opportunities for adoption?
3. What **benefits do implementing research partners perceive** for their programmes or organisations as a result of compliance with open data policies?
4. How well do the selected programmes implement existing donor open data policies? In other words, what is the **quality of open data publication?**
5. How can donor open data policies be **strengthened**, and under what **shared principles?**

To answer these questions, we carried out research in three phases:

1. An **open data policy review** against criteria in the ODI's guide to writing a good open data policy[13]. The guide provides a checklist of policy elements covering data licensing and reuse rights, prioritizing data for release, privacy considerations, data publishing standards, engaging with data users, approaches to consuming data, monitoring commitments, and overall policy transparency.
2. **Key informant interviews and surveys** with a range of stakeholders, including open access policy experts, programme managers from funder organisations and implementing research partners across five jointly funded agriculture programmes.
3. **Assessment of data quality** from the five jointly funded agriculture programmes, using the ODI Open Data Certificates[14] as a framework to review the availability, accessibility, discoverability and quality of open data publication.

It is important to note the diversity of organisational structures and internal constraints facing donor institutions. Therefore we don't expect donors to completely overhaul their existing policies or to all adopt an identical approach as a result of this research. This research is also not intended to provide a comprehensive assessment of donor open data performance across the board. Our small sample of (five) agriculture programmes was intended to draw out illustrative examples of benefits, use and challenges. More holistic assessments of open data can be found elsewhere, e.g. Publish What You Fund's 'Aid Transparency Index'[15].

For a more detailed description of the research methodology, see Annex I.

13 https://theodi.org/guides/writing-a-good-open-data-policy
14 https://certificates.theodi.org/en/
15 http://ati.publishwhatyoufund.org/index-2016/results/

Introduction to selected programmes and stakeholders

Programmes

For the purposes of this research, we wanted to look at a diverse but representative group of jointly funded agriculture programmes which could provide us with lessons about how coordination is currently working in practice, and how it could improve. The programmes also give us a sense of the value of open data to researchers within the sector, and what needs to be done to increase the quality and usability of data. A profile of each of the selected programmes is provided below.

Cereal Systems Initiative for South Asia (CSISA)

Website:	www.csisa.org
Aim of programme:	Catalyse sustainable and inclusive agricultural development using innovative technologies, practices and policies.
Start date:	2009 (Phase 1-3 $33.1m: currently Phase 3 – 2015–2020)
Key activities:	CSISA initiative is helping smallholders to be profitable and productive using sustainable intensification technologies, including water management and mechanisation. The programme is doing this through increasing farming resilience to extreme weather conditions, improving knowledge of best management practices for agriculture, and funding innovative applied research.
Location:	Bangladesh, India and Nepal
Funders:	BMGF and USAID
Implementing research partners:	International Maize and Wheat Improvement Centre (CIMMYT), International Food Policy Research Institute (IFPRI) and International Rice Research Institute (IRRI)

NextGen Cassava

Website:	www.nextgencassava.org http://cassavam.blogspot.co.uk
Aim of programme:	Fund innovative research and development to improve the uptake of disease-resistant cassava, a major crop that is essential to food security.
Start date:	2012 (Currently: Phase 2)
Key activities:	The NextGen Cassava programme funds innovative research and development to improve early detection of major cassava viruses, identify and scale up the production of cassava crops resistant to disease by smallholder farmers, create new breeding tools and techniques for the crop, develop drought tolerance in cassava plants, and establish a functional website about cassava with a tool for browsing the cassava genome, maps and markers.
Location:	Across Sub-Saharan Africa
Funders:	DFID and BMGF (total = $32.4m)
Implementing research partners:	Cornell University, International Institute for Tropical Agriculture (IITA), the National Root Crops Research Institute (NRCRI) and the National Crops Resources Research Institute (NaCRRI), Boyce Thompson Institute (BTI) for Plant Research, US Department of Energy (DOE) Joint Genome Institute (JGI), Makerere University, West African Centre for Crop Improvement (WACCI).

JPAL Agricultural Technology Adoption Initiative (JPAL ATAI)

Website:	www.atai-research.org
Aim of programme:	To develop and test programmes around the adoption of new farming technologies, with a long-term objective to ensure that the poor derive greater benefit from existing and new agricultural technologies.
Start date:	2010 (Currently: Phase 2 – 2016)
Key activities:	The JPAL ATAI programme is funding 15–30 new high-quality randomised control trials, 10–30 diagnostic data collection exercises, forming new partnerships, conducting impact studies around agricultural innovation, and providing evidence on factors that stimulate or hinder technology adoption by smallholder farmers.
Location:	South Asia and Sub-Saharan Africa
Funders:	DFID and BMGF (total = $12m)
Implementing research partners:	MIT Poverty Action Lab (JPAL) and UC Berkeley Center for Effective Global Action (CEGA)

The New Alliance Information and Communication Technologies (ICT) Agriculture Extension Challenge Fund

Website:	http://www.new-alliance.org/
Aim of programme:	Fund aims to use ICT to help smallholder farms and agribusinesses improve farm management, crop yields and food security.
Start date:	2014
Key activities:	Activities vary by country and implementing partner, but include fostering public–private partnerships, supporting the expansion of ICT-enabled advisory services, and supporting the uptake of agricultural technologies by farmers.
Location:	Ethiopia, Ghana, Malawi, Mozambique, Senegal and Tanzania
Funders:	DFID, USAID, BMFG and International Fund for Agricultural Development (IFAD) (Total = $12m)
Implementing research partners:	Digital Green (Ethiopia), Grameen Foundation (Ghana), Catholic Relief Services (Malawi), Concern Universal (Senegal)

Strategic Partnership on Agriculture and Food Systems for Nutrition (Agriculture – Nutrition Impact Studies Programme)

Website:	www.driversoffoodchoice.org
Aim of programme:	Funding research into the impact of agricultural interventions on health to increase investments and policies that improve health outcomes for the poor, particularly young children and pregnant women.
Start date:	2015 (–2022)
Key activities:	This programme is funding research to fill evidence gaps on the impact of agricultural interventions on health outcomes in two areas: • Drivers of Food Choice – initial 8 subgrants concerning food choice and a second round call to include 15 high-quality studies on food choice. • Agriculture–Nutrition Impact Studies Programme – commissioning 4 large-scale impact studies and 2 smaller-scale formative studies.
Location:	Multiple
Funders:	DFID and BMGF (Total = £38.2m)
Implementing research partners:	Lead: University of South Carolina

Stakeholders

Since all of these programmes are funded by multiple donors and implemented by multiple research organisations and individuals, we wanted to gather a range of perspectives to answer our research questions. We spoke to the following groups of people:

- **Strategists** – responsible for designing/overseeing the open data or open access policy, and building capacity of programme managers within the donor organisation;
- **Programme managers** – responsible for overseeing overall implementation of activities within the donor organisation, including supporting implementing research partners to comply with open data policy requirements;
- **Implementing research partners** – responsible for undertaking research activities and directly publishing results and programme information as open data, including on custom open access platforms, open access journals, and donor-specified publishing platforms.
- **External open data/open access policy experts** – experience with designing open data or open access policies in other sectors or funder organisations.

The next section of the report compares the different approaches and publication requirements of the three funder organisations which are the focus of this report (The **United Kingdom Department for International Development (DFID)**, **United States Agency for International Development (USAID)**, and **The Bill and Melinda Gates Foundation (BMGF)**)

DONOR OPEN DATA POLICIES COMPARED

Data exists on a spectrum: it can be closed, shared, or open[16]. Open data is data that anyone can access, use and share[17]. In the past five years a growing number of public and private sector organisations have been drafting policies that outline how they intend to publish data openly and how they intend to consume it.

The ODI and GODAN believe that creation of an open data policy is an important element of developing strong open data practice[18]. From the ODI's experience, a well written open data policy will clearly define the commitment of the organisation to publishing, sharing and consuming data. It will be used by implementing research partners to help prioritize their work on data, and by external stakeholders to understand how an organisation will be handling data so that they may look for opportunities to contribute to and use the data. An open data policy can also help encourage informed reuse of third-party data by researchers and by the donor organisation.

Open data and open access policies take on different goals and forms based on the donor organisation's particular objectives, institutional design, and operational culture. In this section of the report, we explore some of the strengths, similarities and gaps across the three donors we examined for the purposes of this report. The insights are relevant not only to agriculture research work; they apply to work across all sectors and extend beyond research data.

16 https://theodi.org/data-spectrum
17 https://theodi.org/what-is-open-data
18 http://theodi.org/maturity-model

SUMMARY OF DONOR APPROACHES TO OPEN DATA

The United Kingdom Department for International Development (DFID) introduced an Open Data Strategy in 2012[19], which built upon their Information Strategy, and introduced a schedule of new data release commitments covering the period April 2012–March 2014. The strategy sets out DFID's ambition for improving data publication, and driving reforms via transparency, accountability and citizen participation throughout the whole delivery chain. The strategy outlines concrete commitments to redeveloping their aid information platform, complying with IATI, establishing a governance mechanism, and extending the volume and diversity of types of data mandated to be published as open data, e.g. programme evaluations and contracts. To promote use of data the strategy also contemplates a number of practical activities such as pilot programmes to invite feedback from beneficiaries of UK aid, consultations on the usability of data repositories, and a fund to support developers who create innovative tools.

In addition, DFID maintains an open access policy (DFID Research Open and Enhanced Access Policy V1.1[20] with accompanying implementation guide[21]), and Digital policy[22]. The open access policy addresses research that DFID funds, while the open data strategy extends beyond research data to encompass other types of data, such as financial, procurement, operational, and results data (e.g. IATI).

The open access policy aims to increase the uptake and use of findings from DFID-funded research by increasing the number of research outputs that are mandated to be open, and increasing the accessibility of outputs. The policy requires researchers to deposit raw or derived datasets in any suitable and established open access repository within 12 months of data collection, and for researchers to retain and provide the raw datasets for free on request for a minimum of 5 years after programme completion. Researchers are instructed to use the Directory of Open Access Repositories (OpenDOAR) to locate a suitable repository

In addition, metadata for all outputs must be published as open data in the Research for Development[23] repository, while researchers are instructed to use open licences, favouring CC-BY. Research for Development is DFID's own research repository, a free-to-access online database containing information and outputs from research programmes supported by DFID.

19 https://data.gov.uk/sites/default/files/DFID%20Open%20Data%20Strategy.pdf
20 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/181176/DFIDResearch-Open-and-Enhanced-Access-Policy.pdf
21 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/181177/DFIDResearch-Open-and-Enhanced-Access-Implementation-Guide.pdf
22 https://www.gov.uk/government/publications/department-for-international-development-digital-strategy-2012-to-2015/dfid-digital-strategy-annual-update-october-2014
23 https://www.gov.uk/dfid-research-outputs

United States Agency for International Development (USAID) introduced an open data policy in 2015, referred to as 'ADS 579'[24]. In contrast to DFID's focus on transparency and accountability, USAID's policy focuses on the value of open data for decision-making during strategic planning, design, implementation, monitoring and evaluation. This policy provides guidance for complying with data sharing requirements under USAID's Evaluation Policy (January 2011), outlines USAID's approach towards implementing the President's Executive Order (EO 13642) which makes government data open by default, and establishes the policy directives, procedures, roles and responsibilities governing data management, and offers guidance for publishing and maintaining data on USAID's central open data repository, the Development Data Library.

Under the policy, implementing research partners are required to submit research and underlying datasets to the Development Data Library (DDL)[25], in machine-readable format. Published datasets also appear on Data.gov[26].

Implementation of ADS 579 is balanced by USAID's guidance on managing privacy concerns (ADS 508). ADS 508 outlines the organisation, functions, policies and procedures around safeguarding personally identifiable information (PII). It is understood at the time of writing that USAID is also developing an open access policy.

The **Bill and Melinda Gates Foundation (BMGF)** by contrast, as a non-public agency, does not have the same obligations as DFID and USAID with regard to public accountability. Nonetheless, BMGF is committed to transparency and data as a public good, believing funded research should be promptly and broadly disseminated. As the open access FAQ states, *"transforming the lives of the world's poorest people will require the collaboration of many partners, and it is crucial that they can access and use research without restriction."*

In 2015, BMGF introduced an Open Access policy[27] which applies to all new agreements as of 1 January 2015. The policy applies to research publications and the underlying datasets, which should be deposited in specified repositories (e.g. PubMed Central) and published under a CC-BY 4.0 licence. Significantly, in the case of multiple funders for the same programme, BMGF's open access policy must still apply. However, there is scope for discussion with other funders to address any concerns.

When initially introduced, BMGF allowed for a two-year transition period during which grant recipients could embargo their work for 12 months. But as of 1 January 2017, anyone who receives funding from the foundation must make their research and underlying data available immediately, for example, by publishing it in an open-access journal or depositing it in a public repository.

24 https://www.usaid.gov/ads/policy/500/579
25 https://www.usaid.gov/data
26 https://www.data.gov/
27 https://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy

Observations from applying an open data policy review

From ODI's guide "How to write a good open data policy"[28], the key elements of a good open data policy are: clear guiding principles; data licensing and reuse rights; identifying and prioritizing data for release; privacy and ethical considerations; data publishing standards; engaging with reusers of research data; internal use of third-party data; concrete commitments and metrics; and an open policy process. After comparing the donor policies against these elements, we observed the following strengths and gaps.

○ no mention of the criteria ◑ mention of the criteria, but without much detail ● detailed description of the criteria

Policy principles and overview

A strong policy should contain a clear definition of closed, shared and open data[29], outline the benefits of open data, describe the type of data suitable for release, and make a clear declaration of underlying principles and relevant legislation.

Policy principles	BMGF	USAID	DFID
Clear definition of closed, shared and open data	○	◑	○
Benefits of publishing and consuming open data to the organisation	◑	●	●
Description of the types of data that the organisation collects, stores, and releases	◑	●	●
Reference legislation, policies and best practices that are relevant to the application of the policy	◑	●	●
Declaration of the principles that underpin the policy	●	◑	●

We found:

- All policies contain a strong vision of the benefits of publishing and consuming open data for the organisation, but with different areas of focus. For instance, DFID focuses on transparency and accountability, USAID on data for decision-making, and BMGF on advancing scientific endeavour and innovation.
- None of the policies contain a clear definition of 'open data', although USAID categorises data in terms of access rights (public, restricted public, non-public) which mirrors the ODI's data spectrum[30].
- Most policies describe a list of data types intended for publication, but do not provide a comprehensive list of datasets otherwise collected and stored by the organisation – perhaps because the volume is too large or evolving. For example, USAID mentions

28 https://theodi.org/guides/writing-a-good-open-data-policy
29 https://theodi.org/blog/closed-shared-open-data-whats-in-a-name
30 https://theodi.org/data-spectrum

performance monitoring data, surveys, research data and supporting documentation. It is unclear how the list of data types will be updated to take into account new data assets.

- Although some policies provide strong links to domestic legislation for the donor's home country, e.g. USAID, there is scope to situate these policies in a broader international context by connecting to international best practice, standards and principles e.g. Open Data Charter principles, or to national legislation for the countries in which the donors operate.

Data licensing and reuse rights

Clear recommendation of the default open licence[31] under which data is to be released is important to maintain confidence and clarity on how data covered by the policy can be reused. There should also be clear instructions around obtaining rights to publish throughout the entire data production chain – from collection through to publication and use

Data licensing and reuse rights	BMGF	USAID	DFID
Clear recommendation of the open licence under which data is to be released			
Process for clearing rights to publish open data			
Open data embedded in procurement processes			

We found:

- Licensing practice is generally strong. DFID provides that all information published on the DFID repository and data.gov.uk[32] will be in reusable formats and licensed under the Open Government Licence, and their Open Access policy encourages researchers to use open licences such as Creative Commons (CC-BY). BMGF similarly recommends CC-BY 4.0 as it applies to research publications, but it is not explicit regarding the underlying datasets.
- USAID by contrast applies a Creative Commons Attribution-No Derivatives 4.0 International License to data published by implementing research partners on the Development Data Library[33]; by definition this is not an 'open' license because it limits certain kinds of usage.
- The policies vary in clarity when ensuring that rights to publish are properly cleared and understood including by implementing research partners during the data collection process. USAID refers to required clearances during the data publication process, but does not clarify requirements during the collection stage. BMGF's policy is similarly silent on rights during data collection. It may be that guidance during data collection is covered during the research design process.

31 https://theodi.org/guides/publishers-guide-open-data-licensing
32 https://data.gov.uk/
33 https://www.usaid.gov/data

Identifying and prioritizing data for release[34]

Outlining how data is inventoried, reviewed and then prioritized for release helps data users understand which data the organisation must make available, and hold that organisation to account. From the donor's perspective, publishing a data inventory or assets register is good practice that can help to inform new data requests and understand patterns of user demand. From the *implementing research partner's* perspective, having a process to understand user demand for their data can help to motivate open publication, and prioritize where to invest in data cleaning and formatting.

Identifying and prioritizing data for release	BMGF	USAID	DFID
Process for prioritizing data for release, e.g. based on user feedback, FOI requests, etc	○	○	◑
Inventory of internal data assets to help drive the data release process	○	●	●
Ongoing process for prioritizing release when new data assets are created	◑	◑	○
Process by which data will be released including decision points and risk assessments	○	●	◑

We found:

- With respect to donor-generated data, the process for prioritizing datasets for release based on user demand is unclear from the majority of policies. DFID's open data policy does, however, express an intention to publish "further information on common requests made to DFID under the Freedom of Information Act on our website".
- Positively, DFID maintains an internal information assets register while USAID lays out an ongoing process for contemplating and reviewing new data assets for publication as they are created. The process of evaluating new data sources is not clear in BMGF and DFID policies.
- The process by which data is released is not always made clear from the policies themselves, especially regarding key decision points and conducting risk assessments. USAID's policy rates very highly in this area, with clearly defined roles and responsibilities for conducting risk assessments.
- With respect to data generated by implementing research partners the feedback loops with donors and potential data users are not outlined. It could be helpful for researchers to understand more clearly the sources of demand for their data

34 See for example https://theodi.org/guides/prioritise-open-data-to-drive-global-development

Privacy and ethical considerations

Clear directives regarding the treatment of personal data, data protection, and responsible data use during all phases of collection and sharing are important to build trust, and to ensure that private information is not released by mistake. Mitigating activities such as conducting privacy impact assessments, ethics reviews, and applying anonymization should also be outlined.

Privacy and ethical considerations	BMGF	USAID	DFID
Stipulation that personal data should not and will not be released as open data, unless there is either consent from affected parties or other legitimate basis for its release	●	◑	◑
Need to anonymize or aggregate data prior to its release	○	◑	○
Reference to relevant data protection laws and standards that relate to the collection and subsequent sharing of data	○	◑	◑

We found:

- Commitment to safeguarding personal data was generally quite strong. BMGF contains an explicit statement that personal data should not and will not be released as open data, unless there is consent from affected parties, while USAID refers to the organisation's privacy policy (ADS508)[35], which provides substantive guidance on underlying principles, and mitigating steps such as conducting privacy impact assessments.
- Directions regarding the anonymizing or aggregating of data prior to its release in the situation of personally identifiable information (PII) was not uniform, which suggests this might be an area where practices diverge. USAID's policy explicitly mentions a process for redacting data that contains private information about individuals before publication.

35 https://www.usaid.gov/sites/default/files/documents/1868/508.pdf

Data publishing standards

Open data should be available in a machine-readable format. In addition, data should be made understandable through good-quality metadata, with a preference for open standards to encourage wide reuse.

Data publishing standards	BMGF	USAID	DFID
Human and machine-readable formats, with a preference for open standards	◑	◑	●
Metadata and supporting documentation	◑	●	●
Measuring quality of publication against industry best practices	◑	○	●

We found:

- All policies are technically strong on promoting machine-readable formats and open standards, e.g. DFID explicitly refers to adopting the IATI standard. However, data publishing could be strengthened by referring to more open standards including those specific to sectors, e.g. open contracting for public sector procurement, AGROVOC[36] or GACS[37] terms in agriculture.
- All policies contain strong statements regarding providing metadata and supporting documentation. For example, DFID promotes the Open Archives Initiative Protocol for Metadata Harvesting[38] for sharing research data between archives, while USAID requires supporting documentation, submission of a codebook or data dictionary, and metadata for each programme.

36 http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus
37 http://agrisemantics.org/gacs/
38 https://www.openarchives.org/pmh/

Engaging with users of research data

Working with external stakeholders, such as potential data users, can help to guide the release and quality of data, and ensure it can be easily used. A strong policy should outline how data users can engage with the organisation to request data for release, and outline channels for users to provide feedback, e.g. on quality issues. Engaging with data users can also help donors and implementing research partners to understand how data is actually being used, and to assess the impact of their investment in open data

Engaging with users of research data	BMGF	USAID	DFID
Process for users to request and help prioritize data for release	○	◑	○
Channels for users to provide feedback	○	◑	○
Wider strategy for engaging with reusers throughout policy implementation	◑	○	●

We found:

- Across the board, there are weak references to engaging with data users – including ongoing channels for feedback. Beyond technical mechanisms available on data publication platforms like the Data Development Library ('DDL') or Devtracker[39], there appears to be no obvious ongoing engagement strategy to invite ongoing feedback or assess the impact of data releases.
- DFID's open data policy in 2012 referred to a plan for engaging with users of DFID research data during early phases through introducing pilots, digital tools and social media, consultations on the usability of the programmes database (now 'Devtracker'), and surveys asking for customer feedback from public enquiries. However, it is unclear what the current engagement strategy is.

39 https://devtracker.dfid.gov.uk/

Approach to reusing data

An open data policy should ideally also promote reuse of open data sources, and use of reliable third-party data that is appropriately licensed. In other words, it should encourage implementing research partners, and donor staff, to use open data sources within their own work.

Approach to reusing data	BMGF	USAID	DFID
Guidance on how to identify whether third-party open data is appropriately licensed for reuse	●	●	◑
Suggestions for how to find and source reliable, high-quality data	◑	●	◑

We found:

- All policies are quite weak on providing guidance about how to identify whether third-party data is appropriately openly licensed for reuse. DFID's open access policy, however, does specify that researchers are responsible for ensuring they have the necessary permissions to make material available, and are directed to respect third-party copyright, licensing and embargo policies.
- Beyond this, there are minimal instructions on how implementing partners might find and source reliable, high-quality data for reuse from government or industry portals, or how they might ensure that data they procure through third parties has appropriate rights and licences. DFID and BMGF also point implementing research partners towards where they might publish, and presumably find, open data research, e.g. PubMed[40], OpenDOAR[41].

40 https://www.ncbi.nlm.nih.gov/pubmed/
41 http://www.opendoar.org/

Concrete commitments and metrics

Laying out specific and measurable commitments can help donors to measure success of the open data policy over time, and for stakeholders to hold funders to account for their promises. Having clear time-bound goals at a programmatic level can also incentivise implementing research partners to comply. It is critical that donors clearly stipulate to grantees what their expectations are, for example, regarding their budgeting approach to ensure open, quality-assured data.

Concrete commitments and metrics	BMGF	USAID	DFID
Plan for improving capability, including further guidance and training for staff			
Concrete commitments to the publication of particular open data within a timeframe			
Commitments about the quality of publication of open datasets			
Commitments to maintained datasets over time, and for a set period			

We found:

- All donors make high-level commitments toward improving capabilities of data publishers and researchers, including further guidance and training for staff. DFID's open data strategy contemplates capacity building for staff to use data, while supporting citizens, legislatures and media to use data DFID publishes. Information Managers are also purportedly appointed to train staff and monitor how guidelines are being met. How this is rolled out to support implementing research partners is unclear.
- BMGF's support to implementing organisations and researchers extends to paying individual article processing charges and fees to the extent required to enable grantees to comply. The Foundation checks and tracks compliance through Chronos[42], a new service to help researchers manage the process of publishing under the policy's terms.
- DFID also makes strong public commitments around the quality of published open datasets and introduces an ambitious 80% target of open datasets underlying research publication

42 https://chronos.gatesfoundation.org/

Open policy

A policy promoting transparency should ideally also be 'open' itself, for example, for describing the processes by which it will be reviewed, share lessons learned, and invite feedback from stakeholders. This includes providing an opportunity for research partners to contribute towards improving the policy and process of implementation

Open policy	BMGF	USAID	DFID
Defined timespan	○	○	●
Participatory process for policy development	○	○	○
Process for revising policy and providing feedback	◑	○	○
Responsible party for the policy identified?	○	●	◑

We found:

- In general, the policies were quite silent on this point. DFID refers to a process of participatory policy development which informed their original policy design.
- However, it is unclear otherwise how often the policies will be reviewed, how learning will be shared, and what opportunities there are for implementing research partners and other constituents to contribute to new iterations within defined timeframes

SUMMARY FROM POLICY REVIEW

Good Practice

From this policy review we identified **good practice** in the following areas:

- Ambitious visions and widespread recognition of the benefits of publishing and consuming open data to the organisation, and global development sector (ALL)
- Identifying the types of data to be released as open data, with targets around quantity, quality and publishing to open standards (DFID, USAID)
- Clearly defined roles and responsibilities guiding the release of open data, and processes for dealing with personal data (USAID)
- Commitments toward providing support and capacity building to staff (ALL), and through providing additional resource to lower the barrier of adoption for implementing partners (BMGF)
- Clear guidance surrounding the technical elements of open data (machine readable, metadata, licensing) (ALL).

Gaps Identified

We identified the following **gaps**:

- Lack of shared definitions covering key terms ('open data') and publishing principles such as acceptable justifications for not publishing data
- Little guidance around dealing with personal data during data collection phases, mitigating steps to safeguard privacy, and making ethical judgements throughout the data production chain
- Low focus on engagement to prioritize future data releases, capture stakeholder feedback, collect use stories, and assess impact
- No identifiable strategy for promoting reuse of data either internally by implementing research partners and donor staff, or externally by other data user groups
- Absence of clearly defined and time-bound metrics to measure quality of compliance, value for money, and impact of open data and research publication
- Absence of feedback mechanisms to promote ongoing participation, share learning, improve implementation, and revise policy
- Little guidance on budgetary expectations and line items to ensure open, quality-assured data.

Summary Recommendations

To **strengthen and harmonise policies**, donors could:

- **Adopt common definitions of key terms** (e.g. 'open data') or shared justifications to avoid different interpretations by researchers
- **Share approaches towards dealing with ethical considerations** such as justifications for non-publication, and provide tools to make it easier for compliance, e.g. data spectrum, anonymization guide, FAIR Principles[43], ethics canvas[44], GODAN's responsible use of data recommendations[45]
- **Publish yearly implementation plan and metrics, regularly review compliance by implementing organisations**, e.g. via data quality review and inviting feedback
- **Introduce practical projects and incentives to promote data reuse and innovation**, thereby providing evidence of value and impact to motivate compliance
- **Invite feedback from implementing research partners** and take measures to iterate policy and tools based on feedback.

43 https://www.force11.org/group/fairgroup/fairprinciples
44 ODI Data Ethics Canvas, forthcoming 2017. See also Responsible Data Forum (https://responsibledata.io/), and Oxfam's Responsible Data Use Kit (http://policy-practice.oxfam.org.uk/our-approach/toolkits-and-guidelines/responsible-data-management)
45 See http://www.godan.info/documents/responsible-data-agriculture

PUTTING POLICY INTO PRACTICE: LESSONS FROM IMPLEMENTATION IN FIVE AGRICULTURE PROGRAMMES

We recognise that, while fundamental, a strong organisational policy is not in itself a guarantor of good open data practice. It must be supported by appropriate resourcing, capacity building and leadership support to promote culture change across the organisation, implementing partners, and external stakeholders including data users.

In this section we consider how donor open data policies are being rolled out in practice in the context of five jointly funded agriculture research programmes. The insights are derived from interviews with strategists, programme managers, and implementing researcher partners (see definitions in Section 2). We were particularly interested in exploring issues related to:

- **availability, accessibility and quality of data publication from the selected programmes;**
- **emerging benefits and examples of data use;**
- **available guidance and support to implementing organisations and researchers**; and
- **coordination between partners.**

Common challenges during implementation and potential solutions will be addressed in Section 5.

WHAT IS THE AVAILABILITY, ACCESSIBILITY AND QUALITY OF OPEN DATA?

Availability

On the whole we found a significant volume of data is being collected, created and published by implementing research partners. This is particularly the case for the more mature programmes like the Cereal Systems Initiative for South Asia (CSISA)[46] and NextGen Cassava[47].

There is a wide variety in the types of data being generated and used across the programmes, in many cases supporting large research programmes. For example in the CSISA programme, there is evidence of spatial data[48], genomic data[49], socioeconomic data[50], opinion survey data[51] and environmental data. In NextGen Cassava, similar types of data underpin research into cassava virus disease diagnostics, for example into diagnostic tools[52] and real-time crop surveillance[53].

There is also a wide variety of data collection methods on display. For instance, the JPAL Agricultural Technology Adoption Initiative (JPAL ATAI)[54] is funding 15–30 randomised control trials and 10–30 diagnostic data collection exercises. There is evidence of vast amounts of data being used to inform secondary analysis of research[55] and findings[56] on the ATAI website.

46 See http://csisa.org/research-publications/
47 See http://nextgencassava.org/resources.html
48 See http://csisa.org/wp-content/uploads/sites/2/2014/06/Jain-et-al-2016_Mapping-Smallholder-Wheat-Yields.pdf
49 See for example http://csisa.org/wp-content/uploads/sites/2/2014/06/Mahajan-et-al-2014.pdf
50 See for example http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0165924#sec006
51 See for example https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/26930
52 See for example http://wiredspace.wits.ac.za/jspui/bitstream/10539/14018/1/Cyprian%20DISSERTATION%20Final%20version.pdf
53 See for example http://www.gcp21.org/wcrtc/ppt/session05presentationposters/S05-06.DanielMutembesa.SIGNED.ID4663.pdf
54 https://www.atai-research.org/
55 https://www.atai-research.org/our-research/
56 https://www.atai-research.org/emerging-insights/

Accessibility

However, a large volume of data does not necessarily mean it is accessible. The key question is whether data reusers can easily discover, access and use that data. This is dependent on how well data is referenced on the programme website and documents, as well as if the data itself can be accessed and downloaded in a reusable format.

In terms of discoverability, we found few specific mentions of where to find data or links to datasets. Only the CSISA programme has an explicit webpage for data[57] and this only hosts one dataset. Some of the programmes have links to research publications[58], but none have links to the datasets underpinning the research. There are no lists of datasets that are or will be generated by the programme, nor general statements of intent to publish data. This makes it very difficult for data reusers to search and discover available data.

Some programmes publish data as 'supplementary material' in open access journals, for example, the mapping of resistance to cassava mosaic geminiviruses[59] and genetic mapping in the clonally propagated cassava[60] from the Cassava portfolio programmes. However, the majority of content appears to be outcome or summary data rather than the underlying datasets. In other cases, the data underlying research publications and data being generated through the activities of the programme are discoverable directly through the programmes' websites – as is the case of the phase 1 baseline household survey data[61] in the CSISA programme and the CassavaBase[62] platform.

In some cases, data is also published on third-party websites – for example research into farmer's preferences for drought tolerance in hybrid versus traditional rice[63] from CSISA is made available on the CYMMYT's Dataverse[64] platform as the primary repository, which is linked from their dedicated programme website. Publishing in multiple places ('distributed publishing') can increase discoverability. However, such examples were difficult for us to discover because they are not linked back to the programme website or the journal publications.

Logins can can create barriers to access in some scenarios. The CassavaBase platform requires users to sign up for an account to access some of the features and data. While it is usually acceptable to ask users to login to access open data under most definitions, it can discourage users from accessing the data. This may be due to be unclear communication around the signup process. However, logins can also show who is using the data and how it is being used, which is a crucial component of many data strategies.

Quality

Using the ODI's Open Data Certificates[65] as a guide, we examined the quality of data publication in terms of the licence, format, metadata and associated documentation.

Licence

Despite clear policy statements regarding licensing, we discovered a mixed approach in practice. In the majority of cases, research publications and programme documents were made available under Creative Commons licences. However, it is not always clear whether an open licence, or any licence, is applied directly to some of the programme content. In some case they are not open access[66], and in others they are published under restrictive (non-open) Creative Commons licences, such as non-commercial non-derivative[67]. This limits what people can do with the data.

57 http://csisa.org/resources/csisa-phase-i-baseline-data/
58 See for example http://csisa.org/research-publications/
59 http://www.sciencedirect.com/science/article/pii/S0168170213004735?via%3Dihub#upi0010
60 https://dl.sciencesocieties.org/publications/cs/abstracts/54/4/1384
61 http://csisa.org/resources/csisa-phase-i-baseline-data/
62 https://www.cassavabase.org/
63 See for example https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/26930
64 https://dataverse.harvard.edu/
65 See https://certificates.theodi.org/en/about/badgelevels
66 See for example http://www.sciencedirect.com/science/article/pii/S0065211317300172
67 See for example http://csisa.org/wp-content/uploads/sites/2/2014/06/APSIM-in-Asia-Gaydon-et-al-2017-FCR.pdf

For data to be clearly open data, there should be a clear statement of the open licence and a link to the full details of that licence. In some cases, programmes are making data publically available on their websites without any reference to an open licence, for instance the CSISA baseline data[68] and CassavaBase[69]. Where data was published within open access journals, it is unclear whether the licences referenced apply to the underlying datasets or just the article. Not having an open licence means that the release, while publicly available, has a barrier to use, as researchers cannot be confident in how they can use the data they can access, and what they can use it for.

Format

Technical barriers can also impede reuse and therefore the impact of data. For instance, proprietary formats can stop users accessing the data in their desired software package or manner. We also found some other examples of inappropriate formats. For example, the accompanying codebook for CSISA is published as a .pdf file containing tabular data, which makes it much harder to process by machine than a spreadsheet or .csv file would be.

Metadata and documentation

Providing structured metadata and associated documentation provides reusers with context and confidence to make the best use of the data and enables the data to be indexed accurately, improving discovery. Metadata includes properties such as the date published, frequency of publication and contact details, and documentation such as the original survey used to collect data. These enable users to understand when, why, how and for what purpose the data was produced and help to encourage appropriate reuse of data.

We found varied approaches to dealing with metadata from the selected programmes. The baseline survey data published on the CSISA site[70] provides an example where there is lots of supporting information on the page but there is no structured, machine-readable layout. The same data[71] on the Dataverse platform is published alongside clear structured metadata[72]. Similarly the CassavaBase[73] provides good documentation and structured metadata surrounding the different datasets.

EMERGING BENEFITS AND EXAMPLES OF DATA USE

Donors and implementing research partners are already beginning to realise the benefits of publishing open data to the agriculture sector, and to their own organisation.

One of the earliest benefits for donors is the positive reputational effect of publicly committing to open data. While initially motivated by transparency and social accountability, donors are beginning to recognise the value of open data to improve investment decisions and drive innovation. As one strategist from BMGF discussed:

> *"...our first and foremost interest lies around getting value for money and making those data sets available, spurring innovation, and in the long term improving the quality of that data."*

Within donor organisations, opening up data from research programmes has led to greater internal access to valuable new data assets. Reported examples of internal use included scoping out research directions for new programmes, and avoiding duplication of data collection during baseline research phases.

68 http://csisa.org/resources/csisa-phase-i-baseline-data/
69 https://www.cassavabase.org/
70 http://csisa.org/resources/csisa-phase-i-baseline-data/
71 See https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/22220
72 However, it is understood that at the time of editing, the CSISA project website will link to the CIMMYT Dataverse for accuracy and reliability.
73 https://www.cassavabase.org/

One programme manager from USAID also reported an intention to use geospatial data to support more rigorous monitoring and evaluation on crop yield.

The process of introducing an open data or open access policy also prompted donors to reflect on their internal approaches towards data management and use. Several donors reported an intention to build internal capacity of staff to use data being produced via research programmes more effectively for decision-making.

From our surveys of implementing research partners, we found widespread evidence of use and benefits of open research. Uses included measuring the reach of activities; accessing results from other organisations to feed into new research; accessing openly available code and tutorials for capacity building; replicating earlier studies and using existing datasets to inform long-term follow-up studies (5, 10 or 15 years after the original data was collected); and challenging the existing evidence base.

Reported benefits included improving the speed, efficiency and efficacy of research (e.g. via CassavaBase, which serves as a 'one-stop shop' for cassava researchers); triangulation and comparison with multiple sources of data to validate findings; ensuring the reproducibility of results; allowing researchers to extend the analysis of earlier studies or conduct a meta-analysis; creating transparency around farmer organisations; and providing more opportunities to smaller, less well funded research teams to access valuable data. As one researcher summarised:

> *"Much more progress comes when we share our data...sharing how I've done things helps others to learn from my mistakes and gets them 'up to speed' faster. No single person has ALL the knowledge or all the time anymore. So the largest benefit is enabling more people to contribute."*

A number of implementing research partners similarly reported flow-on benefits in changed behaviour. Being compelled to publish open data under the donor policy promotes more collaboration between implementing research partners and the wider agriculture research community. The result is an expanded pool of projects and organisations that want to join and share data. Researchers know more about what experiments are being performed elsewhere, can analyse historical data that was previously unavailable, and pass on the same benefit to others. As one researcher from the NextGen Cassava programme commented:

> *"In breeding communities such as cassava, [open data] provides a huge impetus for other researchers to join and be part of the community. It also prevents useless repetition of experiments and better use of research dollars."*

Increasing the size of collaborative networks can also create new funding possibilities, as one expert from CGIAR observed from their organisation's experience.

Data use snapshots

One implementing research partner from the New Alliance ICT Agriculture Extension Fund project reported **using third-party open data (census and population data) to measure the reach of radio programmes** and estimate broadcasting equipment needs across a population, disaggregated by age group, gender, and rural/urban clusters.

Within the NextGen Cassava programme, research partners were able to **draft a joint paper using data from three different breeding programmes** (NaCRRI, IITA and NRCRI). Another **biotech research collaboration** was enabled between multiple sites in Nigeria and Kenya, as well as a **publication on best practices for regulation of transgenic crops** in Africa – all underpinned by open research data. Finally, researchers were able to **verify the genomic selection accuracy for specific agronomic traits** in cassava through combining and aggregating multiple datasets.

> *"A previous study assumed a specific sample was representative of cultivated cassava in their comparative genome analysis against a putative progenitor species. Using their data and placing it in the genetic context of a larger sample of 60 cassava individuals, we were able to show that this original sample was, in fact, not a cassava, but another species entirely – showing that their research conclusions were incorrect."*

Although it is relatively early days in the JPAL ATAI programme, there are several examples in which JPAL affiliated researchers have conducted **long-term studies using existing open data**. One example provided was a study on early childhood stimulation in Jamaica[74].

Limitations upon further use

It is important to note that several programmes like the Agriculture–Nutrition Impact Studies Programme, are in the early stages of data collection and release, which suggests more examples of benefits and reuse will emerge over time.

There is a **risk data will be underutilised** if there is no effort to promote reuse by engaging potential data user groups. One researcher from the New Alliance ICT Agriculture Extension Fund reported that right now they are only publishing visualisations. A lot of potentially valuable data about interactions with Senegalese farmers (e.g. the alerts they receive, the requests they place, the channels they place them through, how many calls and SMS received, etc.) could be published following some efforts of anonymization and aggregation but is not because there is no perceived demand.

Part of the problem is **lack of systematic follow-up on data use stories** to motivate reuse by other researchers and the donors themselves. As one interviewee from USAID reflected:

> *"There might have been benefits – it's just that we don't have any way of tracking or reporting on them. We're at the mercy of anecdotes as they pop up, and we just haven't had the capacity to go in search of those anecdotes. So a lot of effort has gone into it, but we're not always cognizant of what has come out on the other side."*

This suggests a need to promote what data is out there (and where to find it) to those who might want to use it. As one programme manager from USAID suggested:

> *"a lot of these organisations are actually sharing the data and there's a great deal of data out there, it's just a matter of disseminating more widely, where people can go to find that data."*

74 https://www.povertyactionlab.org/policy-lessons/education-health-labor-markets-crime-violence-conflict/early-childhood-stimulation

WHAT GUIDANCE, SUPPORT AND TOOLS ARE AVAILABLE FOR USE BY IMPLEMENTING RESEARCHER PARTNERS?

Data policies need to be backed up by guidance and tools which help programme managers and implementing research partners meet the desired policy goals. In our interviews we explored which tools and guidance were most useful, and what might be missing. Guidance and tools also offer an opportunity for donors to align their approaches. We examined areas where donors could co-create resources and share best practices to create benefits through greater efficiencies.

Guidance

We found great diversity among the operational guidance offered to programme managers and implementing research partners. Processes for dealing with personal data, such as anonymization, require guidance to ensure consistency of approach. Without clear instructions on sensitive issues like these, donors might be unable to guarantee that their open data releases meet wider policy and legal requirements around data protection and privacy.

Many interviewees mentioned there was **little guidance around data ethics**, particularly during planning for data collection. One respondent from DFID recounted that on several occasions they had been prevented from sharing valuable data because a consent form eliminated the prospect:

> *"I'm sure in hindsight that we would want to have a much more developed understanding of the process for data collection and dissemination, so that we could head off those kind of challenges."*

Several implementing research partners also noted that it was crucial to have guidance around **common standards, naming conventions and vocabularies for data**. For example:

> *"A hugely important factor is common vocabularies/ontologies for phenotypic data, and open data is basically useless if these standards are not defined."*

Tools

Tools and templates enable programme managers and implementing partners to efficiently and effectively comply with policy.

Data management plans were identified by multiple programme managers and implementing research partners as an important tool. We found all donors have experience in using some kind of template. For example, USAID asks implementers *"to submit their data management plan [alongside the proposal and] we follow it on a yearly [basis] or every six months or depending on many different situations."* Likewise, DFID now requires grantees to submit and maintain a data asset register, while BMGF is also building a similar process into their grantmaking. Aligning these processes and sharing best practices could help donors to realise efficiencies across jointly funded programmes and beyond.

Another reported need was a consistent strategy for prioritizing which datasets to publish and how to **validate the quality of publication**. While general commitments regarding broad types of data to publish are mostly laid out in donor policies, a tool for prioritization and validation of individual datasets (such as Open Data Certificates[75]) is not. This need arises from a desire to ensure outputs are useful to people who want to reuse it. As one programme manager from USAID put it:

> *"you don't want to dump all the data [on the DDL], you want to make sure that it's the right data that can be interpreted and that can be useful."*

Several implementing research partners also reported a desire for **clear formats for data uploading**:

> *"it must be searchable, all participants must agree to it, and the data must be curated."*

Ongoing support

Interviewees also raised the need for ongoing support, for example the ability to ask questions as challenges arise (helpdesk style). The interactive support that is available for implementing partners seemed to vary across teams, programmes and donors.

Much of the support available to implementing partners appears to be more informal and delivered through programme managers, although in some cases there are clear issue 'owners'. For example, USAID has introduced the role of 'data custodian', who looks after data management plans and policy across portfolio of programmes. However, even data custodians are being stretched. As one donor put it,

> *"I think we support a lot of data collection but I don't know if we do a lot in terms of supporting the 'how do we improve on these processes' and it's left up to the researchers who are collecting the data".*

75 https://certificates.theodi.org/en/

HOW DO DONORS MANAGE POLICY COORDINATION AND COMPLIANCE?

Implementing open data within jointly funded programmes requires coordination and compliance with multiple donor policies. Developing a more harmonised approach could improve process efficiencies, while providing an opportunity for researchers to benefit from sharing knowledge and technical capacity.

Coordination is an issue that involves time and effort to resolve. Navigation between donor policy requirements is currently happening on a case-by-case basis, based largely on existing relationships and trust between the three donors we investigated. However, this approach might not translate to a wider range of funders, for instance Research Councils or universities.

Techniques we identified to promote policy coordination and compliance include:

- **Memorandum of Understanding** (MOU) between funders of joint research programmes. The MOU provides a framework for agreeing upon policies and processes for programme management, reporting and implementation to minimise any duplication of effort, e.g. double reporting by implementing research partners (e.g DFID and BMGF);
- **Embedding policy directives into contract** of research awards, specifying which policy is to be followed, and the process of data sharing and management between donors; and
- **Informal agreements**, for example funder X agrees for a researcher to publish to the World Bank micro data library, or to comply with Y template for submitting a data management plan.

Some challenges were identified arising from the different approaches towards funding open-data-related activities, with some expecting costs to be built into research design proposals, while others (BMGF) offer additional core funding to cover publication costs. Other issues occurred when donors changed their organisational policies midstream, for example regarding maintaining an asset register. However, it was generally felt that this process was managed well (especially by BMGF) due to the high level of contact they have with implementing research partners in the field.

Nonetheless there was widespread recognition of the need to achieve greater consensus in order to reduce the ‘burden’ of compliance for researchers. As one programme manager from USAID mused:

> *“...it would be nice if all the donors, USAID, DFID, Gates … can have one form [of data management plan] so that implementers don’t have to have different standards for different organisations.”*

To make it easier for implementing researcher partners, one donor strategist suggested agreeing on common data management tools during the design phase. One example of this came from the CSISA programme, where USAID and Gates had a conversation with the implementing partners to agree on which reporting formats and data would satisfy both funders.

The benefits of donors moving towards common tools include more time for addressing the overall data infrastructure, incentives for sharing data, and building up the capacity of researchers and staff to analyse and use data for evidence-based decision making, which will impact the ecosystem on a broader scale.

SUMMARY OF POLICY IMPLEMENTATION

From interviews and desk-based assessment of data outputs from the five selected programmes we found:

- Emerging benefits of data use among donors and implementing researcher partners include informing research scope of new programmes, avoiding duplication of data collection, informing new research design, and increasing research collaboration networks. However, lack of engagement and monitoring of data reuse by donors mean examples are still limited and anecdotal.
- Although significant volumes of data are being produced, there were observed weaknesses in the availability, accessibility, discoverability and quality of data.
- Barriers to access programme data include failure to link to underlying datasets or to signpost where and what data is available, lack of clarity around licensing and use of restrictive licences, and in some cases, proprietary formats and login requirements.
- Diverse approaches towards dealing with personal data and data management plans exist to help implementing research partners, but there is a common gap around data ethics guidance.
- Coordination between donors on policy compliance happens on a case-by-case basis. There is a desire to streamline the processes and adopt common tools and templates to reduce the burden of compliance for implementing researcher partners, for example common budget templates.

This suggests the need for donors to:

- **Actively collect data use stories to demonstrate value of research data** to potential data users, and motivate implementing researcher partners to adopt good open data practice.
- **Provide support to implementing research partners to improve data availability, accessibility, discoverability and quality**, e.g. via automated tools like Open Data Certificates. Provide clear guidelines on accessibility (acceptable standards, repository platforms etc.), format, licensing considerations, metadata schemas, readme files, codebooks, etc.
- **Share, and where possible, harmonise tools and templates** to make the process of open data adoption more efficient and consistent for implementing research partners, e.g. common data management plan templates, budget templates and allocation guide.

These findings confirm the gaps we identified in our policy review in section 3 (lack of shared tools and approaches, data ethics guidance, promoting reuse of data, and ongoing methods for monitoring implementation and capturing use stories), but also extended our findings in the area of data quality and the support that is required to help researchers improve data accessibility and discoverability.

OVERCOMING IMPLEMENTATION CHALLENGES

We identified a number of challenges or barriers towards further adoption of an effective open data policy in the course of our interviews. More often than not these were social challenges, rather than technical ones. They include capacity to implement policy, lack of incentive to publish openly, managing culture change, policy monitoring and maintenance, and making it easy to publish.

CAPACITY OF RESEARCHERS TO IMPLEMENT

Specific gaps in the capacity of donors and researchers to implement were reported around data quality, data management, strategy (e.g. creating data management plans), putting data to use, and understanding responsible data use. These reported gaps align with weaknesses in policy and the quality of existing data outputs (see Sections 3 and 4).

While capacity support needs to be provided at every level – including data managers, information specialists, and academic researchers, etc – there was a recognised need to build the skills of local research partners in low-income countries in order to share the benefits of open data and increase levels of data literacy. It was also recognised that domain experts might not necessarily have the skills of a data scientist, so may need extra support e.g. around good metadata.

Not surprisingly, the capacity to manage and publish open data was higher among implementing organisations that maintain their own open data policy, and already have a strong open culture. As one programme manager from USAID working on the CSISA programme commented: *"...they have always been generally open, they publish their papers faster, they put their data out faster."* It is more of a challenge, however, to bring on board new research partners/institutions who are less familiar with open data or collaborative ways of working.

Although all donors have committed in principle to providing capacity support, specific activities backed by resources are not outlined in the policy. We identified many different ways that donors are currently providing support for implementation:

- **Peer network sharing** between information managers to learn from each other and contribute towards a marketplace of ideas on how to improve open data practice. Mechanisms for sharing include face-to-face and virtual (interactive) webinar series, small working groups, gatherings at conferences (CGIAR, USAID).
- **Data 'sprints' with researchers and data curators** to accelerate data publication. In one reported instance from CGIAR's International Potato Center (CIP), the goal was to publish 100 datasets over a short period, but they exceeded this target and reached 130.
- **Budgeting for training** local research partners within grant designs (DFID), and rolling out a standardised training for programme managers on the open data policy and what it is trying to achieve (USAID). Training was also held for NextGen Cassava implementing research partners to get organisations on the same page regarding naming conventions.
- **Feedback channels for implementing research partners** to explain policy directives, responsibilities, and respond to ongoing questions within proposals (USAID).

Building capacity across an organisation also requires maintaining core dedicated roles within the donor organisation, with responsibility for rolling out policy, monitoring implementation, and offering technical support around good open data practice.

Suggestions varied regarding what shape this might take. For instance, USAID appoints data stewards within the Bureau for Food Security, which is responsible for the 'Feed the Future' initiative[76]. Data stewards work across multiple programmes to ensure USAID policy requirements are embedded into contracts, provide guidance around what type of data should be provided to the DDL, and advise on how to deal with issues during implementation such as privacy, and how to manage different types of data. Even so, there is limited bandwidth to respond to all needs including quality control of outputs and improving processes with researchers.

One suggestion from CGIAR was to budget for part-time data curators to help with quality control, e.g. clean data, instruct implementing research partners on metadata schema etc. One interviewee from DFID also suggested a dedicated role for driving quality improvements in data publication, and advising researchers on where to publish to and how from a technical sense. This supports our findings from the data quality review (see Section 5).

INCENTIVES TO PUBLISH

A common issue raised in interviews was a lack of incentives for implementing research partners or a perceived mismatch between effort and gain in publication. This needs to be addressed at multiple levels: (a) changing the practice of academic publication; and (b) offering 'carrots' to individual researchers and teams to change their behaviour.

It was widely agreed that changing the nature of academic publication is highly ambitious and difficult because of the 'publish or perish'[77] imperative. The pressure to publish research in peer-reviewed journals can encourage researchers to hold on to data for competitive purposes. For publication in some journals, data must be embargoed until the final paper is published, which can disincentivise early release.

Changing the model for academic publication is outside the scope of this paper to consider. However, we found many practical ideas for incentives to encourage individual researchers and teams to modify their behaviour:

- **Evaluating the performance of implementing research partners differently**, recognising that open data publication is a professional achievement and a necessary skill, e.g. via key performance indicators (KPIs), performance reviews, evaluating research proposals based on ability to demonstrate previous open data publication, and promotion.
- **Engaging with researchers on the value of making their data open** and the importance of good data management, e.g. in terms of expanding their research collaboration network, generating research opportunities, and the added value of sharing to the agriculture sector.
- **Make data publication a deliverable of the programme, not just a by-product** and write it into conditions of the research grant ('data as deliverables' approach'). As one programme manager from USAID reflected:

76 See https://www.feedthefuture.gov/
77 https://www.theguardian.com/science/2011/sep/05/publish-perish-peer-review-science

> *"If data is identified as a deliverable...that raises a level of awareness throughout the preparation cycle but most importantly, it gets people thinking about future use of data and the need to potentially share this data with other partners working in the same areas of research."*

One research implementing partner from the New Alliance ICT programme agreed with the mixture of 'stick and carrot' approach:

> *"If projects were forced to publish data generated as part of the funding as open data, and if there was an effort by funders to centralise these data-sets, this could leverage demand and impact."*

- **Crediting quality data collection and publication.** Encouraging implementing research partners to publish underlying datasets, and applying a DOI to those datasets means others can access it easily, recognises the depth of effort in data collection and curation, and can contribute to academic reputation by allowing the impact of data to be tracked. For example, open peer-review platforms like F1000 enable a DOI to be applied to secondary research (articles) and the underlying datasets at the same time.
- **Encouraging or mandating researchers to publish in good open access journals** that require publication of underlying datasets with the research paper, and allow citation of 'pre-prints' (pre-peer-reviewed manuscripts) in future research grant applications to promote timely publication of datasets.

This may involve discussions with research partner institutions to encourage giving recognition to people who collect or enhance the quality of data for other researchers to use. This could help to ease fears of local research teams about the future use of their data and loss of 'ownership', and motivate people and teams to focus on improving the quality of and publishing datasets.

Lack of budget can slow down cleaning and publishing underlying datasets from publications. Budgeting for open data can be challenging, as research articles are often pulled together towards the end of the funded research period. If the research period has ended, donors are reliant on researchers releasing datasets in their own time (unless compelled to produce datasets within a defined period as required by BMGF). To manage this risk, several interviewees referred to the importance of embedding data publication activities within the research programme design over a longer timescale.

Alternatively, some donors cover costs associated with publication out of a separate core budget line (see the BMGF open access policy[78]) to ensure that the activity takes place. One CGIAR Center, Bioversity, also has an innovative budgeting practice to support dissemination by experimenting with a centralised 'pot' of funding which is available to draw down from even after the defined research programme period ends. From their experience, CGIAR also recommends providing clear budgeting guidance to assist with research design, e.g. stating what percentage of the grant budget should be allocated to support data management and publication.

78 https://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy

MANAGING CULTURE CHANGE

Several interviewees referred to difficulties in operationalising policies and fostering an open culture. This is a problem that reaches far beyond the agriculture sector, as noted by the Wellcome Trust in their research on embedding culture and incentives to support open research[79].

Challenges relate to addressing fears that implementing researchers have about making their data open and institutionalising open data as routine practice. This kind of culture change is difficult to achieve quickly and it takes time and resources to ensure that implementing research partners are following through with their data management plans.

Several implementing research partners reported that the major hurdle is often convincing all partners to cooperate, as opposed to any technical component. There is a need to address researcher fears by *"showing people the sky won't fall down if people do make their data open"* (CGIAR), while providing examples of tangible benefits researchers can gain from having timely access to others' datasets, including from other disciplines such as biomedicine and ecology, and the benefits for advancing the sector as a whole by making their own research available to others. As one implementing research partner from NextGen Cassava reported:

> *"Not all participants have the same culture of sharing. This can be learned through positive experience. There needs to be a monitoring aspect of this to ensure that all partners comply."*

There was a general sense across the three donor organisations that implementing research partners are normally comfortable making their research publications open as long as they have the funding to do so and have some choice about where they want to publish. Costs need to be budgeted for making the data available in a standard format, anonymized, and easily discoverable.

However, publishing underlying datasets behind research publications is more complex. Some researchers still feel a sense of threat, with concerns reported surrounding ownership of data, how donors define the point of 'final data collection', and attribution of datasets to the original researcher and data collection team. There is a general fear that if researchers publish data that they collect, they will lose the benefit to others who can make better use of the data. This reinforces the need for donors to offset fears through offering incentives, and building capacity to use open data effectively.

One USAID interviewee also noted challenges of obtaining timely datasets for publication due to embargoes for 'research publication purposes' or restrictions on the implementing research partner institution side (especially where a public international organisation (PIO) or university is involved).

Nonetheless, it was widely felt among donors that the effort and underlying value of the datasets involved is worth the time and resource investment *("that data might stop people dying").* One interviewee drew inspiration from the culture change within the PubMed domain. Transformation in that sector is being driven by funders insisting on open research and data sharing platforms, combined with a network of researchers who see the value of sharing data to solve a collective problem, e.g. worldwide antimalarial resistance[80].

79 https://wellcome.ac.uk/sites/default/files/embedding-cultures-and-incentives-to-support-open-research-oct16.pdf
80 https://figshare.com/articles/Learning_from_the_pioneers_lessons_about_data_platforms_drawn_from_the_WWARN_experience/4476308

One tactic for dealing with sensitivities and policy obstructions around open datasets is taking a phased approach, by introducing data sharing agreements between implementing research partner organisations even before any data is cleared for public release. This can help get around some roadblocks and make researchers feel more protected, while still giving access to those research partners who need it.

Data ownership and equity was raised as an issue for opening research in developing country contexts, for example where research teams responsible for collecting field data may not necessarily have capacity to process that data for analysis, or publish in the required format. Another potential issue, but which is beyond the scope of this research, may be where researchers gather lots of valuable data, but don't make it available or accessible to the communities that they have collected data about.

The role of senior leadership in institutionalising change within both donor and implementing research organisations was regarded as crucial. Senior leaders within donor organisations are expected to deliver clear communication around expectations, and coordinate organisation-wide efforts to support and incentivise implementation by implementing research partners. This high-level commitment to open data must also be backed by funding and support for researchers to follow policy directives. One positive example was offered from NextGen Cassava:

> *"The decision to make CassavaBase completely open without restriction was one of the good decisions to help encourage the culture of open data sharing. The investment by DFID and BMGF in this case has made an important contribution that I hope we can build upon and extend to other projects and other data types."*

POLICY MONITORING AND COMPLIANCE

Monitoring compliance of the implementing research partners to the policy was also expressed as a common challenge.

Programme managers from DFID and BMGF noted a lack of capacity to provide oversight and follow-up, partly due to lack of resources dedicated to monitoring policy compliance. One strategist at DFID noted (in relation to underlying dataset publication under their open access policy):

> *"One of the main lessons was that we set up no central oversight mechanism for that particular policy. There was no policing of it, there was no gathering of data about its use, no systematic approach to implementing it across all of our systems."*

The policy was never fully embraced because it was written, devolved, and disseminated but without ongoing oversight over its implementation[81].

By contrast, USAID has a strong approach to monitoring implementation, with data stewards regularly checking what data publication is in the pipeline, reminding programme managers to request data management plans, and maintaining oversight of open data implementation per ADS579 as a whole.

81 At the time of writing, it is understood that DFID is commissioning a review of its existing Open Access policy

However, once data is published, there is not always effort given to understanding who is using the data and documenting evidence of impact. This is possibly due to the early stage of implementation within several programmes, but it certainly reflects a lack of resources dedicated towards evaluating data use. This step might be more likely to happen if included as part of the data management plan from the start.

To systematise the learning process, USAID has appointed a 'Learning Team Lead' within the Monitoring, Evaluation and Learning (MEL) division of the Bureau of Food Security who is charged with overseeing overall open data policy implementation, and makes sure learning is being acted upon in terms of streamlining processes.

There was a common desire to share lessons learned about policy implementation among donor organisations. This suggests a need for ongoing learning and 'light' policy iteration including the latest thinking and good practice – ideally converged with other donors.

MAKING IT EASIER FOR RESEARCHERS TO PUBLISH AND FIND DATA

Several implementing research partners reported lack of clear directives around where and how to publish open data. For some, thinking about the accessibility of the user interface was an afterthought when designing the data publication platform. DFID's open access policy does not specify where to publish (but provides some recommendations), whereas BMGF lists PubMed Central as the nominated repository[82].

There was desire for common formats and simple software solutions to make the process of data publication easier. Tools that prompt researchers to comply with best practice standards and formats were also encouraged.

Beyond publication, there seemed to be a gap in thinking about how the data, once published, would be maintained after the programme ends ('data stewardship'). Related to this, low effort is generally put into creating awareness that data exists with external stakeholders who could be interested in using it. This finding is supported by few references to engagement within donor policies, and patchy discoverability of data.

There is a need to create more awareness of where potential data users can go to find open data, and clearly signposting where data is being made available. As one programme manager from USAID remarked:

> *"That's one of the things I think we could, in the data community, really focus on — how do we better raise the level of awareness of first of all where the data is and how do you get it and how accessible is it?"*

Ensuring datasets and research publications appear in a variety of locations, including on the funder portal/ library, in addition to a link to the programme website, can help to ensure data funded through research remains in the public domain.

The differing needs of data user groups suggests the need to maintain a degree of flexibility when determining the strategy for publication, while factoring in capacity support and user-friendly tools to make high-quality data publication easier for researchers.

82 See https://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy/Page-2

SUMMARY OF OVERCOMING IMPLEMENTATION CHALLENGES

From the interviews we identified the following challenges:

- Data quality, data management, strategy (e.g. creating data management plans), putting data to use, and understanding responsible data use, especially among local research partners within low-income countries;
- The pressure to publish, lack of proper budgeting, and institutional policies are discouraging implementing research partners from releasing underlying datasets in a timely way;
- Attempts to foster a culture of openness are impeded by fears among researchers around loss of benefit and uncertainty around use[83];
- Among some donors, there is a lack of oversight to routinely monitor how data is being used and to evaluate impact;
- Unclear directives on where to publish open data and low awareness around how data can be accessed are creating barriers for data producers and limiting potential use.

These challenges confirm some of the gaps in policy identified in Section 3 (e.g. relating to signposting data), but also go deeper on issues silent in the policy such as culture change, addressing fears around data control and use, and active engagement strategies to promote uptake of data.

To help overcome these challenges, **donors should**:

- **Resource capacity-building activities to support implementing researchers and programme managers**, including opportunities for peer learning exchange among implementing research partners and between donor organisations;
- **Incentivise researchers to publish** by rewarding good-quality data production, promoting the value of research data, embedding data publication as a deliverable of the programme – including within long-term budgets;
- **Foster culture change through insisting on openness and collaboration**, rewarding open behaviour, and delivering messages from the highest level of leadership;
- **Designate responsibility for monitoring policy implementation**, the quality of data production, and process improvement within the donor organisation;
- **Explore shared tools for making data publication by implementing research partners easier**, and factor in considerations around platform selection and awareness raising within data management plans.

83 Refer also to http://www.godan.info/documents/ownership-open-data-governance-options-agriculture-and-nutrition-0

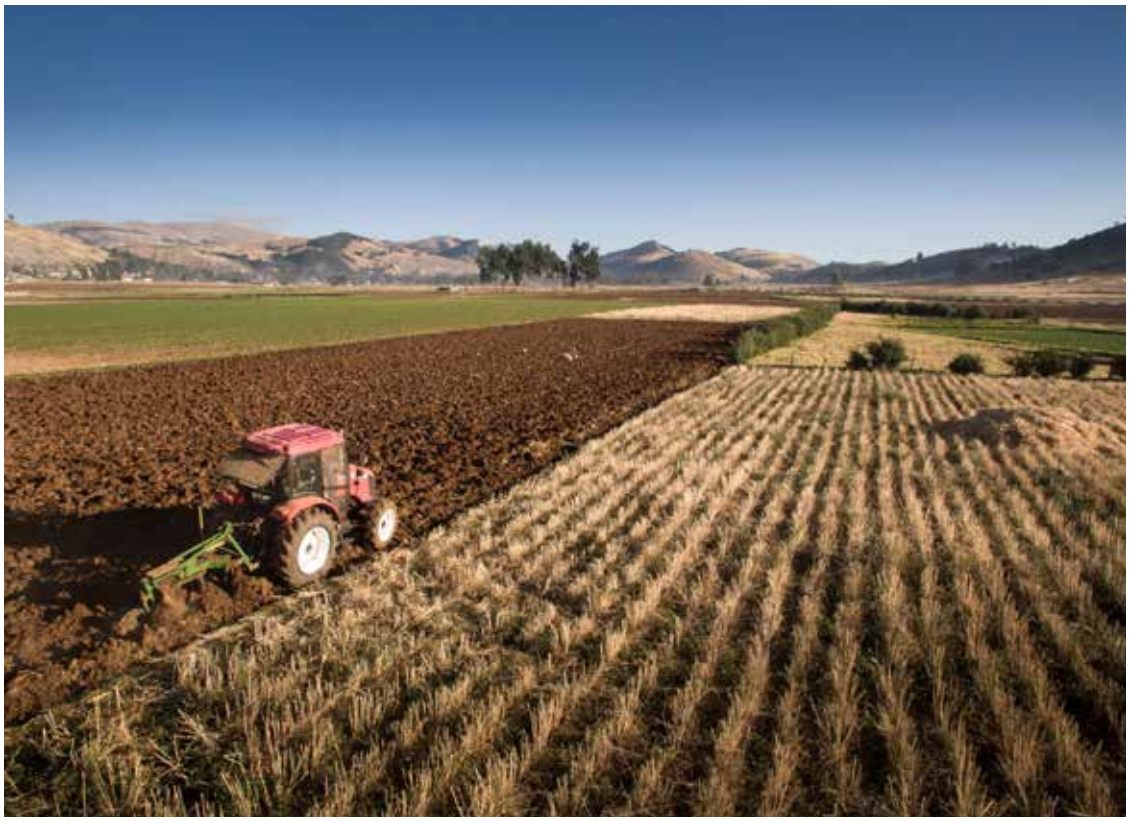
CONCLUSION: TOWARDS SHARED DONOR PRINCIPLES

Reaching agreement on common approaches and principles towards open data within a sector is no easy task. We recognise and respect each organisation's unique culture, legal arrangements, and structure that inform the precise terms of their internal policy.

Yet within these different arrangements there is still a high degree of flexibility. As far as possible, donor policies can employ similar underlying principles, commitments, and terminology. At minimum all organisations should regularly review their policy with implementing research partners, and monitor the quality of data publication and research outputs in accordance with the policy.

A variety of common tools and actions are also required to balance the rights of researchers who have contributed towards data production with the obligations of donors and interests of the wider community that could productively use the data for social, environmental and economic benefit. Donors should review implementation guidance and requirements relating to data management plans, while providing ongoing opportunities for capacity building to support compliance.

In the following section we consider the opportunities for greater alignment of donor approaches towards implementation, and propose steps for strengthening practice based on a shared set of principles.



OPPORTUNITIES TO HARMONISE TOOLS AND APPROACHES

Guidance materials and tools are one area where donors could achieve greater alignment and share best practices. There is a strong desire for more standardisation on process internally and between donors, as expressed by USAID:

> *"... how can we find the commonalities and align our policies so that we can better facilitate data sharing more quickly, more widely and then identify what our common data standards are and practices are … so that we can more quickly facilitate the release of data and share it openly."*

As a first step, donors should share existing tools and data standards that support good open data practice, such as data management plans. This could take the form of flexible templates or checklists, which give the donor the ability to include what is important while covering the minimum core attributes. The data management plan should be accompanied by guidance or examples showcasing how to draft one. As one programme manager from BMGF observed, this would make it easier for donors to enforce, while also making it much easier for implementing research partners to avoid having to comply with numerous different policies from each donor they are working with.

Using existing common tools for validating data publication, such as ODI's Open Data Certificates[84], would help to improve accessibility, discoverability and quality of data. From our data publication quality assessment, clearly there were gaps around choosing which platform to publish on, applying a license, and maximising the discoverability of data. Common tips, advice, and examples of good practice might further strengthen practice in this area.

During the course of interviews, several external resources were referred to as helpful guides for thinking through ethical issues and planning how to manage data responsibly. These included Oxfams's 'Responsible Data Management' kit[85], and CGIAR's 'Open access and open data support pack[86]'. CGIAR's support pack includes information about licensing, staffing information, creating a 'Data Management Plan', budgeting for open data, and creating an implementation plan. The implementation plan[87] and data management plan[88] consider ethics, privacy and data ownership issues.

A SHARED SECTOR-WIDE APPROACH

The second key area where donors can make a difference is around shared commitments towards shifting the culture of data sharing, publication and use across the agriculture sector.

We have seen the power of funders to drive changes in behaviour through clear communication backed by incentives in the area of genomics research. In that sector, breakthrough research on genome sequencing occurred only due to large, collaborative research communities. The Global Alliance for Genomics and Health[89] have developed a 'Framework for Responsible Sharing of Genomic and Health-Related Data[90]', which provides a principled and practical framework for the responsible sharing of genomic and health-related data. It contains foundational principles backed by practical commitments to capacity building, sustainability, increasing accessibility of data, and sharing policy templates.

84 https://certificates.theodi.org/en/
85 http://policy-practice.oxfam.org.uk/our-approach/toolkits-and-guidelines/responsible-data-management
86 See https://sites.google.com/a/cgxchange.org/oad-support-pack/
87 See https://docs.google.com/viewer?a=v&pid=sites&srcid=Y2d4Y2hhbmdlLm9yZ3xvYWQtc3VwcG9ydC1wYWNrfGd4OjczOWRiNDcwYjliMDFhYzM
88 See for example https://docs.google.com/viewerng/viewer?url=http://www.reading.ac.uk/ssc/resource-packs/dms/02-data_management_policies_and_plans/Creating+a+Data+Management+Plan.pdf
89 https://genomicsandhealth.org/
90 https://genomicsandhealth.org/files/public/Framework%20for%20Responsible%20Sharing%20of%20Genomic%20and%20Health-Related%20Data%20-%20Version%2010%20September%202014.pdf

The Wellcome Trust's new data sharing policy[91] also provides a potential model to be emulated, with a holistic view of how all research outputs, including data, can be made open and discoverable at the earliest opportunity. The Wellcome Trust acknowledges the value in *"developing policies and practices that incentivise and reward implementing researchers who support open research"*, and in actively policing programmes with new checks against compliance[92]. They do this through providing funding to cover open access publishing costs, working in partnership with other funders to support Europe PMC (PubMed Central) – an online database offering free access to published biomedical research, making the process of publication easy[93], allowing researchers to cite preprints in grant applications[94] to speed up the publication of datasets, and launching a prize to promote and celebrate new approaches and technologies that facilitate the sharing and re-use of research outputs.

It is time for donors, philanthropic funders, and public international organisations investing in agriculture research for sustainable development to convene around a similar vision for harmonising their approach towards open data. This can send a powerful message to researchers, while generating more awareness of the value of open data within the agriculture sector to inspire other funders.

From our combined policy and implementation review, we found several **opportunities where donors of agriculture programmes can align** around:

1. **Join a global funder dialogue with other donors, researchers, and research institutions** aimed at advancing a shared set of principles and sharing good organisational practices, to underpin more harmonized open data implementation. As noted by the Wellcome Trust in their research on embedding cultures and incentives to support open research[95], discussion and exchange of ideas between different players is essential for effective progress to take place.
2. **Support and adopt common policy principles** such as definitions of key terms, and open standards to achieve greater uniformity in the standards and formats in which implementation partners generate, curate and disseminate open data[96]. Adopting a common lexicon of key terms could also help to promote consistency of approach.
3. **Shared approaches towards dealing with ethical considerations** such as justifications for non-publication and responsible use of data. Draw on best practice and lessons learned in complementary sectors[97].
4. **Promote good open data practice among those receiving funding by regularly monitoring compliance, and articulating clear expectations regarding budget allocations to ensure open data**, e.g. with open access requirements. Ensure measures are in place that support compliance such as data publication quality checks post research grant award, budget review templates, tracking outputs, and ongoing channels for feedback/questions.
5. **Increase engagement and introduce practical projects to promote data reuse and innovation** to create more value from data including between research partners, through additional awards or prizes for data, use/innovation, and public awareness raising about data.
6. **Collect data use stories to demonstrate value and impacts of research data** in the agriculture domain and disseminate results widely, thereby motivating implementing researcher partners to adopt good open data practice.

91 https://wellcome.ac.uk/funding/managing-grant/policy-data-software-materials-management-and-sharing
92 https://wellcome.ac.uk/news/our-new-policy-sharing-research-data-what-it-means-you
93 https://wellcomeopenresearch.org/
94 https://wellcome.ac.uk/news/we-now-accept-preprints-grant-applications
95 https://wellcome.ac.uk/sites/default/files/embedding-cultures-and-incentives-to-support-open-research-oct16.pdf
96 e.g. IATI and OpenAg Funding, DCAT, shared vocabularies such as the Global Agricultural Concept Scheme (GACS)
97 e.g. Responsible Data Use forum, Oxfam Responsible Data Management kit, ODI data ethics canvas

7. **Support the capacity of implementing research partners to improve data availability, accessibility, discoverability and quality** via peer learning, training programmes, and automated tools like Open Data Certificates. In particular, invest in capacities of data collectors and researchers in developing countries to manage data locally and contribute towards global efforts to meet SDGs.
8. **Adopt shared guidelines, tools and templates aimed at reducing the time and cost of policy compliance**, making it easier to publish data and get it to those who need it, and enhance the utility of data held in open repositories, e.g. common data management plan templates.
9. **Incentivise researchers to publish by rewarding good-quality data production** via writing into research team KPIs, embedding data publication as a deliverable, recognising previous efforts in data publication, offsetting publication costs and delivering strong messages from senior leadership about requirements related to data sharing, budgeting and publication. Work with publishers to encourage publication of underlying datasets, and promote measures that extend the reach of data citation and the use of DOIs.
10. **Sustainably resource data publication and management**, and require a strategy for data dissemination and stewardship in place from the beginning of all new projects, allowing for data to be published in any open, standards-compliant repository. Ensure that the full cost of data collection, quality control, management and publication is accurately assessed during research design and budgeting phases. Adequately invest in the 'data infrastructure' over the long term (people, technology and systems, e.g. data repositories) to ensure data remains accessible and discoverable including after the life of the project.

ANNEX I: RESEARCH METHODOLOGY

RESEARCH METHODOLOGY

In this research, we are particularly interested in funder approaches towards supporting implementing researcher partners to publish open data. The research questions we sought to answer through this research report were:

1. What are **common principles and examples of best practice** across the range of represented donor open data policies. What are the **main areas in which the donor policies differ**, and where does this become a problem for compliance?
2. What are the main **lessons the programmes have learnt** through implementing those policies, including main challenges and opportunities for adoption?
3. What **benefits do implementing partners perceive** getting for their programmes or organisations from compliance with open data policies?
4. How well do the selected programmes implement existing donor open data policies? In other words, what is the **quality of open data publication**?
5. How can donor open data policies be **strengthened**, and under what **shared principles**?

Our research was primarily focused on data publishing by the implementing research partner, as opposed to data produced by the donor concerning the nature of its footprint and spending (i.e. International Aid Transparency Initiative (IATI)[98] data). We note that other research is ongoing into the area of how to improve the quality and usability of this type of data by the Open Ag Funding[99] initiative.

To answer these questions, we carried out research in three phases:

First, we **reviewed the different donor policies** against criteria in the ODI's 'guide to writing a good open data policy[100]'. The guide provides a checklist of policy elements covering data licensing and reuse rights, prioritizing data for release, privacy considerations, data publishing standards, engaging with data users, approach to consuming data, monitoring commitments, and overall policy transparency.

The aim of this phase was to understand the individual organisations' various approaches towards implementing open data. Following the desk-based assessment, we compared the three policies to identify patterns of best practice, inconsistency, and cross-cutting gaps. This analysis provided insight into which areas of open data policy might be strengthened, and highlighted commonalities in approach which could be endorsed.

Second, we conducted twelve **key informant interviews** with a range of stakeholders from the three funders representing five selected jointly funded agriculture programmes to test our findings from the policy review, and to explore what implementation looks like in practice. The five programmes that were selected by the three donors were:

98 http://www.aidtransparency.net/
99 https://www.interaction.org/project/open-ag-funding/overview
100 https://theodi.org/guides/writing-a-good-open-data-policy
101 http://csisa.org/
102 http://www.new-alliance.org/
103 http://www.nextgencassava.org/
104 https://www.atai-research.org/
105 http://www.driversoffoodchoice.org/

1. Cereal Systems Initiative for South Asia (CSISA)[101]
2. New Alliance ICT Agriculture Extension Challenge Fund [102]
3. Cassava Diagnostics Programme (NextGen Cassava) [103]
4. JPAL Agriculture Technology Adoption Initiative (JPAL ATAI) [104]
5. Agriculture Nutrition impact studies [105]

We also interviewed two external experts in open access policy from CGIAR and Wellcome Trust to understand good practice within other organisations. Lastly, we surveyed twelve representatives of grantee organisations to invite insights from researchers which are responsible for implementing the grant ('implementing research partners'). The purpose of these interviews was to probe lessons learned, challenges and benefits experienced during implementation.

Third, we completed an **assessment of data quality** from five jointly funded agriculture programmes, using the Open Data Certificates[106] as a framework to look at:

- Evidence of data being created and used by the programme
- Availability, accessibility and discoverability of open data
- Quality of data publishing, including the licence, access process (e.g. logins), format, metadata and documentation

By examining the programme websites, research publications, and programme documents we aimed to understand what data is currently being produced by the programme and how effectively this is being communicated from a data user's perspective.

Finally, we conducted a workshop with representatives of the three donors to feedback initial research findings and to invite discussion around what might be adopted as common donor principles. Feedback from that discussion is incorporated into this report.

LIMITATIONS OF THE RESEARCH

Several factors limit the potential universality of findings:

- Different levels of maturity of the various agriculture programmes, which made it difficult to compare outputs against the same metrics.
- The relatively small sample size of interviewees means it is difficult to determine how widespread and emblematic the challenges we identified are across the organisation.
- The number of programmes nominated for review by donors was also relatively small (five), and perhaps not representative of the entire field.
- It was challenging for the ODI to access representatives of implementing research partners during the research period due to their location in the field and travel demands. As such we had to adapt our original research methodology to use online surveys as well as interviews.

106 https://certificates.theodi.org/en/

Global Open Data for Agriculture & Nutrition (GODAN)

c/o CAB International
Nosworthy Way
Wallingford
Oxfordshire
OX10 8DE
United Kingdom

Tel: +44 (0) 1491 832111
Fax: +44 (0) 1491 833508

Follow us on Twitter:
@godanSec

Thinking of becoming a GODAN partner?
Visit http://www.godan.info/become-a-godan-partner

DOI: 10.1079/CABICOMMS-80-70