A grayscale photograph of an elderly man walking towards the camera in a dry, open landscape. He is wearing a white turban and a light-colored shawl draped over his shoulders. He has a walking stick in his right hand and a bundle of sticks or tools slung over his left shoulder. The background shows sparse trees and a clear sky.

Measuring the value of improving data governance and access in gates foundation programmes

**A case study of the Supporting Soil
Health Interventions in Ethiopia project**

Measuring the value of improving data governance and access in gates foundation programmes

A case study of the Supporting Soil Health Interventions in Ethiopia project

Damian Whittard, Felix Ritchie and Jonathan Nyengani
Bristol Business School University of the West of England

Corresponding author: damian2.whittard@uwe.ac.uk

Contents

ACKNOWLEDGEMENTS

UWE, Bristol would like to thank all the stakeholders who participated in the development of the evaluation framework and assisted in its application. In particular, we would like to thank the many experts who invested their valuable time to take part in the interviews and surveys. Our special thanks, however, go to our funders the Centre for Agriculture and Bioscience International, the Open Data Institute and the Gates Foundation. In particular we would like to thank Ruthie Musker, Deborah Yates and Michael Rose, all of who gave their time generously to support the study, much of which is informed by their insightful contributions.

Executive Summary

Background and Context

The Bill and Melinda Gates Foundation (BMGF) investment, 'Enabling data access to support innovation in decision agriculture related to soil health, agronomy and fertilizer' aims to promote better data governance in decision agriculture investments, and the national systems in which the investments operate. The goal of this investment is to co-create tools, guidance and recommendations for multiple stakeholders within target regions in India and Ethiopia to help them apply good data practices.

In support of this project, in July 2019 the Centre for Agriculture and Bioscience International (CABI) and the Open Data Institute (ODI) commissioned a team of economists from the University of the West of England, Bristol (UWE) to measure the value of improving data governance and access in projects supported by BMGF programmes.

Following an iterative approach when a number of projects across India and Ethiopia were considered, it was decided that the Supporting Soil Health Interventions in Ethiopia (SSHIE) project was the most appropriate project on which to focus. The SSHIE project was a \$1.5 million project led by the Gesellschaft für Internationale Zusammenarbeit (GIZ). The project ran from November 2017 to August 2020.

Objective of the Study

Given the challenges with valuing something so nebulous as data governance and access, and in the absence of any formal frameworks previously applied to measure such concepts, two key objectives were identified.

- To develop a formal framework that could be used in the planning, measurement and evaluation of the value of data governance for future donor funded projects.
- To test the framework by evaluating the value of data governance of a CABI funded project and identifying the mechanisms through which value is created.

In order to identify the best approach, multiple frameworks and evaluation structures were tried and tested. Two separate but interlinked models were selected as potentially helpful in the evaluation of data governance and access. The qualitative model was based on the Five Safes framework which previously had only been used in the design of data strategies. The quantitative framework combined activity-specific logic models (theory of change models) with traditional cost-benefit analysis (CBA) techniques.

Developing the Framework

Five Safes Framework

The Five Safes is a recognised framework used in the planning and design of data governance and access, but previously had not been used in evaluation. This study developed and adapted the framework so that it could be applied to evaluate the value of data governance across a range of CABI projects and programmes. In order to do so, a generic framework with example questions was created. The initial framework was sent to stakeholders for testing. It was then applied to guide the initial round of interviews set up to explore issues in relation to the value of data governance and access in the SSHiE project.

The Logic Model

This framework was specifically developed to guide the approach used to quantitatively estimate the value of data governance in the SSHiE project. The UWE and CABI team initially co-developed a logic model from which the CBA was derived. The logic model was created to map how improvements in data governance and access could lead to improved output, outcomes and wider economic and social impacts. The CBA was subsequently developed so that it could collect data at different points along the logic model.

Combining the Models

Following an initial set of interviews with CABI and GIZ, the Five Safes framework was mapped against the logic model to enable any gaps in information to be identified; those gaps were then addressed in a second round of interviews. This was followed by the creation of a survey administered to twelve project experts.

The two models were then used as the frameworks to analyse the results generated from the interviews and surveys. By applying these frameworks to the SSHiE case study, it was possible to test the effectiveness of the frameworks for valuing data governance and access. This approach also met the second objective by generating qualitative and quantitative insights into the value of data governance in the SSHiE project.

The Qualitative Approach (Based on the Five Safes Framework)

The qualitative approach enabled the research team to:

- Identify the specific types of value to different stakeholder groups
- Describe the mechanisms through which this value is perceived to have been created
- Develop a framework to quantitatively measure this perceived value

Stakeholders' perceptions of value and the mechanisms through which it is generated were multifaceted. The interviewees identified tangible, and easily measurable, outputs such as research papers and changes in policy. They also elaborated on much softer, intangible outcomes such as change in culture and increases in trust, while also discussing potential wider economic and social benefits.

One of the key themes across the qualitative study was the importance of trust (e.g. trust between stakeholders, in the quality of data, or in the institutional arrangements for governance). There was also an appreciation of the need to invest significant amounts of time to build trust before stakeholders could work together in a meaningful way to co-create and co-deliver projects; but also awareness that this was an investment, not expenditure.

Other elements identified as important in deriving value were: engagement of stakeholders, with particular emphasis on local empowerment supported by international expertise; and the identification of high level project 'champions' and willingness to work with potential 'blockers' to the project. Of particular value to the success of the project was the creation of a Coalition of the Willing (CoW), a group of scientists/researchers who came together to help drive the project locally.

Improving data governance and access is sometimes seen primarily as a policy and infrastructure challenge. It was clear, however, that the interviewees felt that full value could only be achieved if supported with investment in building human capital right across the data value chain.

The Quantitative Approach (based on the Logic Model)

Surveys were used to generate exploratory estimates of value of data governance and access. The research team collected and analysed the available data in relation to inputs (resources used to deliver the project), activities (what is delivered), outputs/outcomes (what is achieved) and wider economic and social impacts. As a result of the quantitative research stream the research team were able to:

- develop a methodology to measure the perceived internal value of improved data governance and access within the project itself;
- attribute this value to the mechanisms through which it was created;
- provide indicative estimates of the external value and impact of improved data governance and access; and
- assess where this impact would most likely be felt in the economy.

Two fundamental questions were raised by this work stream:

- How do you quantify, value and monetise what potentially is an unquantifiable concept?
- How do you bracket and measure just the data governance and access elements within the wider bounds of a project?

The research team did develop a (partial) framework and methodology which allows some form of quantitative measurement of data governance and access. This framework, however, was also subject to other practical concerns impacting on the quality of the estimates produced. These include the following:

- estimates are based on a small number of data points;
- answers provided are highly subjective and potentially subject to optimism bias, given that the interviewees had an interest in positive perceptions of the project; and
- there are a number of very broad assumptions about costs and benefits.

While acknowledging these significant and limiting caveats, the process showed that it is possible to generate plausible and credible quantitative estimates of both costs and benefits of data governance and access.

Given these caveats, and acknowledging that the estimates are only illustrative, the case study results suggested that to date, and on a direct cost measure, the SSHIE data governance activities yielded a negative return; but they also show that relatively few 'indirect' benefits (current but unmeasured, or measurable but in the future) are necessary to reverse that view, at least from the point of the economy more generally.

List of Acronyms

ATA	Agricultural Transformation Agency
BCR	Benefits Cost Ratio
BMGF	Bill and Melinda Gates Foundation
BSRC	Bangladesh's Soil Resource Centre
CABI	Centre for Agriculture and Bioscience International
CBA	Cost-Benefit Analysis
CEA	Cost Effectiveness Analysis
CGIAR	Consultative Group on International Agricultural Research
CIAT	International Centre for Tropical Agriculture
CoW	Coalition of the Willing
CSA	Climate-Smart Agriculture
CSISA	Cereal Systems Initiative for South Asia
CWoDAaS	Consultation Workshop on Data Access and Sharing
DIV	Direct Internal Value
DSTF	Data Sharing Task Force
EIA	Economic Impact Assessment
EIAR	Ethiopian Institute of Agricultural Research
ESRI	Ethiopian Soil Resource Institute
EthioSIS	Ethiopian Soil Information System
FAIR	Findable, Accessible, Interoperable and Re-usable
FITS	Farmer's Information Technology Services
GIZ	Gesellschaft für Internationale Zusammenarbeit
HICs	high-Income Countries
HCD	Human-Centred Design
ICfTA	International Centre for Tropical Agriculture
ICT	Information, Communication and Technology
IIV	Indirect Internal Value
LMICs	Low- and Middle-Income Countries
MCA	Multi-Criteria Analysis
MoA	Ministry of Agriculture
NPV	Net Present Value

NRM&FSS	Natural Resources Management and Food Security Sector
ODI	Open Data Institute
OECD	Organisation of Economic Co-operation and Development
PIR	Policy Implementation Report
ROAMEF	Rationale, Objectives, Appraisal, Monitoring, Evaluation and Feedback
SADS	Soil and Agronomy Data Sharing
SCW	Stakeholder Consultation Workshop
SDR	Social Discount Rate
SRDI	Soil Resource Development Institute
SRI&MD	Soil Resource Information and Mapping Directorate
SSHiiE	Supporting Soil Health Interventions in Ethiopia
UWE	University of the West of England

Introduction

In July 2019, a University of the West of England (UWE) research team began working with the Centre for Agriculture and Bioscience International (CABI) and the Open Data Institute (ODI) to agree an approach to measure the value of improving data governance and access in the Bill and Melinda Gates Foundation (BMGF) programmes.

BMGF works worldwide in five programme areas. BMGF's Global Growth & Opportunity programme focuses on creating market-based innovations to stimulate inclusive and sustainable economic growth. All their programmes emphasise collaboration, innovation, and risk-taking, but, most importantly, results. This study attempts to provide some exploratory estimates of the value of data governance in the Gates Foundation funded Supporting Soil Health Interventions in Ethiopia (SSHIE) project, using economic appraisal techniques based on, but not limited to, traditional cost-benefit analysis (CBA) techniques.

Given that relatively little research had been completed in this field, the first challenge was to agree working definitions of key terms. For the scope of this study, data governance is defined as follows:

'A framework for assigning decision-related rights and duties in order to be able to adequately handle data as an asset' (Otto, 2011, p.47)

As value is a nebulous concept, another challenge was to agree how 'value' would be measured in this study. Following a review of the literature and after discussions with key stakeholders, the following criteria to assess value were agreed and applied throughout the study:

- To what extent has the project delivered value by meeting its commitments (e.g. project outcomes in relation to cost savings/investments)
- What is the additional value to users (i.e. value to be specified by users)
- What is the additional value to other stakeholders (i.e. value to be specified by other stakeholders)

From the beginning, the research team have followed an action research methodology¹ which focusses on bringing individuals together in open dialogue, to engage in cycles of action and reflection with subsequent ideas examined in phases of active experimentation. This approach also aligned with CABI's preferred Human-Centred Design (HCD) methodology which it applies across its development projects. HCD is a creative approach that places the human perspective at the centre of the design process for new solutions to meet the users' specific needs.

¹ Action-research practices aim to open communicative spaces where people can come together in open dialogue to address issues of concern and to engage in cycles of action and reflection, so that ideas that are tentatively articulated in reflection can be examined systematically in phases of active experimentation (Encyclopaedia Britannica 2020, online)

In line with this methodological approach, an iterative approach to specifying the project aims was followed. The final aims of the project were agreed as follows:

- Develop a formal framework that could be used in the planning, measurement and evaluation of the value of data governance for future donor funded projects.
- Test the framework by evaluating the value of data governance of a CABI funded project and identifying the mechanisms through which value is created.

Initially this study focussed on developing a formal framework to evaluate the value of data governance, with the aim of providing CABI and its partners with a standardised approach for developing and evaluating any future investment. This approach also offers transferable insights that go wider than just the case study under investigation. Such findings are particularly helpful to the wider donor community and therefore the study has paid close attention to identifying benefits and the mechanisms through which they work.

To test its usefulness in evaluating the value of data governance, this framework was used to estimate the value of data governance in CABI's SSHiE project. This provided an additional benefit to the SSHiE project of a greater understanding of the value of data governance within their project.

This study has made a number of important contributions to the under-researched area of the value of data governance and access; all of which should be of value to CABI, its partners and the wider donor community. In particular, it has successfully:

- developed and applied the popular 'Five Safes' framework to evaluate the value of data governance.
- applied the framework using a mixed method approach by combining both qualitative and quantitative methodologies and data sources.
- provided insights in to how to evaluate the value of data governance and access in future studies.

The qualitative approach enabled the research team to:

- Identify the specific types of value to different stakeholder groups
- Describe the mechanisms through which this value is perceived to have been created
- Develop a framework to quantitatively measure this perceived value

Quantitative methods were used to generate formal, if exploratory, estimates of value of data governance and access. The research team collected and analysed the available data in relation to inputs (resources used to deliver the project), activities (what is delivered), outputs/outcomes (what is achieved) and wider economic and social impacts. The team:

- developed a methodology to measure the perceived internal value of improved data governance and access within the project itself
- attributed this value to the mechanisms through which it was created
- provided indicative estimates of the external value and impact of improved data governance and access
- assessed where this impact would most likely be felt in the economy

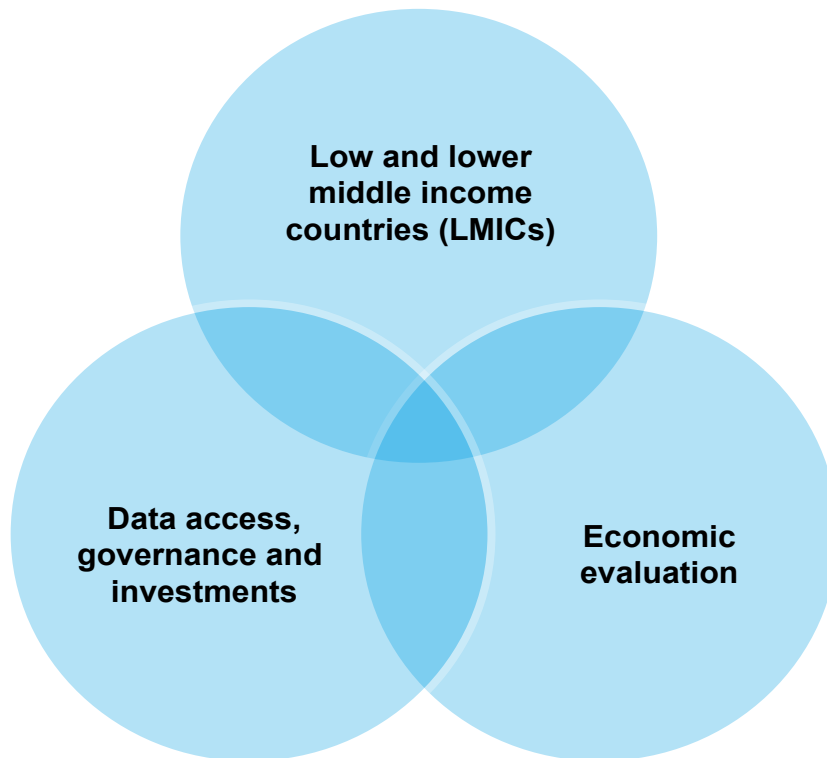
This report is structured as follows:

- Chapter 2: review of the existing relevant academic and practitioner literature
- Chapter 3: frameworks to be used
- Chapter 4: methodological approach
- Chapter 5: qualitative findings from stakeholder interviews
- Chapter 6: quantitative findings from survey data
- Chapter 7: reflections on the research process
- Chapter 8: Conclusions

Literature Review

A structured literature review was undertaken to identify papers using various interactions of the three key themes below:

Figure 1: Literature Review Structure



Part 1: Economic Evaluation

When undertaking an economic evaluation using a CBA approach there is a well-defined structure to follow (further information on the specifics of this approach are included in Appendix 1). However, when completing a CBA for a specific project, the underlying assumptions depend on the specifics of the intervention such as geography and sector of investment. Therefore, in order to undertake a detailed CBA, the evaluation will need to access considerable amounts of well-defined, measurable, high quality data; all of which can be converted to a market price.

In this study this is potentially problematic for a number of reasons. First, data, governance and access are all intangible in nature and therefore difficult to define and measure (Corrado et al., 2009). For example, Wdowin and Diepeveen (2020) report that data is neither a good nor a service and it is not homogenous; nevertheless, data can be viewed as a 'public good' where it is freely accessible (non-excludable) and one person's use does not prevent others from doing so (non-rivalrous). In contrast, Savona (2019) suggest that data generally displays the characteristics of a club good (non-rivalrous but excludable); this may lead to data hoarding, and as such be a barrier to sharing data. 'Data governance' is a further level of abstraction from data as a public or club good, making it even harder to evaluate.

Second, even if the concepts were easy to define and measure, this would need to be supported by a robust system of data collection, both within and external to the project. The experience of most economic evaluation experts is that the planning for such evaluation starts too late in the project lifecycle, therefore missing the chance to collect the necessary data. At the same time, end-of-project evaluations can be carried out too soon after project completion for outcomes and impacts to manifest. This issue was elaborated on in Belmana's (2018) interim evaluation of the UK Research partnership Investment Fund programme. Helpfully they offer suggestions to improve the end-of-project economic evaluation, although the usefulness of these suggestions is limited by the fact that they lack any clear practical guidance on how this can be achieved.

Both the issues of starting the planning too late and evaluating too early were a challenge for the case study (the project was still ongoing at the time of the evaluation). These challenges are further compounded when undertaking evaluations in low- and middle-income countries (LMICs) where the supporting statistical infrastructure may not be sufficiently developed to support the detail required to complete such an evaluation. The following sections further explore the challenges and issues of evaluating investments in LMICs in more detail.

Part 2 Evaluating Agriculture Projects in LMICs

Gross National Income (GNI) is used by the World Bank to classify economies into four groups: low, lower middle, upper middle, and high income countries. Ethiopia has one of the fastest growing economies in Africa, but is classified as a low income country with per capita income of \$790 in 2018. The World Bank report, however, that it aims to reach lower-middle-income status by 2025 (World Bank, 2020).

There is a raft of research that has applied CBA techniques to estimate the economic value of a wide range of agriculture sector projects in LMICs. The following section provides an analysis of a range of papers, highlighting a number of the practical challenges they faced.

For example, Sain et al. (2017) provide a comparative analysis of eight climate-smart agriculture (CSA) practices and technologies associated with the smallholder maize-beans production system. The paper outlines a well-defined methodology in order to assess the cost-benefit profile of the introduction of CSA options into farm production systems. The particular strength of the paper is the use of Monte Carlo simulations to present a probabilistic CBA considering a range of values and the likelihood of these values. They also used innovative approaches to incorporate estimates of environmental and social externalities into estimates of economic profitability of CSA practices – these include biodiversity, carbon sequestration and soil and water contamination.

The paper argues that

“a thorough understanding of the costs and benefits of potential CSA options is needed to channel investments effectively and efficiently towards both short- and long-term interventions and should be coupled with broader assessment of trade-offs between CSA outcomes.”

Lemaga et al. (2005) conducted a CBA assessing the viability of sweet potato post-harvest technologies. The study collected both quantitative and qualitative data. Data collection methods include 23 individual interviews with farmers, observing farmer records and accounts and participatory techniques. The questionnaire was pre-tested before the wider rollout to farmers. The data they collected directly from the farmers included direct and indirect cost and benefits, input requirements, prices of inputs and outputs and availability of inputs, as well as qualitative data on the acceptability of the technologies by the farmers. Focus group discussions were also carried out with farmers to capture qualitative information regarding the sweet potato technologies.

The authors concluded that sweet potatoes have the potential to improve household income and be instrumental in fighting poverty. A strength of this project is its simplicity of approach and transparency. There is a detailed discussion on methodology and the authors have included a copy of the questionnaire. Overall, however, the results are based on a limited sample and the approach taken is more akin to a financial rather than an economic evaluation.

NKang et al. (2009) undertook a statistical analysis of cocoa production in Nigeria, looking at the costs and benefits of three different cocoa management production systems – owner-managed farms, lease-managed farms and sharecrop-managed farms. They concluded that irrespective of management system, cocoa production is profitable. The Net Present Value (NPV) was highest for lease-managed farms; however, owner-managed farms had the highest Benefits Cost Ratio (BCR).

A strength of this paper is that the authors provide a detailed description of methods and data sources. They describe a two-staged sampling procedure; the first stage involved the purposive selection of the two largest cocoa producing areas, while the second stage involved the random selection of 150 farmers - 50 from each of the three management systems. The research was conducted using a structured survey sent to each of the farmers; this was augmented with secondary data from farmer records as well as information from local and national government institutions and international programmes. The analysis also attempts to estimate wider environmental costs and benefits from the project by applying the Gotsch and Berger (2001) model, which takes into account the age of trees and expected revenue per hectare. The authors explain the importance of this approach

“Benefits and costs are linked to the age of the trees. At the early stages, there are heavy costs which are then followed by annual benefits that continue over the full life of the trees once they have reached maturity.” (p.716)

The results of CBAs are sensitive to the discount rate applied. However, there was a lack of clarity as to the reason behind their decision to adopt a discount rate of 10%.

Baranchuluun et al. (2014) examined the costs and benefits to farmers in Western Mongolia of using drip and sprinkler irrigation systems compared with furrow irrigation in potato, radish, headed cabbage and tomato. The authors concluded that the benefits of drip irrigation systems are higher and enabled the farmer to save water and labour compared with furrow and sprinkler irrigation. Drip irrigation also reported the highest NPV and BCR.

The CBA was well developed in as much as it included financial (economic), environmental (e.g. water loss and water saving) and social (labour saving and social insurance) estimates in its CBA. The data was also collected directly from farmers and augmented by project sources and other variables (e.g. prices, growth rate) from the National Statistical Office. However, the paper lacked transparency in relation to precise methodology and sample sizes.

Mittal (2018) presents a CBA of agricultural interventions in Rajasthan. In particular, the author focussed on four areas: seed production and promotion; crop diversification; soil health; and improving and extending services via ICT (information, communication and technology). The author notes that

“Research, extension services, literacy and infrastructure have been identified as the most important sources of growth in productivity” (p.31)

While the paper presents a clear and simple methodological approach, much of the data is reliant on assumptions from previous studies. For example, in seed production they compute the benefits cost ratio of achieving a desirable seed replacement rate. The main costs are those of producing and marketing more seed. The major benefits are improved crop productivity and thus improved farmer incomes. For marketing costs, Mittal uses estimates from Birthal et.al (2015) who provide cost per hectare for knowledge transfer resulting in enhanced adoption. The main benefit is yield gains, which have been assumed to be 10%. This assessment was derived from reviewing several other studies (GoAP, 2015; Singh, Singh and Singh, 2016; Abebe et al., 2017; Clayton et al., 2009) who estimate yield gains from similar interventions to be between 15-20%.

Of particular interest is the author's approach to evaluating the effect of improving and expanding extension services via ICT. In order to do this they estimated the cost and benefits of reaching all farmers who have access to mobile phones (see Yanno et al., 2017 for detailed methodology) over a period of 5 years with advisory services. Costs included the cost of delivering agricultural advice through SMS, cost of integrated voice recording service and other costs of operation. The main benefit was increased farmer income. Estimates of increases in income were informed from findings by Rathore (2011) and Cole and Fernando (2014).

Much of the work on economic evaluation is undertaken at the individual project level. However, in September 2003, David Raitzer undertook a programmatic benefit-cost meta-analysis of 56 investments funded by the Consultative Group on International Agricultural Research (CGIAR). The methodological approach was repeated and applied to 23 CGIAR investments in Sub-Saharan Africa (Maredia and Raitzer, 2010).

Both of the studies offer a 'best evidence' approach to summarising the cost and benefits of agricultural research. To do this the studies scaled-up quantified economic impacts of a number of different projects, grouping them in terms of their credibility.

In order to make this assessment, the authors developed a framework for assessing the credibility of impact assessment related to agricultural research. The methodology for both studies was similar and based around the following criteria:

- Transparency
 - Clearly derived key assumptions
 - Comprehensive description of data sources
 - Full explanation of data treatment
- Demonstration of causality
 - Representative data set utilized
 - Appropriate data treatment
 - Plausible counterfactual developed

From these criteria, each individual study was then assessed and attributed to a group. In the initial study, there were five possible groups, but this was condensed down to three in the later study:

- Potential
- Plausible
- Substantially demonstrated

To undertake a group analysis of independent projects, a number of adjustments were made including converting everything into U.S. Dollars in real terms and discounted using a 2% real social discount rate, with sensitivity analyses lowering the rate to 0% and raising it to 10%.

The rationale for assessing the aggregated value from a group of interventions was that if the returns of the most credible CBAs were greater than the overall investment costs for the whole programme, then the investments are justified even if zero gross benefits came from all other interventions. For example, in Raitzer's (2003) study, the benefits from just four interventions that were classified in the most credible group produced a benefit-cost ratio of 1.94. This means that the benefits from just these four interventions more than justified the costs of all 56 CGIAR investments. This does, of course assume that projects cannot show a negative gross benefit e.g. increased soil acidification.

Part 3 Valuation of Open Data, Data Governance, Access and Investments

Value of Open Data

A McKinsey (2013) report on Open Data focused on quantifying the potential value of using open data in seven domains (i.e. education, transportation, consumer products, electricity, oil and gas, health care, and consumer finance). They estimated that open data had the potential to produce an additional \$3 trillion annually across these domains globally.

Commissioned by Omidyar Network and building on the approaches used in the McKinsey report, Lateral Economics (2014) provides an economic estimate of the value of open data to the G20 economies. Although recognising that providing exact estimates is challenging, even in developed economies with well-developed statistical systems, they stressed that the breadth and depth of value highlighted in the report are dramatic and provide a compelling case, even in times of austerity, for policy makers to extend open data and benefit from the substantial rewards on offer. Through the review of literature, modelling and use of case study approaches, they suggest that open data promotes many of the themes of the G20 (e.g. trade, finance, fiscal and monetary policy, anti-corruption, employment, energy, and infrastructure) and potentially could contribute to over half the G20's 2% growth target.

Later studies have reported similar positive outcomes although there is some variation in terms of the range of impacts. For example, the European Open Data Portal (2017) provide estimates ranging from 0.1 to 7.2% of GDP, whereas the Organisation of Economic Co-operation and Development (OECD) estimates a range of 1% to 2.5% (OECD, 2019). In the UK, HM Treasury (2018) reported that data driven decisions can result in 5-6% higher output and productivity.

In order to understand how value of data is generated and what approaches can be taken to measure this value, Wdowin and Diepeveen (2020) conducted a wide-ranging review of literature. They concluded that initially it was important to make sense of how value from data is

created and to identify the value for the different stakeholders across the value chain. They also recognised, however, that this is challenging due to the heterogeneous nature of data and the fact that its value is context-dependent. They therefore recommended that further study is required into the different types and uses of data in order that typologies can be developed.

The Government Office for Science (2018) identify a number of benefits to data holders including: reputational gain; benefit to humanitarian causes; reduced operational costs; enhanced innovation; social welfare; enhanced public service delivery; and in some cases direct revenue returns. Bergemann and Bonatti (2018), however, remind us of the importance of market structure (e.g. monopoly, oligopoly, competitive) and resulting pricing strategy in understanding how the value will be shared along the value chain.

Wdowin and Diepeveen (2020) summarise that there are also a number of potential costs associated with open data, such as: loss of freedom; legal liability of making data more open and the legal uncertainty surrounding its management; and the investment needed to measure and analyse data in order to inform decisions for both consumers and business.

Wdowin and Diepeveen (2020) suggest that it is helpful to think in terms of a 'value data chain'. Indeed, the OECD (2019) concluded that the value depends on how and where in the value chain data is used. Short and Todd (2017) elaborated on the challenges of understanding value along the data chain as it will be influenced by various components such as: usage type and frequency; content; age; author; history; reputation; creation cost; and revenue potential. This idea was reinforced in a report by PWC (2019) who extended this analysis and reported that the value is linked to issues such as exclusivity, timeliness, accuracy, completeness, consistency, use restrictions, interoperability, liabilities and risk. Mawer (2015) commented that the value of the data is believed to increase as it progresses further along the data value chain. They identified six steps that lead to the ultimate value of the data (i.e. raw data; data processing; integration; analytics; actionable insights; action – all leading to overall potential value). However, Wdowin and Diepeveen (2020), citing Spiekermann et al. (2015), conclude that measuring along the 'data value chain' is further complicated as heuristic and psychological biases can affect preferences and therefore individual valuations of personal data.

In order to measure the value there have been a number of different approaches used. Where data can be reproduced or replaced, PWC (2019) report that the cost approach can give useful upper and lower valuation bounds. Mawer (2015), however, suggest that data should be viewed as an intangible asset from which an income-based valuation approach can be taken based on estimating future cash flows. One of the key challenges with this approach is that it is challenging to estimate a service length life and therefore rate of depreciation of such an intangible.

The third and most common approach is to use a market-based approach to value data; this can be achieved by either directly using a market price or indirectly taking the market price of an equivalent product. However, this can be problematic due to the characteristics of the data. For example, public goods (no one is excluded from using it and all can use it at the same time) by definition do not have a market price, while to measure the indirect value of a club good (where people can be restricted from using it but all those that have access to it can use it at the same time) can be challenging as the value of the data will depend on the use value to each individual user.

Slotin (2018) broadens the issue past that of just an economic valuation and proposes an impact-based approach where the value is based on the assessment of the causal effect of data availability on outcomes. What is clear from the literature, however, is that most estimates of valuation are economic and market-based, and there is limited discussion on the social and public evaluation of data (Wdowin and Diepeveen, 2020).

At a more micro level, a number of studies have looked at the value of open data in relation to particular projects. For example, Deloitte (2017) assessed the value of Transport for London's open data and digital partnerships. They estimated the time savings for network passengers to be between £70 and £90 million per annum, which increases to £130 million per annum when wider effects are accounted for.

Similarly, the ODI commissioned a number of interim evaluations (London Economics, 2019; Frontier Economics, 2019; London Economics, 2020) to estimate the economic and social impact of its R&D programme, which was developed to support and build upon the UK's strengths in data. In reality, however, the main focus of the evaluations has been on valuing the economic rather than social impact. All three evaluations undertook a case study approach, supported with reference to quantitative data, where available. For one particular project, a service delivery model in North Lanarkshire, if scalable, was estimated to have a potential return on investment of £26:1. This estimate was based on a net present value over five years. The challenge with this estimate, however, is that it is based on the assumption that similar projects run in North Lanarkshire could be rolled out to all other councils and therefore extrapolated to generate similar returns.

Valuing Governance in LMICs

Very little attention has been paid to data governance issues in developing countries. Basu (2004), for example, explicitly considers 'e-governance' in developing countries. However, although he does note the importance of property rights over one's data, the paper fails to distinguish between 'government' (the practical management of data) and 'governance' (the legal and ethical management). As a result, the paper's conclusions on data governance and access amount to a restatement of the OECD principles on public sector use of data.

Two rare examples are both commissioned reports rather than academic papers. The Wellcome Trust (2015) looked specifically at the impact of data access arrangements on research performance in LMICs. It argued that restrictions on data sharing were generally unwarranted (that is, they were due to personal/institutional barriers rather than legal or technical ones) and did significantly limit data value.

The Wellcome Trust (2015) report concentrated on research data collection, rather than the deployment and re-use of operational data, and so it is not clear how far this is applicable to BMGF operations. Other studies on public sector data use in high-income countries (HICs), such as Hafner et al. (2015), suggest the institutional barriers are fairly universal. These studies, however, are concerned with public sector bodies maximising data use for public benefit, which may not be relevant to all agricultural interventions.

GODAN (2016) directly addressed the practical and ethical issues surrounding better use of open data in LMICs. As in Wellcome Trust (2015), they observed that institutional and operational factors are more relevant than law, but argue that, in the case of commercial exploitation of agricultural data, access limits are not the problem. On the contrary, they argue that inadequate or unclear rights to ownership have limited the gains to farmers, have allowed inappropriate exploitation of data, and so have damaged trust in data sharing per se.

The difficulty is that improved protection for farmers may not be compatible with open data goals, and they suggest that Fairtrade-type schemes might be a way to square the circle.

This conclusion arises from GODAN (2016)'s preference for open data outcomes where the benefit of data sharing is experienced as a public good. Other perspectives might suggest other solutions. For example, there is concern over the financialisation of development aid in Africa. The logical conclusion of the financialisation model is that farmers and data aggregators should be looking to commercial agreements to unlock value for both parties so that data sharing is in everyone's interest. However, as GODAN (2016) argue, this sort of model assumes a balance of power and information between the negotiating parties, which is unlikely to be the case.

Finally, Kuteesa and Kyotalimye (2019) promote the thesis that data sharing in Africa is primarily limited by resource constraints: inadequate investments in data collection and distribution, and in the training of data holders and users to extract value. In this perspective, the governance of the data is almost an irrelevance; the exploitation of the data, for public benefit or private interest, is practically limited by skills and infrastructure.

In summary, Wellcome Trust (2015) articulates a public-sector perspective which argues that institutional factors are key barriers to data sharing; GODAN (2016) argues that over-exploitation by self-interested data aggregators limits sharing by damaging trust; and Kuteesa and Kyotalimye (2019) argue that there are substantial practical limitations on anybody to exploit data, for good or ill.

Access to Data through ICT Investments in LMICs

Given the limited number of explicit studies on governance, we instead consider how data sharing investments more broadly have been seen; this might give some insight into how improved access to data has affected development or other social goals.

In terms of LMICs, early studies in this area looked at the broad relationship between ICT investment and productivity or growth. This is well founded in the economic growth models of Romer and Solow, which have been widely used to study development; such studies are also aided by the ready availability of high-level economic statistics for many countries over long periods.

Lio and Liu (2005) model aggregate production functions across multiple countries, and argue that ICT does provide a positive return on investment (albeit lower in LMICs than in HICs). Ngwenyama and Morawczynski (2009) investigate macroeconomic effects of ICT investment in Latin America, using a variety of techniques including econometrics models. They find that the effectiveness of such investment is strongly determined by the existing economic conditions and infrastructure at the time of the investment.

However, as with most macro-econometric studies, the analysis in these papers is of strictly limited value: countrywide averages are the data source, and there is no useful information to be gleaned on the impact of investment timing, type and context. To get useful information we need to focus on the case-study literature.

Van Zyl et al. (2012) identify six types of ICT investment affecting agricultural production:

- Information systems such as geographical information systems (GIS)
- ICT-enabled learning and knowledge exchange
- Modelling solutions
- Sensory and proximity devices
- ICT-enabled networking solutions
- e/mCommerce tools

Different elements apply at the pre-cultivation, cultivation and post harvesting stages. Although Van Zyl et al. suggest there are many examples of transformative ICT, as the source report² no longer appears to be available, the generalisability of these examples is unknown.

GODAN (2018) used data flows to structure their conceptual analysis of the opportunities from improved data access. They distinguish between 'local' flows (within the farm's own information network, 'imports' (data coming from outside to benefit farmers), 'exports' (data sent from farms to aggregators for tailoring of services, and 'incidental' data (exported data in aggregated form primarily used for modelling and policy analysis).

Deichmann et al. (2016) use a different structure, focusing on the outcomes of improved data access rather than inputs and using the standard economic theory of transaction costs to show how change can occur. They classify the outcomes as: improvements in access to information resources ('inclusion'), better use of information to achieve outputs ('efficiency'), and the market opportunities offered by being able to scale up applications to very large audiences ('innovation'). As both Deichmann et al. (2016) and GODAN (2016) note, the exporting of data can cause concerns for the farmer who sees potential commercial gains for others rather than his or her own benefit.

Carletto et al. (2014)'s detailed study argues that there is a fundamental problem to be considered before the impact of improved data access can be established; this is the lack of quality agricultural data. Moreover, collecting that data is just a part of the puzzle; significant problems exist in sharing that information with organizations that can exploit it for the benefit of farmers. Kuteesa and Kyotalimye (2019) argue that smallholders face substantial time pressures in addition to the uncertainty over the value of data sharing; hence they suggest that ICT projects that require the co-operation of farmers need to ensure that their participation provides a direct return in terms of the time investment.

Turning to the question of ways to facilitate data sharing and access, much attention has focused on the use of mobile phones as a transformative technology. Aker (2011) was one of the earliest articles to discuss the potential of mobile technologies. It noted that government attempts to invest in ICT for agricultural development were accused of being both too limited and too unwieldy, unscalable, and difficult to hold to account. The wide-ranging meta-analysis of Deichmann et al. (2017) seven years later came to the same conclusion: "these innovations often fail to scale-up and achieve wider acceptance" (p.32). Panganiban's (2018) case study of the Filipino government's ICT programmes drew much the same conclusion,

² The full report no longer appears to be available. Yonazi, E., Kelly, T., Halewood, N. and Blackman, C. (Eds.) (2012). *eTransform Africa: The transformational use of information and communication technologies in Africa*. Washington DC: World Bank and African Development Bank, with the African Union.

Aker (2011) highlights the massive growth in mobile phone use in developing countries since 2007, and argues that SMS and other mobile-phone based communication substantially leverage agricultural extension strategies. However, she also finds that there is not a single path to the benefits: some mobile-based solutions (such as weather reports) can replace traditional information transfer, but others (such as planting advice) need to be seen as complementary to traditional agricultural extension programmes.

At the time of writing (2010) this was largely speculative; the smartphone market, driven by the launch of the iPhone in 2007, was still well below maturity and dominated by high-end consumer products for high-income countries, and the development of apps focused on wealthier consumers. However, by 2016, 70% of the poorest 20% in developing countries had access to a mobile phone of some description (World Bank, 2016; cited in Deichmann et al., 2017), with SMS messaging to support agricultural extension being the main component (World Bank, 2016; cited in Deichmann et al., 2017). Of the 26 successful ICT projects listed in Deichmann et al. (2017), all but eight exploited the availability of mobile phones.

Aldosari et al. (2017) also found that respondents to a survey of Pakistani farmers were more likely to see mobile/internet communication as helpful to agricultural extension programmes; unfortunately, they then only asked follow-up questions about the value of TV and radio broadcasts, and so the impact of other communication channels was not evaluated. However, Panganiban (2018) interviewed scientists, farmers and policy makers, and concluded that innovative internet-based technologies are best seen as complementary to traditional media, rather than substitutes.

Panganiban (2018) did show that data access underpinned by better ICT support improved the return on agricultural extension. A large part of this depended on the development of 'farmer's information technology services (FITS) centres'. These were both physical facilities with ICT resources and online/passive learning tools for the harder-to-reach areas. Panganiban (2018) notes that over half of the municipalities had FITS centres set up, but it is less clear how well these worked in more remote areas, or whether the model could be translated to sparse settings of Africa, for example.

Summary

This chapter has reviewed the literature from three perspectives; first it examines the theoretical foundations of economic evaluation in order to identify best practice; second there is a review of the practical applications of economic evaluation in low and lower-middle income countries; finally the review focusses on issues in relation to valuation of data governance and data access.

In terms of an economic evaluation, it is clear that for the 'best' results, it is important to plan the evaluation in the design phase of the project and, as such, collect the required data throughout the length of the project. However, for many projects, evaluation is often considered ex post and as such there are a number of practical approaches which can be used to generate useful estimates; these include undertaking a case study approach, using a mix of methods and triangulating results where possible.

The literature in relation to economic evaluation for low and lower middle income countries demonstrates a mix of approaches used to evaluate both programmes and projects. The studies reviewed use a variety of quantitative and qualitative approaches to generate their estimates. It is clear that there is no single solution that can be applied in all situations, and that different methodologies and approaches should be applied depending on the context

and data availability. The work reported here is of limited value to this study as it is very much focussed on evaluating agriculture projects as whole. This is in contrast to this evaluation which is only evaluating the data governance sub-element of a project, which itself brings in additional complexities. The work of Raitzer (2003) and subsequently Maredia and Raitzer (2010), however, has been helpful to this study. Their framework which assesses the credibility of previous economic evaluations was a useful lens from which to assess the credibility of the design approach taken here.

The final part of this chapter discussed of the range of work undertaken on deriving the value of data. It is clear, however, that this is a very challenging and complex area with much of the research focussing on developed economies where both data and statistical infrastructures are more developed.

When looking at research into the value of data governance and access in LMICs, the review revealed relatively little research in to the valuation of data governance in particular. Therefore, this study aims to go some way to address this gap in the literature. In terms of data access, much of the literature focusses just on the role of the farmer as end producer beneficiary. Although farmers are an important group of users along the data value chain, in the SSHiE project it is clear that there is value to a number of other important stakeholders (e.g. policy makers, researchers). This study goes some way to providing a fuller understanding of value to different stakeholders along the data value chain.

The following section outlines the approach taken in developing a formal framework which could be applied in the evaluation of the value of data governance in donor funded projects.

Objective 1: Develop, Test and Refine the Formal Framework

Logic Model/Theory of Change

The UWE and CABI team initially co-developed a logic model/theory of change upon which the cost and benefit framework would be built. Where reasonable to do so, it was agreed that quantitative estimates would be generated for the following:

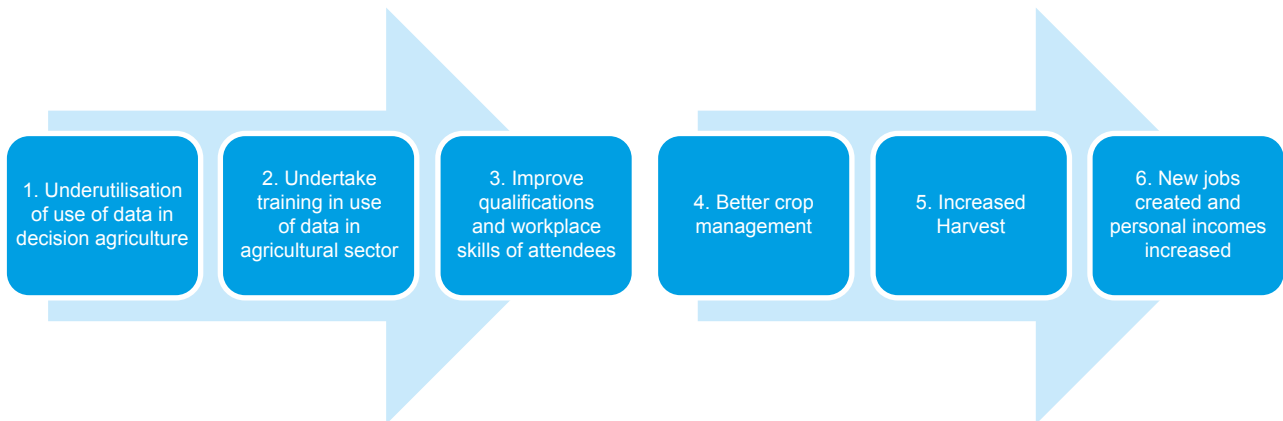
Table 1: Logic Model

Term	Definition
Inputs	Resources used to deliver the project
Activities	What is delivered to the recipient
Outputs	What the recipient does with the activity
Intermediate outcomes	The intermediate outcomes of the project produced by the recipient
Impacts	Wider economic and social outcomes

Source: HM Treasury's Magenta Book

The cost/benefit framework would be used to evaluate a theory of change/logic model created to map how improvements in data governance and access could lead to improved outcomes. As part of the mapping process, the team identified the start point (defined the challenge) and the end point (the overall objective), and then built the pathway between the two. Figure 3 provides an example of a logic model/theory of change created for the project to understand how capacity building investments in human capital could lead to improved downstream economic impacts.

Figure 3: Logic Model



Source: Authors' model

This framework was used to guide the approach used to quantitatively estimate the impact. However given the limited availability of data and the specific challenges of evaluating the value of data governance, it was recognised that any quantitative evaluation would be partial and at best illustrative of the added value of improved data governance. Therefore, it became clear that it would potentially be of more importance to gain a qualitative understanding of the mechanisms through which value was created. To do this, however, it was recognised that there would be a need for a more specific framework to be developed.

‘Five Safes’ framework

Following iterative developments, the UWE team redesigned Ritchie’s (2017) ‘Five Safes³’ framework in order that it could be applied to monitor and evaluate the value of data governance and data access within CABI/ODI projects and programmes. The ‘Five Safes’ is used in the design of data strategies for organizations as diverse as the UK Data Archive, the Australian Department of Social Services, the Wellcome Trust, and the Nepal Injury Research Centre; however, it had not been formally applied to evaluating data governance within projects.

The following sections provides a brief overview of the UWE ‘Five Safes framework’, before detailing developments made in order that it can be applied to evaluating the value data governance and access for a range of CABI projects and programmes.

The Five Safes is a recognised framework that can be applied to plan and design data governance and data access.

Table 2: The Basic Five Safes Framework

Safe projects	Is this appropriate use and management of the data?
Safe people	How much can I trust the data users to use it appropriately?
Safe settings	How much protection does the IT environment afford to the data?
Safe outputs	How much risk is there in the outputs of the access breaching confidentiality?
Safe data	Is the level of detail in the data appropriate?

³ For a full discussion see Desai, T., Ritchie, F. and Welpton, R. (2016). Five Safes: designing data access for research and Ritchie, F. (2017). The ‘Five Safes’: A framework for planning, designing and evaluating data access solutions

The five dimensions within the framework can embody a range of values. ‘Safety’ should not be seen as a specific state, but rather as a measure. For example, ‘safe data’ does not mean that the data is non-disclosive; rather it is the dimension under which the safety of the data can be assessed. The model also does not specify how the dimensions should be measured. For example, it may be appropriate to assess the safety of the data using subjective scales (e.g. from very high to very low), or alternatively more formal statistical models could be applied. The outcome of which, however, is that the user has a better-informed idea of ‘more safe data’ and ‘less safe data’.

When considering any data access solution, it is important to consider all five dimensions, even if it is just to note that a specific dimension is not relevant. However, within the broader framework each element should be evaluated individually to identify its own risk characteristics and evidence of appropriate practice.

To enable it to be applied to evaluate the value of data governance across a range of CABI projects and programmes, a generic framework with example questions was created. Table 3 below describes the first iteration of the framework developed for this project. This reconfigured framework, which for the first time includes additional sub-domains, was redeveloped in order that the value of data governance and access can be evaluated across the whole data lifecycle, including planning, design, collection and dissemination.

The redeveloped framework also includes examples of the types of questions that could be asked. The precise nature of the questions asked should vary depending on the type of project/ programme being assessed, the stage of the data governance and access lifecycle being assessed and the roles and responsibilities of the specific individual interviewed.

Table 3: Specific Case Study Questions

Domain	Sub domain	Example questions
Safe projects	Project planning	<ul style="list-style-type: none"> • What was the approval process? • How did it facilitate or delay the project? • What processes have been set up to make the next iteration more efficient?
	Data management plan	<ul style="list-style-type: none"> • Did you have a data management plan at beginning? • Were all stages identified? • How much did this change?
	Approvals process	<ul style="list-style-type: none"> • Who was responsible for the approval process? • Was the approval process developed from scratch? • How was the advice of specialists used?
	Public Engagement	<ul style="list-style-type: none"> • Is explaining your role to the wider public part of your project’s objectives? • How do you engage with the public?

Domain	Sub domain	Example questions
Safe people	Governance	<ul style="list-style-type: none"> • What model and structure of data governance and access is used? • How many people are employed in data governance functions and in what capacity? • What proportion of their time was spent on data governance issues?
	Training	<ul style="list-style-type: none"> • What, if any, training is provided for the following groups: <ul style="list-style-type: none"> - Data collectors - IT staff and data processors/stewards - Users - Is this training developed and delivered in house?
	Access	<ul style="list-style-type: none"> • How do you differentiate access privileges by type of users? • What systems and procedures are in place to ensure data users operate in an appropriate manner?
Safe data	Data quality	<ul style="list-style-type: none"> • How is the flow of data processed in the project? • How is data quality managed? • How is disclosure risk in the data managed?
	Data need	<ul style="list-style-type: none"> • How do you limit the level of detail available depending on the type of user and level of need? • What is the most detailed level of data made available to researchers? (e.g. geography)
	Compliance	<ul style="list-style-type: none"> • What are the mechanisms through which breaches of data governance procedures are enforced? • What are the range of sanctions that can be applied in the case of a breach?
Safe settings		<ul style="list-style-type: none"> • How do you make data available from a safe setting? <ul style="list-style-type: none"> - From where can data be accessed? - What IT systems do you use to limit unauthorised access? • What are the challenges to users safely accessing the data from a safe setting? • How do you enforce standards and policies in relation to accessing data in a safe setting?

Domain	Sub domain	Example questions
Safe outputs	Disclosure risk	<ul style="list-style-type: none"> • Are clear standards set and adhered to in relation to disclosure control? • What type of output checking for disclosure control is undertaken before release? • Are researchers trained in checking outputs for disclosure risk?
	Building capacity	<ul style="list-style-type: none"> • Is data created specifically for a project retained after the project is completed? • Are users allowed to archive their workspace (including code) once the project has finished? • What additional products and services have been developed as a result of using the data?
<p>Five Safes: questions to be asked for each domain</p> <ul style="list-style-type: none"> • How much did the processes you commented on contribute to the costs and outcomes of the project? • What have you learned and what you would do differently next time? 		

Test and Refine the Combined Framework

The initial framework was developed and sent to CABI and its partners for review ahead of its formal application in a case study context. In line with the action research methodology followed in this project, the intention was that the Five Safes framework should be tested, reflected upon and developed in partnership with all engaged parties throughout the length of the project.

The initial framework was accepted as appropriate ahead of the first interview and following a brief discussion at the start of the meeting, no further amendments were suggested prior to its pilot application. Within the interview the participants were questioned on an understanding of the framework, and as an initial framework it elicited little opinion one way or the other.

The results from the initial round of interviews were written up using the Five Safes framework to present the results. This presentation allowed for the identification of gaps in the data and therefore informed the subsequent planning of the second round of interviews and follow-up questionnaire. In order to prepare for the second round of interviews a questionnaire matrix was created, mapping activity, outputs and outcomes from each work package against the Five Safes framework. This further enabled the identification of additional information needed. Table 4 provides an illustrative snapshot of the data matrix used to map work packages to the Five Safes framework, to evaluate data already collected and show where additional information was required.

Table 4: Work Packages (partial) mapped to logic model and Five Safes

Logic Model						Five Safes				
Work Package	Inputs	Activity (What was done)	Outputs	Intermediate outcomes	Long-term outcomes	Pro.	Pep.	Set.	Out.	Dat.
WP 1. Develop Policy	Number of people Total cost Other tangible inputs (including from previous stages)	Open Data Sensitization Workshop (awareness raising)	Number and type of attendees Policy documents/ documents guidance produced	Common basic understanding of data access issues among relevant decision-makers so that decision-making is improved	Capacity: individuals who understand value of open data		x			
		Development of Coalition of willing	Identification of partners Identification of objectives Identification of responsibilities	Agreement on project aims, providers, beneficiaries, operating methods, so that future discussions do not need to revisit basic points	Example of pro-active governance to wider community	x	x			
		Consultation workshop on data access and sharing	Number and type of attendees Policy documents guidance produced	Common basic understanding of data access issues among relevant decision-makers reducing time to get agreement in future	Guidelines for the wider community	x	x	x	x	x
		Establishment of data sharing taskforce	Taskforce membership, longevity Policy documents/ guides produced	Recognised authority for project decisions reducing approval time. activity	Recognised authority for decisions on data sharing in the wider community	x			x	x

A simplified version of the full framework was then sent to the interviewees ahead of the interview to indicate to them the topics that would be discussed during the interview. Although the framework was helpful to the researchers in planning for the interviews, following feedback from the interviewees it was clear that the matrix was less so for them.

Following interviews with a number of stakeholders in the country, it became clear that alongside the challenge of developing and ratifying the policy, there was also the need to develop a framework of how the policy could be implemented. The CABI team assessed the challenge under three priority areas. For the sake of this project, these priority areas have been remapped against the Five Safes framework:

- Technology Infrastructure [Technology] – safe setting
- Policy Implementation Guideline [Processes] – safe project, safe process
- Upskilling Plan [People] – safe people, safe outputs

Objective 2: Evaluation of value and the mechanism of diffusion

Objective 2 is the “identification of stakeholder value of data governance and the mechanisms through which this is created”. In simple terms the project aimed to identify:

- What was done
- How much it cost
- What it achieved
- Lessons learned (including future avoidable costs)
- Capacity needed/built (for project, for organizations, for sector generally)

This section describes the case study chosen and the methods used for the qualitative and quantitative evaluation.

SSHiiE Overview

The SSHiiE project is a \$1.5 million project led by GIZ. The project was initially due to run from November 2017 to April 2020; however a short extension was granted until August 2020.

GIZ’s role was not to directly deliver the outputs but to coordinate the various activities associated with each of the work packages. The overall objective of the project was to provide proof of concept for the impact of leveraged geo-spatial soil, agronomic, and health data on transformative agricultural development.

The specific aims were:

- increase the quality, efficiency, and reach of government and private sector services;
- improve national and regional resource allocation and policies; and
- increase the quality, availability, and utility of data and evidence.

In order to meet the aims, the project was further broken down in to six work packages:

- Analyse data needs and bottlenecks, assess gaps for the “Use Cases,” and mobilize stakeholders.
- Craft a policy for stakeholders across organizations to share data and collaboratively develop solutions to easing Ethiopian soil system bottlenecks.
- Provide technical and operational support for, and facilitate inter-institutional consultation needed to establish, the Ethiopian Soil Resource Institute (ESRI) as the national IT host for a centralized data repository.
- Help convert the Ethiopian Soil Information System (EthioSIS) into a geo-spatial database with online access to serve users from a variety of sectors and with differing needs.
- Develop IT-assisted Use Cases to provide proof of concept for the impact of geo-spatial soil, agronomic, or health data.

- Link up and collaborate with several currently active investments that have complementary initiatives relevant to soil health, fertilizer use efficiency, and transformative agricultural development.

The first three work packages were included in the evaluation, but not all activities within these packages were included. This was because certain activities within the first three work packages, as well as all activities in the three omitted work packages, were judged not to be primarily focussed on issues of data governance.

In order to evaluate the value of data governance in the SSHiE project, initially a number of reports were analysed. These include:

- GIZ Ethiopia Soil Health Proposal (July 2017)
- GIZ Progress Report (August 2018)
- Soil and Agronomy Data Sharing (SADS) Policy Implementation Report (October 2019)

The main stakeholder groups, both in terms of delivering the project and/or benefitting from the project are

- the donor community
- the project delivery team
- the Government of Ethiopia (enabling it to make better evidence-driven decisions)
- development partners and agricultural investors (providing better information for land use and management plan formulation)
- extension staff including the scientific community (providing better site-specific information and advice); and
- farmers (improve how they manage land resources).

Methodology: qualitative approach

Due to budgetary and time limitations, it was not feasible to interview individuals from all stakeholder groups or to identify a representative sample. Instead, a convenience/ snowballing approach was taken. The agreed approach was initially to interview key personnel from the donor community (CABI Project Representative) and the Project Delivery Team (GIZ). This would directly generate information on the value and mechanisms of data governance in the SSHiE project. Equally importantly, these initial interviews helped identify other key stakeholders to consult.

Four interviews with stakeholders were completed, involving seven individuals. As well as the individuals from the donor community and the project delivery team, interviewees were from the Natural Resources Management and Food Security Sector (NRM&FSS) for the Ethiopian Ministry of Agriculture (MoA); the International Centre for Tropical Agriculture (ICfTA), also representing the Coalition of the Willing (CoW), a voluntary stakeholder group set up to support the delivery of the project; and the MoA's Soil Resource Information and Mapping Directorate (SRI&MD). The interviews were recorded and transcribed. Interviews are indexed I1-I5 below for verification purpose, but note that these may relate to the comments of more than one individual at an organization.

The interviews were structured around the Five Safes and the basic model of change to evaluate inputs, activities, outputs and outcomes. The aim of the interviews was to identify the following issues in relation to data governance and access:

- technical solutions that have been shown to be useful and transferable to other contexts
- non-technical options for capacity-building (training, templates, standards)
- institutional and/or financial barriers to accessing data
- where possible, the costs and financial returns associated with any of these
- whether change is best implemented at the level of the agent or through strategic direction from funders

Following an analysis of the interviews, the project team developed a questionnaire (see Appendix 2) to generate more specific measures of the value of data governance. The questionnaire was sent to twelve experts, identified by the CABI team as being able to provide detailed insights in to the project. The aim of the questionnaire was to produce direct and indirect estimates for the different activities in the three work packages under review.

To produce some tangible estimates of the costs and benefits of particular activities in the absence of hard data, the questionnaire asked individuals to make subjective judgements. For specific questions they were also asked to consider the counterfactual (i.e. what would have been the likely outcome had this not happened). To assist in the analysis of those subjective judgements, individuals were also asked to rate their level of confidence in providing their estimate. These confidence estimates were then applied to weight sample averages.

The questionnaire and guiding instructions were initially tested by three members of the CABI team, two in the UK and one in Ethiopia. It was particularly important to get feedback from an Ethiopian resident in order to test the questions for clarity and cultural appropriateness. Following comments, the questionnaire was revised ahead of final circulation. In particular, the full questionnaire was not sent to all individuals because different stakeholders would only have knowledge of some aspects of the project and the research team wanted to make the questionnaire as parsimonious as possible.

Prior to sending the questionnaire to the twelve experts, the UWE research team created a You Tube video to help the experts respond to the questionnaire. In the video, the researcher explains the objectives of the study, talks through the structure of the questionnaire, and completes an example response to one of the questions.

The participants were sent their personalised version of the questionnaire shortly after being sent the link to the video. Following the receipt of the questionnaire, a CABI representative in Ethiopia contacted the experts individually to offer support and encourage them to return the questionnaire. The experts were chased on at least two separate occasions. Ultimately, eight of the twelve experts provided a response.

Other information sources

The next two chapters present the qualitative and quantitative analysis, respectively. Both chapters are supplemented with information from other project documentation sources including:

- GIZ Progress Report - August 2018
- SADS Policy Implementation Report - October 2019
- Summary of expenses for CoW and SIRM to improve Data Framework In Ethiopia

The GIZ progress report is the regular form used to provide updates to the Gates Foundation program officer regarding progress made towards achieving the project's stated outputs and outcomes. The SADS Policy Implementation Report (PIR) was produced as a result of a mission to Addis Ababa. The summary of expenses to improve Data Framework in Ethiopia was provided by GIZ. However due to concerns in relation to supplying personal and commercially sensitive data, only high level budget overviews were provided. Data was restricted to just high level summaries of the major fields of activities that were carried out to improve data governance.

Qualitative findings

In order to evaluate the value of data governance in the SSHiE project, documentation was analysed and supported with reference to interviews with key personnel.

Safe Projects

The interviews revealed that the genesis of the project went back to when the government was preparing its ten-year soil sector strategy. Three main issues were identified:

- they could not find comprehensive data sets to make the strategy evidence-based
- data collected through public funds remained with researchers and was not necessarily shared
- data was collected in different formats, for different purposes, and quality was not consistent (11)

The inference is that, at least from the government perspective, the full value of the project will only be achieved once all three issues have been resolved.

What was clear early on from all the discussions was that there was no agreed definition of what value was and therefore how it should be measured. Within a standard economic analysis, value is usually measured in monetary terms by primarily looking at the direct costs and benefits where the concept itself and resulting impact can be clearly traced. In the project this caused particular problems as it attempted to value and monetarise something which by its nature is unquantifiable and its impact challenging to isolate and trace. To overcome this challenge, the study focussed on value as defined by the different stakeholder groups. Conversations with various stakeholders revealed that they interpreted this concept more widely and that the value created was multifaceted. For example, one interpretation focussed on the individual and the research process.

“the first collector of the data extracts value... The value starts to be seen as data starts to be shared in a more consistent way and gets used... Some of the difficulty in using the word value, is people see value as monetary value immediately, so I have been trying gently to introduce the different concept of research value” (12)

Part of this value was the change in culture that the project was engendering. It encouraged individuals and organizations to open up to a more formal process of data sharing.

“So the value of this initiative is really, is very good for me... it’s creating the value... to data and it’s creating... some sort of awakening for others on how data is important and how it adds value when we share data among stakeholders” (11)

Others focussed more on the project’s impact across the whole of society, with its potential to inform better policy.

“I think this goes far beyond the value for individual researchers and the paper... decision makers at federal and regional level become increasingly aware of the value of data... they base their decisions... on hard facts. Ethiopian citizens... will in the end benefit from better policy decisions” (15)

Indeed, the wider economic, social and environmental value created was seen by ministry officials and other key stakeholders as an important benefit of effectively implementing the data sharing policy.

“data is very important... for security and economic growth.” (I3)

“by managing our soil resources we can improve production and productivity and combat the climate change problems nationally” (I1)

Despite the difference in views on what defines value, the concept of the project creating value was widely recognised.

“this data sharing policy I think is... valued by all the stakeholders across the country.” (I3)

It was also evident that some stakeholders perceived the spillover value to be far beyond just the scope of the original project.

“people at the ministry were saying ‘why only for soil and agronomy?’ And everybody was saying we need the policy initiatives for other data too, like, for the livestock, economy and other things... So I cannot really... underestimate the value of the policy” (I3)

This suggests that the indirect spillover value, triggering change in other sectors of the ministry, could be considerably larger than the direct value of the project itself.

One significant development in the project was the formation of the ‘Coalition of the Willing’ (CoW). The CoW was set up by a group of interested scientists/researchers to help inform the development of the data sharing policy and informally to share data to demonstrate the value of doing so. This had the direct effect of improving communication between stakeholders, ensuring buy-in for the project goals, and providing a forum for momentum about the project to be generated.

“the coalition is... bringing some new insights, new ideas, in relations emerging issues... It became broader and inclusive... supporting any emerging issues, big issues that the government is trying to achieve.” (I4)

An unexpected spillover effect was that this informal coalition had a life beyond the project and was able to contribute to the evidence base needed to address other contemporary challenges such as the Covid 19 pandemic.

“this coalition... has data scientists inside... so when Covid came for example, we said... we can help [the Government] map hot spots.” (I4)

This illustrates the difficulty of ascertaining value; a solution to improve within-project data governance found a substantial secondary purpose supporting the wider economy. The evidence was that long-term engagements and regularly bringing the diverse Ethiopian stakeholder groups together were breaking down barriers permanently.

“Incrementally you can definitely see and feel a difference in how people are talking about data sharing now... the whole [data] ecosystem is moving to a more permissive and a more... trusting approach.” (I2)

The role of the external donor community and international expertise were also critical factors to facilitating this change.

“You need some finance... You need expertise in the overall topic... we were very lucky to have... CABI and... the Gates Foundation who have the international linkages to bring in the right people... [their] involvement was critical to give it credibility” (15)

The project team had a critical role in providing support to identify, draft and engage stakeholders with the development of processes and process documentation needed for the successful implementation of the data policy. It was noted that, where the direct project team did not have the in-house expertise, partnering with internationally renowned organizations gave them access to a wider pool of expertise that could be used on the project.

“I also think it is crucially important to recognise that this type of expertise we don’t have in house... to make sure that we engage with an institution who can provide that expertise and who has the international credibility in doing so” (15)

It was recognised that value is generated by access, but levels of access to data can, and should, be restricted depending on levels of confidentiality. For example, the ODI data spectrum suggests that National Security data could form the ‘closed’ end of the data spectrum, whereas data such as wildlife images would sit at the other, ‘open’, end of the spectrum. But much data would fit somewhere between these two extremes; as such it could be shared, with protections if necessary, and used to demonstrate the value of doing so.

“we talked about open data, data sharing...even help them to go in their project. And that was a very good basis to show all real data, real evidence so that we convince people” (14)

Although the ODI’s data spectrum is a conceptual model rather than an empirical framework, a measure of increasing value can be inferred if movements along the ODI spectrum towards more open data are demonstrated.

“when we first started engaging in Ethiopia we talked about the data spectrum... this is very much a journey from closed data across towards open data... but this is definitely not... moving to a massive open data culture.” (12)

This statement suggests, however, that although the SSHiE project has contributed to moving towards a more open data culture, the journey is still in its early stages.

The project report in August 2018 notes that one of the early key achievements was the delivery of its Open Data Sensitization Workshop. This enabled the project to create awareness of the need to make soil and agronomy data accessible and thus help to transform agriculture and promote national economic growth. The workshop attracted 56 participants from 21 organizations including the national agricultural research system, governmental and non-governmental development organizations, higher learning institutions, international agricultural research centres, and local and international consulting firms and donor organizations.

This early public engagement was important to set the tone for the project and bring together the potential delivery partners for the project – the CoW. It also allowed the project team to identify potential barriers to the delivery of the project and openly address any concerns stakeholders may have had in engaging with the project and its aims and objectives.

“I think the Coalition of the Willing and the process of engaging with a coalition and researchers in using the data, to develop papers... to do on-the-job training and analyse and utilise this data... was probably one of the success stories of this project.” (15)

This sentiment received further support from representatives who identified the engagement as a critical factor in initially getting agreement and subsequently empowering stakeholders to take ownership of the project.

“the main thing... is engagement... We got to a point where everybody agreed. Everybody... that was involved from the Coalition of the Willing, in the development of the policy and the development of the standards and the development of all the protocols that exist so far, feel that they have got a stake.” (I2)

An additional success factor seems to be related to the choice of partners to work with. Combining experienced international and local institutions ensured that learning from previous projects was taken into consideration.

“Before this project there was one year on another project... that we implemented and it had a bit of a data sharing component.” (I4)

Working with experienced government and non-government stakeholders ensured that the correct skills, knowledge and levels of influence were in place to deliver the project.

“We prepared various documents, strategy, policy guideline documents, we follow the same procedure... because previously we have prepared various documents.” (I1).

It was also noted that there are a number of key challenges which potentially could limit the value of the project. One potential source discussed was the political and institutional instability in the country. The evidence here was mixed, as support for the project did not seem to be linked to a particular administration.

“We... started before the change [of the political landscape in Ethiopia]... It’s not really related with the change of leadership at the high level” (I3)

However, challenges with institutional arrangements were potentially more limiting and led to considerable delays as considerable time and money was spent on detailed engagements with one of the ‘championing’ departments.

“the Ethiopian Soil Resource Institute which this project was supposed to be supporting. And we started that process but then basically, politically the whole concept collapsed and if I would have known that, I would not have put so much energy and also financial or other resources into this process.” (I5)

One of the key drivers to the project's success was the momentum it created and maintained. The development of the SADS policy was cited as a good example of this, taking less than two months from the first inception meeting to delivery of the final draft output. However, there is some evidence that members of the project team felt that a broadening of stakeholder engagement could have increased the value of the project.

“the work to develop the policy itself last year was done over a... quite intense two month period. Ideally I would take longer and have more stakeholder engagement. There was not enough in my view engagement with the regions... There just wasn’t ... necessarily the budget and time to do it.” (I2)

The update report in October 2019 and subsequent interview in April 2000 confirmed that, even following this extensive programme of engagement, the policy itself was still to be ratified by the Ministry of Agriculture.

“with the help of CABI we are working on... an agreement of a national data policy and implementation procedures and guidelines...there still needs to be an official endorsement by the ministry of agriculture.” (15)

The delay in ratification was primarily attributed to Ministry advisors wanting to broaden the scope of the policy to include data for the whole of the agricultural sector.

“other ministers of agriculture’s advisors thought: this policy is good, why don’t we roll it out to whole of agriculture rather than just soil and economy? So that delayed it.” (12)

However, the majority of stakeholders, including the Ministry, were clear that the SADS policy needed to be implemented just at the level of the agronomy sector. That way lessons could be learned and applied a later date to data for the broader agricultural sector. Broadening the policy to cover the wider agricultural sector would just delay the implementation of the soil and agronomy policy. This was because it would require additional significant investment in both time and money, in order to bring together all appropriate stakeholders including from various government ministries.

Lessons learned for the organization of projects and investment in setting up processes

The first key finding was the importance of effective working links between local and international partners. The mechanism through which this works is threefold. First, it directly helps to secure sufficient funds needed to facilitate high level change. Second, it engenders trust with local stakeholders who are encouraged to take ownership of the project and co-deliver its objectives. Finally, it provides a conduit to share knowledge and bring in external expertise to deliverer specialist services.

The second key to maximising the value of this project was engagement with a diverse set of stakeholders. The challenge with stakeholder engagement is that it is time consuming and potentially costly. The evidence here, however, is that with more time and with wider engagement, particularly in relation to regional policymakers, the value of the project could have been increased.

Both of these findings suggest that time spent building initial relationships, clarifying areas of expertise, identifying and involving potential project ‘champions’, and agreeing goals should be seen as an investment. Although institutions, structures and individuals can change over the course of the project, leading to loss of focus or project creep, early investments in relationship management can help to mitigate against such risks.

There may be a conflict between aiming to achieve the direct project goals, and extracting a wider benefit to society. Maximising the direct value of the project may be achieved by tightly defining the scope at the beginning and wherever possible, resisting calls to widen the focus of the project. This approach, however, runs the risk of limiting the overall value the project could achieve by limiting the potential spill-overs to other areas of the economy and society. Again, an initial investment in agreeing project goals with a wide range of stakeholders can balance these internal and external objectives.

Safe People

The GIZ progress report notes the value of developing relationships with institutions who can then provide financial and other support. It was, however, noted that this can take a considerable investment in time and effort.

- “from the institutional side that is quite tedious, it needs a lot of patience and it needs many meetings, and ideally also a high level government representative supporting this.” (I5)

For example, in the initial workshop to launch the event, an address was given by the Honourable Chairperson of the Agriculture Standing Committee of the House of People’s Representatives. The event was attended by a number of other government officials and with 40 participants from 14 different national and international partner organizations. The project manager reflected that the quantity and quality of the representation at this event was influenced by securing such high level support for the event.

In the interviews, this point was picked up and elaborated on; in particular, the need to identify and work with a high ranking official to champion the project

“So it needs not a sponsor, it needs a champion...the national government within the policy is seen to really push it through.” (I5)

As well as key influencers who are willing to champion the project, the October 2019 report comments on the necessity to identify key blockers. By understanding and addressing concerns the blockers may have, these individuals can potentially become ‘powerful advocates for the policy itself’.

“because there were some individuals from some organizations were not able to share data so there were critics, some fighting, very noisy discussions but ultimately I think they were very productive discussions” (I4)

Also believed to be of importance in building momentum and attracting a diverse set of local stakeholders to support the project, was the attendance of donor organizations and international project partners at the initial events.

“we had representatives from... Gates and from US AID who are both funding this initiatives and that helped greatly to convince and influence the partners that there really is a need to work on more harmonised and unified protocols.” (I5)

Potentially of greatest value to the project was the creation of the CoW. This coalition was formed during the open data sensitization workshop to take forward the idea of enhanced accessibility and sharing of Ethiopian agronomic data and information.

“a group of people and institutions who were willing to do this on a voluntary basis, a coalition of the willing... this [was] the first step to populate these data repositories at the ministry and the research institute” (I5)

The CoW consisted of 26 participants from 17 major Ethiopian research institutions. It worked because there was a clear incentive for individuals to take part, inasmuch as they would be given access to other members’ data.

“we started as coalition of the willing...[working] to see how we can share data voluntarily.” (I4)

The importance of establishing the CoW was that it was able to take a 'bottom-up' approach and drive forward with localised data-sharing initiatives.

“the first big lesson learned for me is... engaging the ‘grass roots’ labelled stakeholder. We started with the data generators, data users themselves. We didn’t start from above and try to tell, you know, tell or inform the ones who publish the data.” (I3)

Simultaneously, however, the project also drove forward with the 'top-down' development of a formal policy framework developed in partnership with the Ethiopian government. This was overseen by a taskforce.

“The purpose of forming a task force is to identify the components of the data sharing policy... so that we can have a common vision and in order to utilise the policy for minimising data duplication efforts in the country.” (I1)

The project benefited from tight feedback loops between the bottom-up and top-down approaches. This was helped because the CoW was overseen by a committee which fed into the government taskforce. Indeed in many cases the membership of the CoW committee and Government task force overlapped.

“the ministry of agriculture created a task force... a lot of them are from the coalition of the willing... so we are even given a mandate to support on developing policy like data sharing policy for the government.” (I4)

It was clear that working with such diverse stakeholders and vested interests can be complicated, but ultimately through engagement, discussion and negotiation it was possible to align interests in order to jointly deliver the project.

“there is competitions in hand, in different organizations... so it’s a matter of balancing... they will have problems to solve and the way to solve it is like this... and if everybody agrees so we are on the same page you can move on.” (I4)

This 'bottom-up' and 'top-down' approach is helpful in this regard as it is mutually reinforcing. It takes the form of a carrot and stick approach: the carrot is demonstrated by what can be achieved when a CoW work together, which gives further impetus for the stick to be developed (i.e. national data sharing policy) and implemented.

“So it’s the carrot and stick approach. It’s all been about carrot at the moment. And we are waiting for ratification of the stick” (I2)

Despite the delay to the full implementation of the data policy, the SADS PIR details evidence of progress towards this aim. For example, there had been revision and modification of the Ethiopian Agricultural Transformation Agency (ATA) data sharing protocol and the Ethiopian National Agriculture Research System issued a letter requesting data to be shared by various stakeholders. Feedback from those interviewed also revealed the belief that CABI/ODI could help drive forward the implementation of the SADS policy through their knowledge, experience and high levels of trust they had built up in Ethiopia.

“[International organizations] have developed this trust with the research community here... they are sincere about this and that they are not here out of personal interest but that they really want to set up a system in Ethiopia that helps the country and helps Ethiopian researchers to make better use of their data and basically contribute to national development.” (I5)

The necessity of building trust, forming strong relationships and empowering stakeholders was clearly key to developing and potentially implementing the SADS policy in Ethiopia and therefore maximising the value of data governance delivered by this project.

“That’s the other resource, and is probably the biggest resource that is needed, that is trust... Data has a value, researchers value what they produce and they will not share it easily unless they trust people, and they have the trust that this trust or this data will not be misused or misappropriated and be analysed and utilised without them knowing, and without giving due credit to them.” (15)

The project team recognised that in order to address long-held cultural and social norms (e.g. in Ethiopia, historically data had been shared as a result of personal relationships rather than through any systematic process), it was important to empower stakeholders to develop their own solutions rather than implement them from afar.

“we also co-develop ideas... There are a lot of things that come... from discussions and the task force and also in the coalition.” (14)

It was evident that co-created solutions were enabled through an agile and flexible approach to managing the project.

“It needs a person who is willing to listen and to adjust the agenda and topics to the needs of the audience and not to bulldoze an idea through. It needs to be a participatory process, sort of co-development of this entire process” (15)

The project team were able to do this by following a three stage process with stakeholders - engaging in a discovery phase, developing and testing an ‘alpha product’, and then re-developing the product to meet stakeholder needs. This process was valuable for developing a product to meet the local needs, building trust and reinforcing personal relationships, developing a delivery partner network and transferring knowledge. The process was also valuable for building capacity to enable stakeholders to build on the work after the initial CABI project has completed.

The project team also had a significant role in building capacity through the identification of skills gaps and then co-developing materials and workshops to fill such gaps.

“we had a series of workshops where basically data scientists came in to provide training and to highlight the options that exists. What could be done with this type of data? And then basically helped the national scientists to analyse data with these novel approaches.” (15)

However, Ethiopia has strong regional ties, both institutionally and culturally. Training would not only be needed at the national level, but there would also need to be training focussed on upskilling researchers in managing local datasets; these can then be combined to piece together the regional and national pictures.

“The regions in Ethiopia have a lot of autonomy and obviously they have an infrastructure all of their own. So there is an ambition... [for a] network... regional coalitions of the willing” (12)

The issue of regionalism was a recurring issue within the interviews; engagement with regional as well as national stakeholders identifies as one of the factors of success.

“They come from the different regions within Ethiopia and different institutions particularly the research community. So... people from almost all regions... would... have been participating actively” (I3)

Building capacity through training was also considered from the initiation of the project.

“Capacity building started with popularising this policy, I mean creating awareness of the policy started with training of researchers” (I3)

However, much of the focus on training was for the long term sustainability of the project. One of the aims was to be able to train individuals who in turn would be able to share the knowledge within their institutions and with future generations.

“the sustainability of the institutional set up: we hope to achieve this through the on the job training. I mean these are people coming from national institutions who will return to their institutions and the hope is that they will also bring this newly acquired knowledge into their institutions.” (I5)

Without investment in these people’s skills, changes to the policy in relation to accessing, sharing and reusing data may be difficult, and gains may be limited. While developing training packages, however, the SADS PIR, emphasised the need to ‘keep it simple’. For example, the PIR suggests that training packages on ‘data management’ or ‘data governance’ should be broken down into short components, which would allow the testing of individual parts rather than the whole package. It was also recognised that there are already processes and guidance in place, and that anything created should complement these and be able to be adapted and incorporated into other processes should stakeholders wish to.

Lessons learned for organization of and investment in people

To extract the maximum value from a project, it is important to be able to initially identify top-level, key personnel who can ‘champion’ the project; this is an investment, to build trust and form long-term relationships with these key individuals. Local knowledge is extremely important in helping to identify key supporters and ‘champions’ as well as potential ‘blockers’. Building sufficient time and resources in to the planning of the project enables these relationships to develop.

Value is further created through the building of relationships and empowering a diverse group of lower-level stakeholders who can help drive the project forward. Being able to demonstrate high level support for the project, regionally, nationally and with the international community, can be extremely effective in supporting such developments at an operational level. It is also important to link this top-down and bottom-up approach by putting in feedback loops from one to the other.

There is also a need to invest in training and knowledge transfer activities, but these should initially be targeted at local individuals with specific expertise. They can become the conduit for future within-country knowledge transfer activities. In order to maximise value, there are a number of core areas for which training should be provided (e.g. data management, data processing); however, the approaches undertaken need to be flexible and targeted at local need. In order to deliver specialist training, expertise outside of the project may need to be drafted in.

The unifying theme for this domain is the need to invest significant time to build trust, be able to respond flexibly to meet the local need and empower all stakeholders to work together to co-create and co-deliver projects. Although difficult to evaluate, the benefits of this investment should not be underestimated, and it seems likely that this generates strongly positive returns.

Safe Settings

Different types of data and access rights require different approaches to making the data available within open, restricted or closed settings⁴.

The principle of appropriate access was central to the development of the policy.

“it is a policy document. It is clearly described that there is a confidential data section in the policy.” (I1)

In terms of practical implementation, one way to develop a safe model of delivery is to learn from other institutions who have similar data handling needs. Members of the Ethiopian Soil Resource Institute (ESRI) visited the Bangladesh Soil Resource Development Institute (SRDI) to learn how it uses a safe setting to manage restricted-access data.

Safe settings are often considered a technical or process matter. However, the SADS PIR highlighted the need for this to be accompanied by a cultural shift in relation to data access in Ethiopia. Prior to the commencement of the project, data sharing was often only advanced by the use of official letters to request that data be made available for a particular purpose. This process potentially created bottlenecks, particularly for those datasets where this level of formality may not be required.

“it was much more informal. So to get access to data was very much who you knew rather than the process that you followed, and you may or may not get access to data.” (I2)

There was evidence that this project had shifted perceptions of at least some organizations.

“far more organizations didn’t want to share their data because of... national security [and] various reasons... I think most people now can understand the benefits of data sharing individually as well as nationally” (I1)

To create a safe setting for restricted access, investment in technology is required; however, the consensus within the project team was that the technical issues were the easiest to solve as it was mainly a financial issue.

“I think the technical problems are the smallest. I mean these are technical issues, if you have money you can solve it. That’s not an issue. And money in this case is probably not really a limiting factor.” (I5)

The technical challenges of the project had largely been overcome with investment in two databases. These two databases were available to use in two government departments, with a further ambition of the project to link these two databases together.

⁴ We use these terms in the sense defined by the Open Data Institute’s Data Spectrum. ‘Open’ is unrestricted; ‘restricted’ means access is available to appropriate parties with controls in place; ‘closed’ means only the data holder/collector can use that data.

“there are 2 databases that are being built that are effectively the core repositories and 2 types of data, the agronomy and soil data. They are in effect the safe settings. They are taking data from sitting on somebody’s laptop or flash drive and actually putting it in a place where there are management protocols around it” (I2)

The creation of these databases, and subsequent sharing of data that was previously only available on local systems, has enabled restricted access to researchers, but has been limited to only those from the CoW.

“use information that otherwise would not have been available at that extent without this type of intervention because it was scattered among a whole range of different stakeholders and this initiative now this data has come together in these two data repositories and can be used for analysis.” (I5)

It was reported that the investment in technology could provide a ‘safe setting’ as issues around access and confidentiality would be automated.

“We have automated accordingly in the national soil information system. All this access, confidentiality and so on, it’s automated” (I1)

This may partially be true, but it is clear that not all aspects can be automated and further investment in people and skills will be required to build capacity to support and administer the technical solutions.

Within the Ethiopian setting, the evidence was that investment in technology was not the problem. The technological solutions were there for the data to be accessed in a safe setting; however, the supportive ecosystem (e.g. legal, cultural, ethical) needed development.

“The legal issues are critically important because you need a legal framework especially here in Ethiopia where the government and governance is very strong and people adhere to what rules and regulations are being set up... Intellectual property rights need to be protected and need to be enforced... but that’s a process which takes time and we are not yet quite there.” (I5)

Although recognising that there was more work to do on this front, there was some evidence of involvement of the wider ecosystem.

“we invited some lawyers... from data protection so that they can listen and we invited different fields... the discussion created good awareness among the different fields.” (I4)

It was clear that the focus of the project was the first-order challenge of creating and implementing the SADS Policy and providing the technical infrastructure; but project managers were also aware of second-order challenges around building capacity in skills and developing robust processes and protocols for accessing data.

“the first thing that’s been focused on is the technical investment and the hardware. The bit that’s missing... how do we put the rules and processes that need to sit and exist around that hardware in place? We are... not quite there yet” (I2)

This need to focus past the technical solutions was acknowledged in the SADS PIR. The suggestion was that in future, the role for CABl in Ethiopia would not be to invest in technology, but rather to provide advice around infrastructure, connectivity and the minimum requirements to enable data sharing in line with policy and procedures.

It was evident from the second round of interviews that focus on capacity building was now a key priority for the project.

“So the training, capacity building on how and which data should be handled... will be done now this year.” (14)

Lessons learned for the set up and management of safe settings

The key lesson learned is that value can only be achieved if individuals trust the processes to ensure that data will be handled appropriately. This may require investment in safe settings to provide trust through automation (as well as the efficiency that this approach can generate). Technical issues are probably the easier to resolve: from the discussion it was clear that consideration needs to be given to the wider ecosystem (e.g. legal, cultural ethical), and this may be a more tricky problem.

In the planning phase consideration is always given to the practical aspect of developing a solution. Our evidence suggests that there is also considerable value in the investment in second-order activities such as building capacity, both in terms of upskilling of staff and development of robust protocols and procedures for managing data in a safe setting. Much work has already been done on these issues globally and therefore in order to maximise the value creation in future similar projects, there is a real opportunity to learn from other agencies with experience of providing access to restrictive data in a safe setting.

Safe Outputs

Promoting the value of safe outputs (both the production of statistical analysis, and the production of databases for further use), can act as a ‘carrot’ to complement the legal ‘stick’ of introducing a new data sharing policy. For example, the GIZ Progress Report refers to the Data Mining project with the Ethiopian Institute of Agricultural Research (EIAR) and the International Centre for Tropical Agriculture (CIAT). These two organizations launched an initiative among interested researchers/institutions to pool agronomic datasets on fertilizer response trials. With expert support, Ethiopian scientists were trained to use data management and statistical analysis methods (SAS and R) to analyse large and varied agronomic datasets and subsequently publish the data in international journals. The training, however, was conditional on a willingness to share data and allow it to be uploaded into the EIAR database. This idea around providing incentives for stakeholders to take part in initiatives was also a recurring theme in the interviews.

“That was the underlying concept that you need, some sort of incentive for people to share.” (15)

The SADS PIR also reiterated the dual approach of carrot (value of output, kudos from sharing) and stick (data sharing policy). Real value could be created by demonstrating the ‘clear line of sight’ from collection to output and then to the re-use of data. This approach of incentivising others to take part through demonstration of the value that could be created was further supported in the interviews. Additional value can be created in terms of building capacity, with the potential for new datasets and code to be retained and shared, encouraging replicability and extensions to the study.

“this policy also guarantees you know, to produce results from data and this also incentivises the data generators for sharing the data. People are encouraged, institutions are encouraged to share because sharing is not only giving the data. The ones who give the data also get data from other sources, institutions and individuals” (I3)

The SADS PIR report also stresses the importance of being able to demonstrate the value of how data collected at a local level can be brought together to form important national datasets. An example provided was the dataset created by the National Agriculture Research System Council on fertiliser application to soil. This dataset was made available to test and validate current advice and guidance at both the local and national level.

A further illustration of the power of this approach was reported in the interviews when the workshops were used to present the results of using larger, matched datasets.

“we didn’t do a workshop just to say ‘OK let’s share data’. We collected the data from secondary sources, from published and we did some analysis... So that the researcher also tried to demonstrate the benefits of having larger data sets for analysis.” (I4)

This approach, developing use cases to promote the value of sharing data and the types of safe outputs that can be generated from such an approach, was further explored with the project team. Outputs from the shared data source generated 15 working papers, targeted at high quality journals.

“we had a couple of use cases where we are trying to demonstrate the benefit of research ... one use case was on the rehabilitation of acid soil... another use case was the use of georeferenced data for meter analysis... this so far has led to some, I think 15 draft papers which we hope that can be published.” (I5)

Publishing journal articles is a value in itself to academic analysts, but the wider value to society can be demonstrated through documenting the transition mechanism following output to impacts. For example, over time, citations of the work can be monitored and impacts traced. Alternatively, it may be possible to trace the impact of the data/research as it was cited in the evidence base to inform policy. More directly, however, impact studies may be able to trace and detect when output studies have informed and changed working practices, e.g. farming methods. This type of value chain approach to estimating returns, however, requires detailed mapping and may only have an impact years after publication.

The interviews revealed that the policy ‘stick’ would be there to enforce compliance from those yet to be convinced by the ‘carrot’ approach.

“the policy is ... mandatory which will be followed by every ministry who generate soil and agronomic data related data sets.” (I1)

However, it was suggested that the incentive works best when the carrot and the stick combine and work in tandem.

“from this year researchers will follow the harmonised protocols. Then if we produce data following these harmonised protocols and then the data sharing policy is there, then we will have standardised better quality data, then we will establish the governance and management issues. And then I feel very positive actually of outcomes to come out of the implementation of this policy” (I3)

Given the sensitivities of working with different types of data, it is important for clear data standard controls to be set and enforced. Training should then be provided on awareness of the issues and in relation to checking outputs for disclosure risk (when the underlying data is confidential, statistical outputs can inadvertently reveal confidential data; basic training can make this risk negligible; for example see Eurostat, 2016). Training in disclosure checking was not practised, but there was support for training in relation to outputs in terms of maximising the value of statistical analysis.

“this process was supported through a series of workshops and by hiring data scientists who could help to run more sophisticated analysis, so encouraging them to share and analyse data and write papers and then get publications out of it. ” (15)

However, rather than following a formal programme of training, an element of the learning was through ‘on the job’ training.

“It was basically learning by doing. That we used the collected, the existing data to train Ethiopian researchers on these quotas and at the same time we used some 12 or 15 potential papers that came out of this” (15)

In terms of developing the project, much work has been put into developing a national framework, establishing the databases and enabling the CoW to access the data to produce outputs and demonstrate the value of improving data governance structures and sharing data. There is evidence of investment in developing protocols and training stakeholders and researchers. However, as the project matures, further investment should be considered to develop robust procedures for ensuring the safety of outputs, supported by a training programme for data managers and researchers. By doing so, this should engender further trust in the process and help advance the process.

Lessons learned on producing statistical outputs and shareable datasets

The main lesson learned from this domain is that the safe outputs generated from the investment in data governance provide tangible direct evidence of the value created. In terms of quantitatively reporting value, the number of research outputs produced and published provides a specific and measurable metric of the value created. It is also possible to further track the additional value created by monitoring impacts of the research in terms of further citations and changes to policy and working practices.

To support the delivery of safe outputs, further engender trust in the process when using restricted data to generate output, and increase the overall value of the project, additional investment should be considered in building skills and capacity to monitor and support the publication of safe outputs.

Safe Data

The formation and support of the CoW was a key success factor of the project. However, it was recognised that they were experts in soil, rather than experts in data science, and therefore it was important to bring in additional expertise to help with issues in relation to safe data.

“protocols have come from the coalition of the willing, they have self-organised to develop them. I mean they are the experts in Ethiopia in soil. So they are coming at it from much more the soil side rather than the data science side” (12)

During the inaugural workshop of the CoW, a Data Sharing Task Force (DSTF) was formed with the primary objective of developing ‘informal’ guidelines for data access and sharing, as well as to promote partnership and communication among the CoW members (Progress Report, August 2018).

“it is critically important that different stakeholders also agree on how the data is being generated so that some experimental protocols and procedures are being standardised and data recording procedures are being standardised and that is currently going on.” (15)

It was hoped that the task force could add value to the project by advising on issues around data quality, accessibility and governance procedures. The group were an important contributor to the development of a national soil and agronomy data sharing policy.

In terms of limiting value, the lack of consistency in data standards was considered to be a barrier to sharing data.

“The data are not standardised...That is one problem. In order to organise that data for scientific purposes” (11)

Others also commented on this while highlighting a lack of trust in institutions, people and systems.

“More consistency would allow more interoperability and more sharing would allow other to extract more value. And it’s getting the researcher to trust that other people won’t just gazump them, manipulate them or try and take their credit but actually [see that] there is benefit in sharing data. (12)

The importance of developing harmonised standards and protocols was often discussed during the interviews. It was clear that some progress on this had been made; an example was cited where core treatments had been standardised across a number of individual experiments.

“In the past every institution, every donor had its own protocol for running these experiments; and now through a series of meetings and consultations (also bringing in expertise from the international fertiliser development consortium, I think) it was possible to agree at least on a number of core treatments which will be similar in all trials. Which again then will make it possible analyse the data across the stakeholders, across geographies and across different projects.” (15)

However, there was recognition within the project team that as well as empowering stakeholders to take ownership and drive forward with this type of work, there was also a critical role for the project team to contribute their expertise in relation to these issues.

“we are... supporting in-country working... to put in place some of the documentation and process around it. So there is not the consistent documentation yet around data licensing, data sharing agreements, the suchlike. That’s part of what we do” (12)

When discussing issues in relation to collecting, managing and accessing confidential data, it was recognised that there are protections built into the policy and already in legislation, but that there was more to do on this front.

“Confidentiality. The policy that’s been written applies to all data that relates to soil and agronomy whether it’s held by researchers, government or anybody else... I think that we will need to think about that more... The data is on, you know, 50,000 farmers with their telephone numbers in a database somewhere... But there are existing data protection rules in Ethiopia, not necessarily a specific bit of legislation, but there are rules” (I2)

Discussions in relation to issues of confidentiality were mixed; at the highest level these issues had been considered.

“The confidentiality and all those things are addressed in the policy. What kind of data can be shared, what kind of data cannot be shared because of sensitivity” (I3)

Indeed, even at the operational level, it seemed that the principle for overseeing access to confidential data had been determined.

“Therefore the ministry itself generates data and determines whether this data is confidential data or not. And if we take data from third party or other organizations who generate data, they determine if this data is confidential data. Because there are data owner rights and data stewards.” (I1)

With work on implementation also being well advanced.

“We have submitted both the policy as well as the implementation guide. They are already submitted and awaiting approval. In the implementation we have also mentioned embargo period. So when and how the data is shared to end users or immediate users” (I1)

From other conversations, it was suggested that issues around privacy and confidentiality were important, but developing detailed guidance in relation to process and implementation had been less developed, as they were considered second-order issues. The initial focus had been on developing and implementing the policy and focussing on standardisation of datasets in order that they could be shared. The view from the project team was that these issues were important in generating value, but that there was a process to run through and issues would be dealt with in a sequential order.

“I think that is absolutely an important point that it’s in the work that you are doing looking at the value... [but] it’s not a ‘here’s all the things you need to think about’ at one time... It’s actually a kind of bespoke process depending on the circumstance that we are working through” (I2)

However, with the increasing pace of developments within Ethiopia, brought about by internationally funded projects such as Farmstack by the Gates Foundation, then it was clear that these issues would be promptly moving up the priority order.

“I think this issue might be coming faster than we think... ADA... have established a farmer database with tens of thousands of farmers. The new investment from Gates, Farmstack... I think they are also planning to build a farmer database... I think that’s where the real issue comes in... I am not quite sure yet how this data could be managed or shared or utilised. I think there will a number of issues coming up... where we might need help to solve those questions”(I5)

As well as the legal and technical issues in relation to sharing safe data, potentially the more significant issues have been around the political and cultural landscape of Ethiopia and its regional dimension.

“The personal data confidentiality probably a minor rule but what plays a much bigger role are political issues... Ethiopia, it is a federal state [with] border issues... which piece of land belongs to which ethnic group... regional state ... [all] very, very sensitive... That is one of the reasons why the government is so restricted in sharing data... and that’s why many people are so afraid of sharing the data also because this can easily be misused” (I5)

This need for cultural change was recognised by the project partners and indeed it was suggested that as a result of the project and its wide engagement with stakeholders, there had been some success in this area.

“We want to have a national consensus and we work as a team to push this further so we can transform the culture or whatever. So that is a bit different than I think than we initially were” (I4)

In the SADS PIR, the authors commented that a potential barrier to the implementation of the policy was the lack of clarity when it comes to understanding key terms. As much of the data collected in Ethiopia is as a result of a research project, it is seen as research data and as such is viewed differently by some, in terms of sharing, from national datasets. Therefore, the authors argue that it is important to develop a clear language and clear definitions for different types of data.

It was suggested that using internationally accepted frameworks, such as FAIR (findable, accessible, interoperable, re-usable data), can help with issues of clarity through developing a shared understanding which can be applied in other contexts.

“while developing this policy, we have learned much from CABI because we have little knowledge about FAIR principles at the beginning... the process helped us to think critically how can we make data findable... accessible, interpretable and reusable... it helped us to think critically and also developed an approach to learn how to prepare a policy for other general purposes... I think we learned much from this process” (I1)

To compound this challenge of shared understanding, the authors also point to the lack of clarity over those datasets that should be funded, collected and maintained nationally, compared to those that are only required and collected locally. They argue that being able to clearly identify datasets of national interest would be helpful. This distinction between national and regional datasets, and how they can be brought together, was also commented on in the interviews.

“The biggest problem... is that that data can’t be brought together to make national or regional or local decisions based on that data because the data is not interoperable. So that’s the pain point.” (I2)

In order to pool local data and enable it for use and re-use, it is important that data collected at both the local and national level followed common standardised procedures and policies. For example, one issue raised by stakeholders was the importance of local soil researchers understanding the importance of accurate geo-referencing of their data in order to enhance its usefulness to others. In order to address such issues around improving the quality of data, it was suggested that simple, short packages of learning could be developed to inform those collecting the data.

The iterative nature of projects such as these recognises that not all issues, even when known about in advance, can be planned and addressed from the start.

“It’s kind of like steps. The first step was to get the policy in place. The next step was to get an implementation manual in place to make it clear what you need to do to implement the policy. And then the next step after the policy and the implementation manual are ratified and signed off, then get all the guidelines lined up and brought together. But it’s almost you have to do it sequentially just to get the momentum” (12)

This process is further complicated when working with a diverse stakeholder group who will be generating ideas not initially considered at the inception of the project.

“We have data updated, and new things came, new ideas from the coalition meetings. Even so we know how we have to go in the beginning but new things have arrived.” (14)

The timing of the implementation is also important given the complexities and interlinked nature of the process.

“Both the policy and the system go hand in hand, because the system is required to successfully implement the policy. The policy also helps us, national soil information system implementation. So one is a component of the other”

It was suggested that in order to make progress, it is important to work on issues as and when they become apparent and with the resources that are available.

“The underlying principle really is to try to do what is possible with those people who are interested in the topic rather than trying to resolve the whole issue at once at national level” (15)

As such, in the SADS PIR, it was recognised that there are still many issues that are yet to be resolved, such as compatibility, metadata and understanding what data is where. Having identified these potential barriers to maximising the value of the initial investment, a compelling case for additional funding could be made.

Lessons learned in respect of investments in data quality and usability

There are a number of key lessons to maximising the value of data through improved data governance.

First, knowledge of the local political and cultural context within which data is generated is extremely important; working through trusted organizations and individuals who have experience of working in the locality can therefore improve the realised value of the data.

Second, much of the value is generated through empowering local stakeholders to take ownership and develop systems and methods that work in a local context. However, specialist statistical/data science skills are required. To maximise the value of data, bringing in external expertise can provide the scaffolding to support the infrastructure for local stakeholders.

Third, standardisation (or, at least, inter-operability) of data collection and processing is particularly important. Given the broad range of stakeholders involved with this process, attention should be given to making the guidance and processes as clear and simple as possible in order to avoid confusion and misinterpretation.

Fourth, projects should be flexible, take a step by step process and adapt iteratively, allowing learning by doing.

Finally, maximising value requires planning for second-order activities, such as putting robust policies and procedures in place for safe managing confidential and restrictive data. These help provide a framework within which consistent data design decisions can be made.

Quantitative Results: Exploratory Estimates of the Value of Data Governance

Section 4 introduced a cost-benefit framework, focussing on input, activities, outputs, intermediate outcomes and impacts, to assess the added-value and cost-effectiveness of data sharing. In this section, we use this framework to illustrate how assessments may be made, generating some exploratory estimates as part of the illustration.

Input costs: data gathered from financial reports

The costs presented below provide a high level overview of the inputs for (1) the CoW and (2) activities undertaken to support the establishment of a national IT host for a centralised data repository, both important factors in the full evaluation. It can only provide an imprecise measure of the activities assessed under the benefit frameworks. For example, the level of inputs may be undervalued as other activity evaluated within the benefits framework may not be accounted for in the full costs presented. Conversely, the input may overestimate the full cost as it may include elements that are not evaluated within the benefits framework (e.g. investment in facilities with a productive life beyond the project).

In the absence of any further detailed data being made available to the research team, total costs have been used to give an indicative measure from which to compare the estimated benefits. Given the uncertainty with the coverage of the data, however, these the costs should be seen as indicative rather than relied upon as any accurate measure of the input costs.

Table 5: High level breakdown of input costs

GIZ-ISFM+ BMGF KOFI		
Financial Report		
Summary of Expense for CoW and SIRM activities to Improve Data Framework in Ethiopia (Excluding Personnel and indirect cost)		
Reporting Period: November 2018 to July 2020		
Project NO: 14.0156.1-307.00		
Expense Category	Amount USD	Amount Euro
Capacity Building – Training**	272,253.09	229,072.86
Capacity Building – Facilities (IT facility for ESRI/SIRM)	157,115.14	132,196.16
Consultants	145,839.49	122,708.87
Workshops	47,648.28	40,091.11
Travel	13,186.81	11,095.34
Grand Total	636,042.82	535,164.34
**In collaboration with CIAT		

Source: GIZ

Table 5 reports that the input costs of activities in relation to improving data governance and data access in the SSHiE project was approximately US\$640,000. In order to gain an understanding of the added-value generated by improved data governance and access, the value of benefits generated as a result of this investment has been estimated.

Benefits: data gathered from questionnaire

To estimate benefits, and hence value-added, from improved data governance and access, the UWE team developed a questionnaire to capture expert information. The CABI team identified 12 potential experts who had particular knowledge of the relevant work streams and would therefore be best placed to answer questions in relation to each activity. Given that there were nine independent work streams and most respondents would only have knowledge of certain activities, individual questionnaires were tailored for each respondent to minimise the size of the questionnaire and encourage the highest response rate. Of the 12 questionnaires sent out, eight were returned.

The strength of approaching experts is that those individuals have in-depth knowledge of the subject matter and so are best placed to provide quantifiable estimates as required (e.g. number of attendees, number of publications). However, a number questions rely upon subjective judgement (e.g. potential time saving, likely impact). Given that all the expert respondents had a vested interest in the success of the project, a weakness of the methodology is there may be an increased risk of optimism bias in their answers and therefore any benefits may be somewhat inflated.

The data on benefits relate to activities completed for three of the six work packages for the SSHiiE project. In line with the quantitative framework, these activities were chosen as they were assessed by the research team as being central to improving data governance and access within the project. The activities evaluated for the three work packages were as follows:

- Work package 0 - Inception phase:
 - Activity 1 - Inception Workshop
 - Activity 2 - Sub-grant agreement
- Work package 1 - Craft policy to share data to ease Ethiopian soil system bottlenecks:
 - Activity 1 - Open Data Sensitization Workshop
 - Activity 2 - Coalition of the Willing
 - Activity 3 - Workshop on Data Access and Sharing
 - Activity 4 - Data Sharing Task Force
- Work package 2 - Provide support for inter-institutional consultation needed to establish national IT host for a centralized data repository
 - Activity 1 - Organizational Structure Stakeholder Consultation Workshop
 - Activity 2 - Human Capacity and Development
 - Activity 3 - International Exposure Visits

The following results are presented in four separate sections:

- Internal value directly created (output)
- Internal value indirectly created (output)
- External tangible value (intermediate outcome)
- Wider economic and social outcomes (impact)

Output: Direct Internal Value (DIV)

In order to calculate the benefit a number of assumptions were made when collecting and processing the data. For example when converting value between US\$ and Ethiopian Birr an exchange rate of 1:39 was used (as recorded in xe.com at 14th Jan 2021). For a full list of the assumptions made, please refer to Appendix 3.

Estimates of direct internal value (DIV) are calculated by multiplying the number of Activities (A) undertaken (e.g. workshops) by the number of people from each sector (broken down by six sectors) who have engaged with the event (Ps), hourly wage rate for the sector (HRs) and length of time in engagement, measured in hours (H). The assumption that underlies this value is that attendance is voluntary, and so attendance implies that the expected value from the event is at least as high as the cost of attending (measured in this case by paid-for hours). This may be an overestimate as (1) individuals may be under pressure to attend, (2) individuals may not consider the cost to their employer of attendance, and (3) the value gained may be systematically lower than expected. Alternatively, it could be an underestimate as it only seeks to establish a minimum value of attendance.

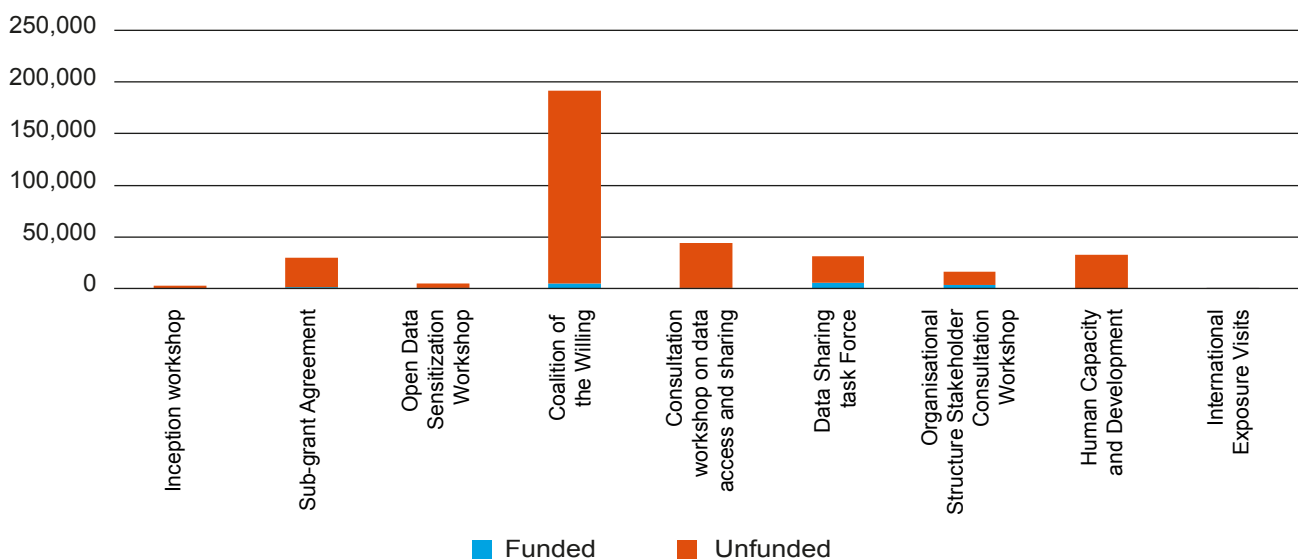
Equation 1: Direct Internal Value

$$DIV = \sum_{s=1}^6 PS_s \times HR_s \times A \times H$$

Following the estimation of the total DIV created, it is necessary to identify the *added value* over and above that which had been directly funded by the project. The overall figure was therefore allocated between those on the project payroll, and those external to the project.

Of the nine activities assessed within this framework, it is estimated that they generated over US\$353,000 DIV; 95% of this was added value to the project, as it accrued to individuals external to the project. Figure 4 shows the breakdown of the DIV across the nine activities.

Figure 4: Direct Internal Value by Activity



Source: Authors' calculations

Over 54% of the overall benefit – equating to US\$191,000, or over 13.1 million Ethiopian Birr - was generated by the CoW. The aim of this activity was to establish a group (made up mostly of scientists, academics and researchers) whose members would benefit from sharing and accessing data among themselves. It was also anticipated that it would help towards establishing a national soil and agronomy database and data sharing policy.

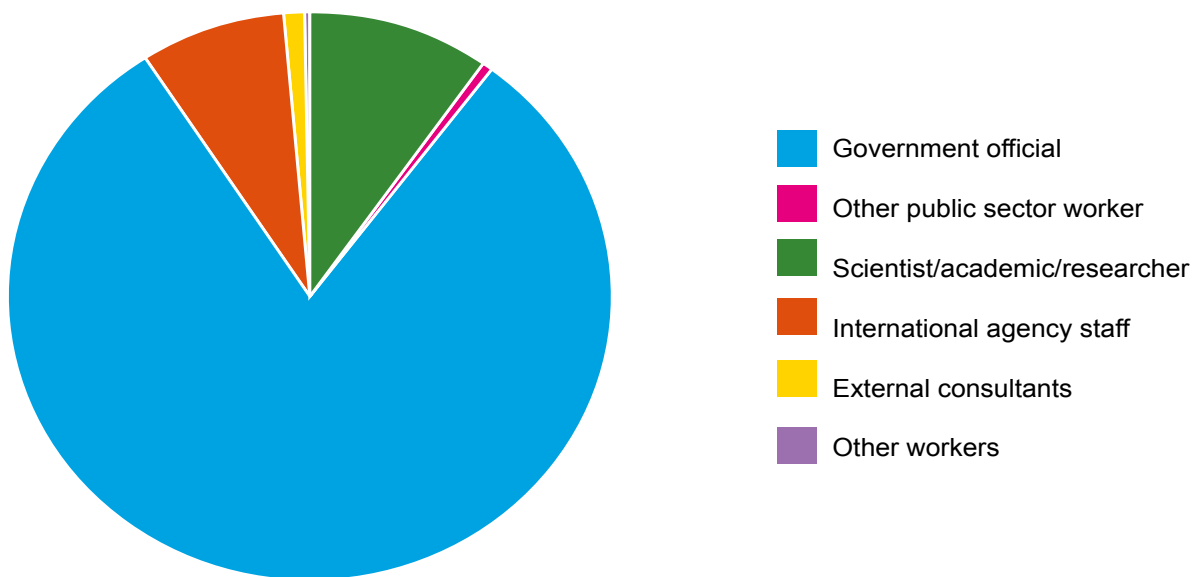
The consultation workshop on data access and sharing built on the value created by the CoW, and was estimated to have generated an additional US\$44,000 of DIV. Other strong contributors were activities in relation to Human Capacity and Development (9.1%), Data Sharing Taskforce (8.8%) and Sub-grant Agreement activities (8.5%).

The inception workshop, open data sensitization workshop and the international exposure visits are jointly estimated to have generated just 2.4% of the direct internal value. However, it is somewhat misleading to present value for each activity separately. This is because much of the value is generated through the package of interventions working together. A good example of this was the inception workshop, which was used as the launchpad for the project and was the linchpin from which all other activity derived. A further example is the CoW which

produced the greatest direct internal value but was only established as a result of the open data sensitization workshop. While ‘international exposure’ similarly appeared to generate limited DIV, the aim of these activities was to build institutional capacity and forge long-term partnerships with similar international organizations. Therefore the benefit of such visits is likely to occur in the longer-term and their real value is unlikely to be captured using the DIV methodology.

Figure 5 shows the breakdown of DIV by sectors. Unsurprisingly, given their central role in the CoW and the fact that scientists, academics and researchers are generally among the first professions who are able to derive additional value from improvements in data governance and access, this group produced over 80% of the overall DIV generated.

Figure 5: Breakdown of direct internal value by sector



Source: Authors’ calculations

The two other sectors who have contributed most to the DIV are government officials and international agency staff. Given the three main aims of the project (increase reach of government and private-sector services; improve resource allocation and policies; and increase the quality, availability, and utility of data and evidence), it is clear that these sectors have a crucial role in delivering a successful outcome. Therefore it is not surprising that they are the most significant contributors to delivering DIV.

Output: Indirect Internal Value (IIV)

IIV was estimated through asking experts to judge how an activity impacted on the delivery of subsequent activities within the project. This concept was one of the most challenging sections of the survey for respondents as they were asked to consider the ‘counterfactual’ – e.g. what would have happened to the following project activities and work packages in the absence of this activity.

In particular, respondents were asked to consider the effect on the time to complete future activities had the current activity not taken place. Respondents were then asked to attribute this increase or decrease to one of seven potential mechanisms. Six mechanisms were identified as potentially being of importance for generating value in the qualitative research. The seventh ‘other’ category would pick up any residual effect.

By asking the respondents to identify both potential increases and decreases in time separately, the research team were able to individually identify which activities were thought to have saved or increased the time necessary to complete subsequent work packages. It provided the possibility to calculate a net time saving for each activity. Although this is extremely difficult to measure accurately, it is conceptually the most straightforward measure in a cost-benefit analysis as it directly answers the question “what impact occurred which would not otherwise have happened?”

IIV is calculated by multiplying Net Time Saving (NTS) (which is created by summing a weighted average of the time saving (TS_w) minus a weighted average for the extra time needed (TN_w)) and multiplying it by average hourly wage (AHW). All weightings are apportioned by the level of confidence recorded by each respondent in the accuracy of their estimate. Average hourly wage is created by averaging the total of all six types of wages. The full steps are documented in the following equations.

Equation 2: Indirect Internal Value

$$NTS = TS_w - TN_w$$

$$AHW = \sum_{s=1}^6 HW_s \div 6$$

$$IIV = NTS \times AHW$$

Due to a lack of data availability it was not possible to generate IIV estimates for International Exposure Visits; however estimates for all other activities are presented in Table 6.

Table 6: Indirect Internal Value

	Time Saving (Days)	Extra Time Needed (Days)	Net Time Saving (Days)	Value of time saved - US \$
Inception workshop	141	-	141	9,610
Sub-grant Agreement	65	-	65	4,453
Open Data Sensitization Workshop	60	-	60	4,082
Coalition of the Willing	142	10	132	8,980
Consultation workshop on data access and sharing	105	-	105	7,175
Data Sharing task Force	84	-	84	5,745
Organizational Structure Stakeholder Consultation Workshop	60	-	60	4,082
Human Capacity and Development	60	-	60	4,082
Total	719	10	709	48,209

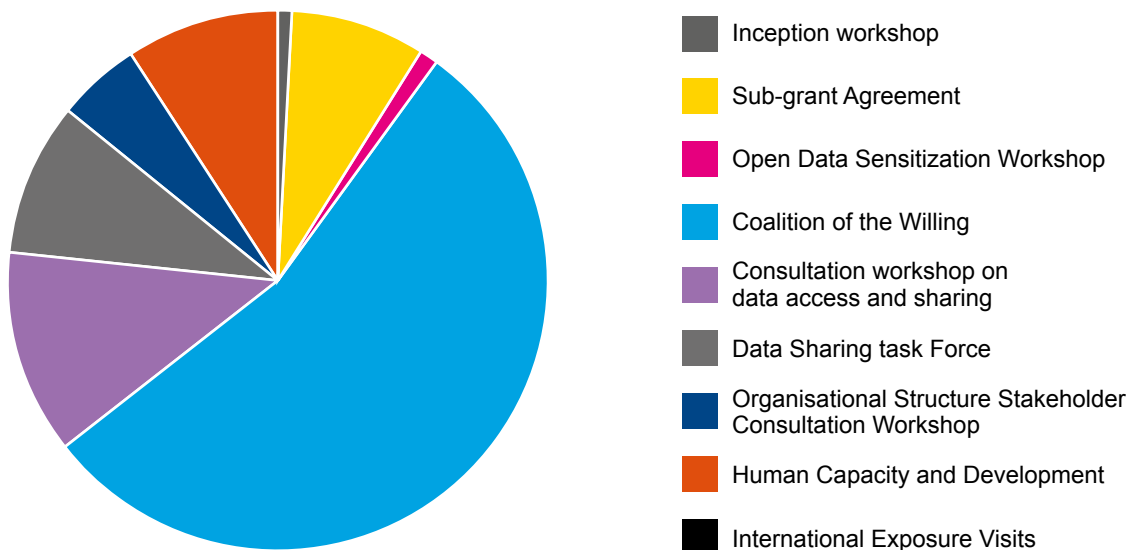
Source: Authors' calculations

The estimates reveal that every category contributed to an indirect time saving, which in total was equivalent to one person working for 719 days or approximately three years. The single largest net contributor was the inception workshop; this demonstrates the importance of this event in engaging the local community to support the project to deliver subsequent activities and achieve its overall objectives.

The CoW was the second largest contributor. This was the only activity which led to subsequent activities requiring additional time to complete (i.e. additional costs), estimated to be equivalent to 10 days. This made relatively little difference to the overall net effect however; in total the CoW was estimated to have saved the equivalent of 132 days' work for subsequent activities.

Unlike the DIV, which was dominated by the CoW and with three activities barely contributing, in terms of generating IIV contributions were much more evenly distributed across all activities. This is demonstrated by Figure 6 which plots the proportion each category made to the overall contribution.

Figure 6: proportion of IIV by activity

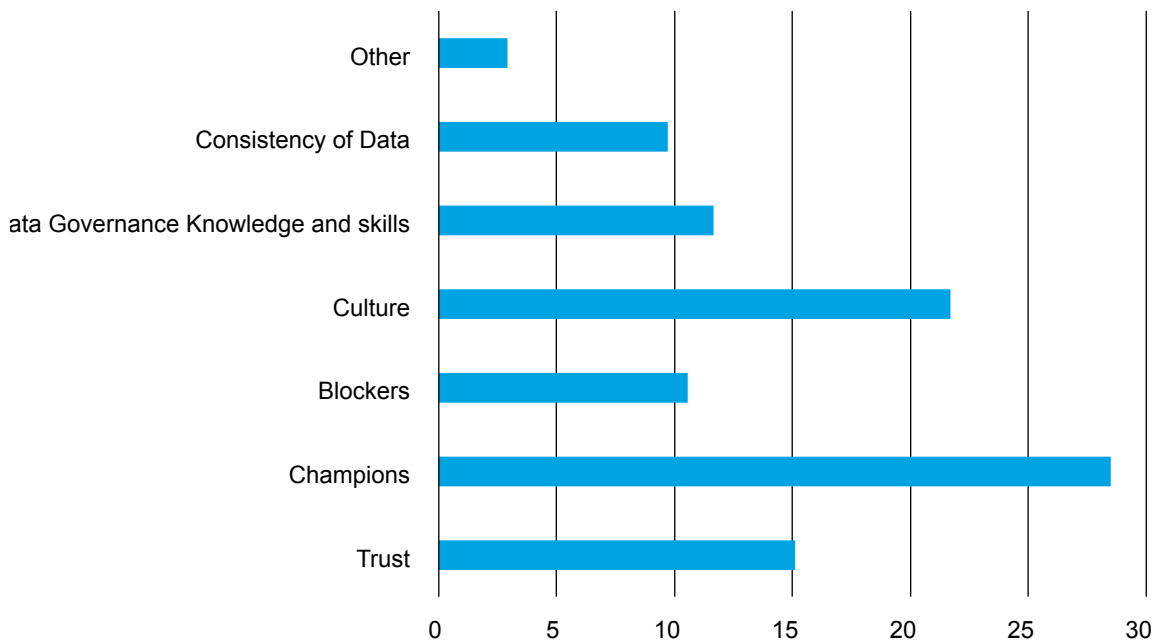


Source: Authors' calculations

The top three categories each contributed between 15-20% of the total time saving; the bottom five each contributed between 9-12%.

Having provided an estimate of the additional time saved and extra time needed as a result of each activity, the experts were then asked to apportion their estimates to the mechanism through which it was most likely to occur. Figure 7 provides a weighted average of the proportion of total time saved (719 hours) attributed to the different mechanisms identified as potentially responsible for driving this saving.

Figure 7: Weighted average of the proportion of time saved allocated to different mechanisms



Source: Authors' calculations

In support of the findings from the quantitative interviews, working with 'Champions' was judged to be important; in the quantitative estimates it accounted for 28% of the time saving and hence 28% of the IIV. This was followed by a change in the culture in relation to data governance and access and an increase in trust between the parties involved in the process. These were estimated to account for 22% and 15% of the IVV respectively. What is of particular interest about these two indicators is that any positive change in culture and trust would be expected to extend past the length of the project and therefore there is potential for much greater longer-term impacts. This illustrates one of the difficulties of assessment where structural change is concerned: how does one draw limits around what is to be assessed as the benefit of a specific project?

Improvements in data governance knowledge and skills, working to address concerns identified by potential 'blockers', and improvements in the consistency of data were jointly estimated to account for approximately one third of the IVV. The remaining 3% was attributed to the residual 'other' category. To understand exactly how the mechanisms led to time savings, respondents were encouraged to further elaborate on this, but no further information was provided.

Intermediate outcomes: External Tangible Value

In any evaluation, the costs and benefits estimated are determined by the parameters by which the evaluation is set. Theoretically, there may be reasons why an evaluation may wish to measure outcomes along the data value chain. For example, as this project's overall aim was to lead to changes in soil and agriculture development in Ethiopia, it is not unrealistic to expect the outcome of changes made to data governance and access would lead to downstream changes, such as modifications to farmer behaviour.

However, given (1) the need for the evaluation to be proportionate and (2) the limited availability of data, the parameters for this evaluation of the (tangible) outcomes were limited to just measuring publications produced as a result of these activities. The strength of this approach is that the output is tangible and measurable, and can be directly attributed to the activity. The main weakness of this approach is that it is narrow in focus and does not capture the full outcomes, and therefore will likely undervalue the full value (on the reasonable assumption that outcomes are generally positive).

Although it is theoretically plausible to estimate a monetary value of the publications, in practice generating meaningful estimates proved infeasible. There is no information on inputs (i.e. time taken to develop the publications), or outputs (e.g. quality and use of publication). To aid any future evaluations based on publications, it would be helpful to collect information on input costs, collect data on the type and quality of outputs (e.g. citations, journal rankings) and map publications through to use. Although conceptually more accurate, these output-based indicators may only be available only some significant time after the completion of an evaluation and therefore it may be more appropriate to adopt an input based methodology.

Table 7 lists the publications produced as identified by the experts. Twenty-one separate publications were generated as a result of three activities, but all but two were attributable to both the sub-grant agreement and CoW. The links between the individual activities make it difficult to accurately allocate the publications to one or other activity.

Table 7: Publications Produced

	Guidelines	Policy Report	Mapping report	Journal article	Not specified	Total
Sub-grant agreement/ Coalition of the willing	2	1	1	15		19
Data Sharing task Force			1		1	2

Source: Authors' calculations

Wider economic and social outputs (impact)

During the qualitative interviews the respondents indicated that the overall impact of improved data governance and access in the SSHiE project was likely to generate broad-based and wide-ranging impacts, and that these could extend beyond the soil and agronomy sector. However, given the challenges of estimating a monetary value for wide-ranging impacts resulting from an activity which is difficult to define, measure, and attribute causality, the research team opted to use a more flexible framework to estimate the potential effect. As a result the experts were all asked to estimate the potential likelihood and impact based on scenarios generated from the qualitative interviews.

In the following analysis, the impacts recorded are generated at the level of the project as a whole. This is a departure from the methodology used to produce the results reported in the previous sub-sections, which focussed just on the activities aimed at direct improvements in data governance and access activities. However, given that a core aim of the SSHiE project was to improve data governance systems and widen access to data for the soil and agronomy sector, combined with the difficulty in bracketing the specific data governance and access activities when considering the wider impact, the respondents were asked to consider the following potential impacts at the level of the project as a whole.

The likelihood/impact matrix methodology adopted required that the respondents score eight statements against the following five-point Likert scale – the full questionnaire is listed in the final section of Appendix 2.

5 Point Likelihood Score		5 Point Impact Score	
1	Not likely to occur	1	No impact
2	Possible but in rare circumstances	2	Minor impact on a limited region or sub-group of people
3	Probable to occur	3	Major impact on a limited region or sub-group of people
4	Very likely to occur	4	Minor impact nationally
5	Almost certain to occur	5	Major impact nationally

The first five statements explore potential impacts of the key mechanisms discussed previously in this research. They do this through examining the potential impact directly on the soil and agronomy sector, and indirectly on agriculture and other sectors through spillover effects. Respondents were asked to judge both the likelihood and potential impacts of the following ‘internally focused’ scenarios:

The SSHiE project will either directly, or indirectly through spill-over effects, lead to...

- improved level of trust between stakeholders
- a change of culture towards data sharing and move to a more open data framework
- improved levels of human capital in data governance and data access
- improved infrastructure and technology
- improved data sharing and access policies

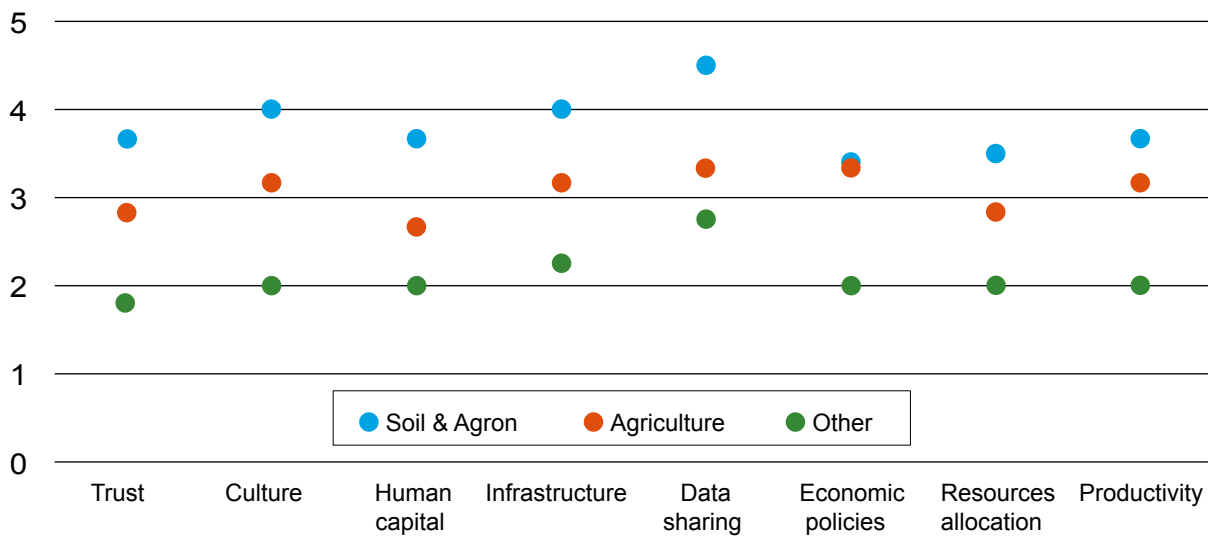
The final three statements have a broader focus by asking respondents to consider the likely impact of the SSHIE projects on the wider economy. They do this for three sectors (agronomy, agriculture, other). The questionnaire includes the following ‘externally focused’ statements:

The SSHiE project will either directly, or indirectly through spill-over effects, lead to...

- improved economic policies
- improved regional and national resource allocation
- improved productivity

Figure 8 shows the likelihood of the impact happening for each of the statements, while Figure 9 displays the potential impact such an outcome would produce. All the figures are calculated as the average response to each statement. The information is presented separately for the soil and agronomy sector, the agriculture sector and other sectors. Because each data point is based on only eight responses, confidence intervals (not shown) are wide, and differences may not be statistically significant. However, as a purpose of this research paper is to illustrate how such estimates may be calculated and used, we assume below that the differences are real.

Figure 8: Average likelihood of impact by theme and sector



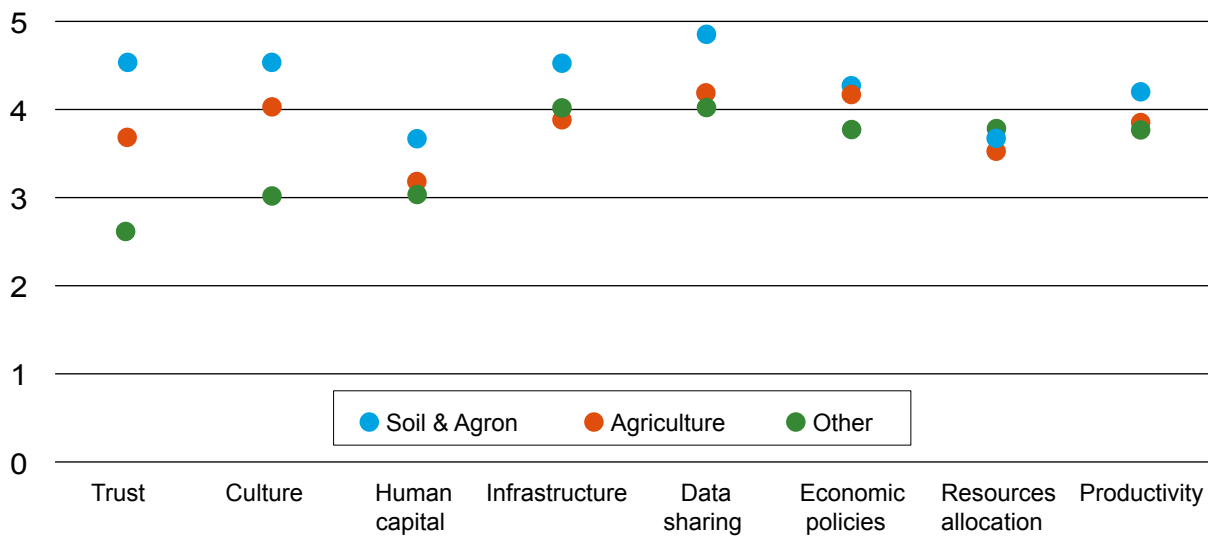
Source: Authors' calculations

Given the relatively low number of respondents, it is important not to over-interpret. With this caveat in mind, however, Figure 8 shows that with just one exception - expected *a priori* - the likelihood of impact is greatest for the soil and agronomy sector, then the agricultural sector and then 'other' sectors. The one exception is that both the soil and agronomy sector and the agriculture sector are estimated to have the same likelihood of improved economic policies. This may partially be explained due to the fact that economic policies are likely to be targeted at a higher level and as such economic policies focussed on the agricultural sector will also cover the soil and agronomy sub-sector.

In the soil and agronomy sectors all eight scenarios are estimated as at least 'probable to occur'; improvements in data sharing and access policies, culture and infrastructure are all 'very likely to occur'. All indicators for the agriculture sector cluster around 'probable to occur'. For 'other' sectors most of the indicators cluster around 'possible to occur but in rare circumstances'. The exception for 'other' sectors is improved 'data sharing and access policies', where the likelihood is estimated to be close to three (2.8) – 'probable to occur'. Indeed for all sectors improvements in data sharing and access policies records the highest rating. This is encouraging as the longer-term economic and social impacts of this outcome are potentially substantial.

For both the soil and agronomy sector and the 'other' sector, the average score of the five internally focussed indicators was higher than for the externally focussed indicators (agronomy 4.0 to 3.5; other 2.2 to 2.0). Given the tighter link between the five internal mechanisms identified as important for generating value in the SSHiE project, this result is as one would expect.

Figure 9: Average scale of impact by theme and sector



Source: Authors' calculations

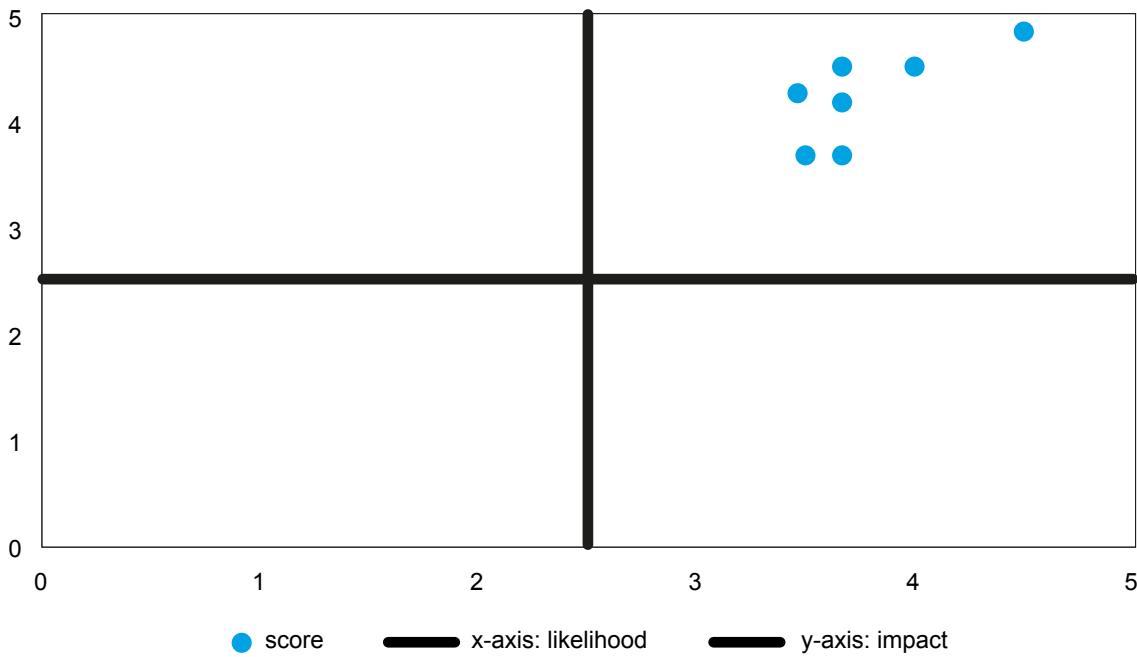
Figure 9 again shows that the greatest impact was identified for the soil and agronomy sector. The most variation between sectors was shown in improvements in 'trust' and 'change of culture'. This may be somewhat explained by the fact the impact of the SSHiE project may be more localised when considering internal mechanisms for change (i.e. trust and culture) which may be more based on personal interactions.

Overall however, the difference in the range between impact (figure 9) and likelihood (figure 8) was considerably reduced between sectors. Indeed when just looking at the three indicators in relation to the economy for impact, all three sectors are extremely close. These results are reassuring as *a priori* we would expect the potential impact to be more equal across the sectors, especially in relation to wider economic impacts.

Figures 10, 11 and 12 plot the likelihood/impact matrix for the three sectors – soil and agronomy, agriculture and other sectors respectively – which provides us with estimates of the potential long-term impacts of the SSHiE project.

When two or more of the eight categories share the same score, this is shown by presenting only one data point, but including both categories in one label joined by an ampersand (&). For example, in Figure 10 'Infrastructure' and 'Culture' share the same score which is indicated by the label "Infra & Culture". The result of this is that only seven data points are listed to describe the full eight categories.

Figure 10: Potential longer-term outcomes, soil and agronomy sector



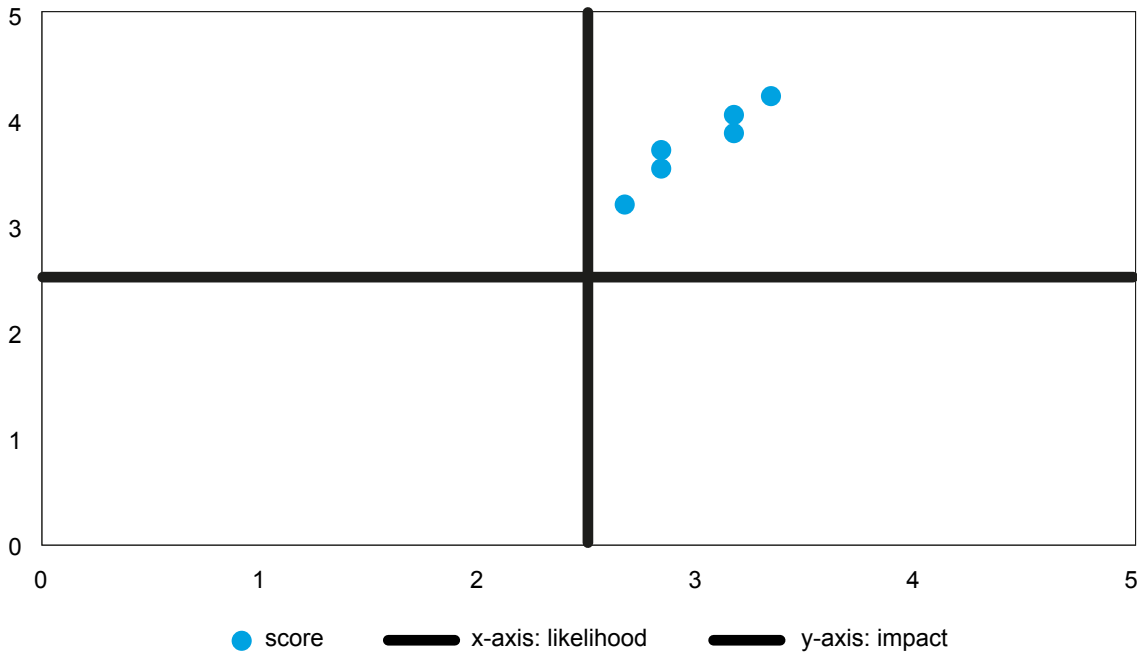
Source: Authors' calculations

For the soil and agronomy sector, the matrix reveals that for all indicators the likelihood/impact score is located in the top right hand quadrant. The average score (when multiplying likelihood by impact) for all eight scenarios is 17 out of 25. Given the consistency of results for all scenarios, this would strongly suggest that the SSHiE project is likely to have considerable long-term impacts.

Of all the scenarios, the likelihood and impact of 'improved data sharing and access policies' is greatest – it achieved a likelihood/impact score of 22. The worst performing indicator from the group of five 'internal' mechanisms identified is 'improved levels of human capital' in data governance and data access. This suggest that continued support may be needed in development of these skills after the completion of the project.

The worst performing indicator for the external' focussed statements (and overall) was 'improved regional and national resource allocation' (13.0), although the likelihood of the impact was estimated to be greater than that for improved economic policies. The likelihood/impact scores for the wider external economic scenarios are broadly below the score for the five internal mechanisms identified within the project, but given that they are all securely located in the top-right hand quadrant of the matrix, it does indicate that noticeable effects should be felt at a national level.

Figure 11: Potential longer-term outcomes, agricultural sector



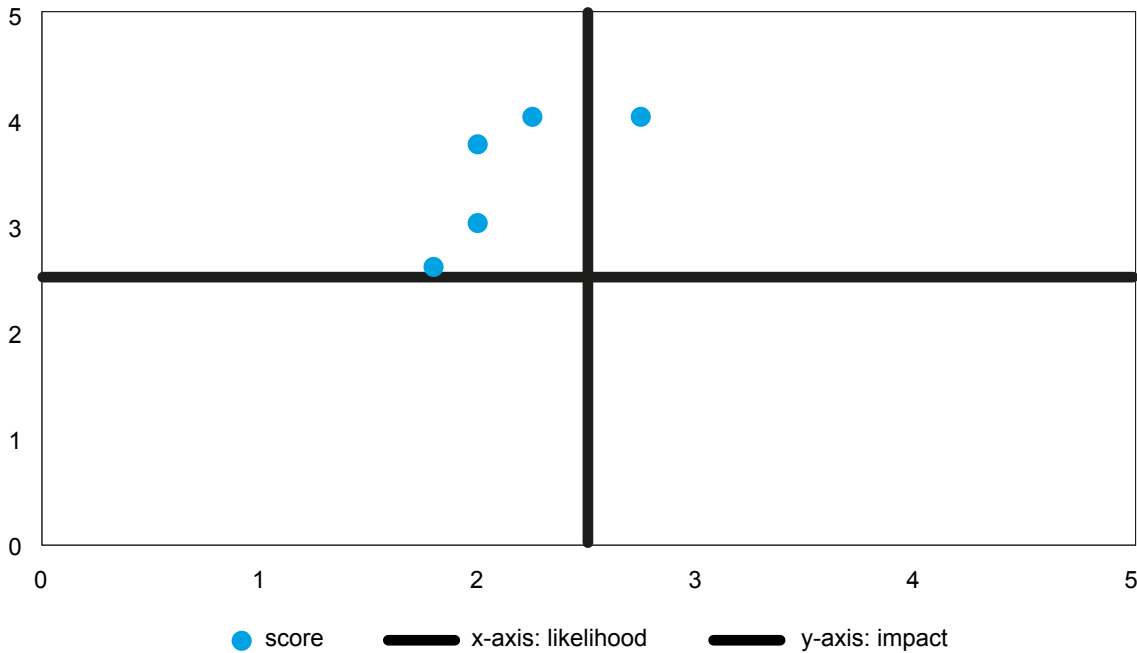
Source: Authors' calculations

For the broader agricultural sector the likelihood/impact score for all scenarios is once again positive and lies within in the top right hand quadrant, but the overall likelihood/impact score of 12 out of a possible 25 is noticeably lower than the soil and agronomy sector. On average the estimates for each indicator are reduced by about 30%. The reduction is more pronounced when comparing the five internal mechanisms (33%); when comparing the three wider economy measures, the difference between the soil and agronomy sector and the agriculture sector is reduced to 20%. This suggests that the wider external effects on the economy are more likely to spillover from the soil and agronomy sector than the internal mechanisms such as improvements in trust and changes in culture. This is understandable as economy-wide changes will primarily be enacted through changes in policy. Given that the soil and agronomy sector forms part of the agricultural sector, combined with the fact that our qualitative analysis revealed that there was government interest in widening the soil and agronomy data sharing policy to cover the full agricultural sector, it is not surprising to see relatively more top-down effects compared to bottom-up effects.

The worst performing scenario again was that in relation to 'improved levels of human capital' in data governance and data access. This adds further weight to the evidence that further post-project support may be needed in this area.

Although the overall score is noticeably lower compared with the soil and agronomy sector, the consistently positive responses do indicate that there are expected to be meaningful long-term impacts of the SSHliE project at the broader level of the agricultural sector. Support for this assessment is reported in the qualitative interviews, which highlighted the opportunity to broaden the development of policy from the soil and agronomy sector to the agricultural sector.

Figure 12: Potential longer-term outcomes, other sectors



Source: Authors' calculations

As expected *a priori*, the results for all other sectors are considerably reduced with only one scenario (improved data sharing and access policies) now situated in the top right hand quadrant. This suggests that, even though the impact (y-axis) still appears to be meaningful for all scenarios, the likelihood of this impact occurring is considerably reduced. With a mean likelihood/impact score of just 7 out of 25, it would seem sensible to conclude that there is only very limited evidence that the SSHiE project will have broader impacts on sectors outside of the agricultural sector.

At the risk of over-interpreting the results, it is however noticeable that the channel of improving data sharing and access policies remains in the top-right quadrant. This would suggest that data sharing policies that are implemented at the level of the soil and agronomy sectors, and which should generate meaningful impacts, potentially could spill-over in to other sectors. If so (and given the support in the academic literature for open data leading to economic growth) there is some support for the counter-argument that the spillover effects of the SSHiE are likely to be felt in other sectors.

Quantitative Results: Summary

In this section we reported some illustrative results on the costs and benefits of data governance and access within the SSHiE project. Given the challenge in defining and accounting for data governance and access elements of the SSHiE projects, combined with the fact that it is extremely challenging to quantify, value and monetise what potentially is an unquantifiable concept, the estimates here should only be considered as illustrative. The intention of producing these estimates has been to test some concepts but mainly to inform discussion as to how data governance and access could be measured in future programmes. What is clear, however, is that for it to be done in any meaningful way, then the evaluation and data collection plan needs to be implemented at the inception of the project.

Given these caveats, Table 8 reveals that the illustrative methodology estimated the following results:

Table 8: Total cost and benefits

Activity	Cost (US\$)	Benefit (US\$)	Net Cost/ benefit (US\$)	Benefit (net time saved – days)	Additional benefit (number of publications)	Long-term impact score: (average score for eight categories – likelihood multiplied by impact)
Input: project expenses	636,042					
Output: internal direct value		353,291				
Output internal indirect value		48,209		709		
Intermediate outcomes					21	
Impacts: soil and agronomy sector						17
Impacts: agricultural sector						12
Impacts: other sectors						7
Total	636,042	401,500	- 234,542	709	21	

Source: Authors' calculations

The results suggest that purely based on an assessment of the monetised costs and benefits of inputs and outputs, investments in data governance and access activities in the SSHIE project would return a net benefit of minus \$234,542 dollars.

This means that in order to break even at the level for intermediate outcomes, the monetised value of the 21 publications would have to equate to approximately \$11,000 dollars per publication. In order to estimate this a further research project would be required. To get a sense of whether this is meaningful or not, it may be useful to consider an input based approach applied to wage rates already used in this evaluation. Assume that the output is valued by the researcher as at least the time spent on writing it; then a researcher in Ethiopia working for approximately 2.5 months full-time on each publication would generate that benefit. This is probably a conservative estimate for publishing in high quality academic journals, and so there is a case that, on an input-costs model alone, the project generates a positive return.

If this were to be the case, any wider benefit, which we have attempted to capture in the long-term impact factor, would then be directly attributed as a surplus of benefits over costs. The evidence supports the suggestions that these wider benefits potentially could be considerable, particularly for the soil and agronomy sector and potentially the wider agricultural sector. The evidence for impacts on other sectors is less convincing.

Ultimately an attempt to put a financial value on these impacts is beyond the scope of this project, and potentially beyond the scope of what is reasonable to do given the difficulties in identification and measurement of such fluid concepts. However, a strength of this work is that the mixed method approach has enabled the research team to identify specific mechanisms through which value is perceived to be created and have subsequently developed a quantitative framework that can be useful in understanding such value. Further work is needed to unpack these mechanisms in order to develop a systematic approach to valuation. As well as estimating the costs and benefits to the project, the flexible framework has enabled the research team to generate estimates of the likelihood and size of the potential impacts. Future research should explore in more detail whether it is possible to provide quantitative values from right across the value chain and in particular methods for mapping, measuring and estimating wider economic impacts.

Reflections on the evaluation process

The initial aim of the project was to produce a quantitative (econometric) study in order to estimate actual and potential return on investment to data governance and access. The results of the study were intended to feed into recommendations for the FAIR data working group, as well as provide valuable insights for external engagement, policy coordination, and recommendations for donor alignment.

Given the difficulty in defining and measuring a potentially unquantifiable concept – the value of data governance and access – combined with the lack of data generated as part of the project, following a series of meetings and workshops between UWE, CABI and ODI it was agreed that the main focus of the research would be to develop a framework which could be used to understand the mechanisms through which value was generated. At the workshop held on the 8th October 2019, it was agreed that UWE researchers should undertake a case study, using mainly qualitative approaches, of a CABI/ODI funded project, in order to explore how improved data governance generates and increases value within projects, and identify practices and procedures which can limit this.

Following a review of potential projects and programmes, it was agreed to undertake a case study of the Cereal Systems Initiative for South Asia (CSISA) Phase III. UWE proposed undertaking a ‘macro’ project level approach to explore the value of data governance and data access. This would then be followed by a ‘micro’ approach, exploring the value of data governance at a sub-project level. The benefit of the two-tiered approach was that it would better enable the research team to understand the mechanisms for how value is created at the different levels of a project. A fuller description of the methodological approach is included in Appendix 4.

However, after the initial round of exploratory interviews, a number of issues were identified. The most important were concerns from interviewees about data governance being evaluated, and the difficulty of identifying individuals who would be able to provide relevant information. Simultaneously, as a result of the literature review it had also become clear that there were no framework templates to carry out this type of economic evaluation. At this point the purpose of the project changed.

The revised ambition was to return to first principles and identify a useful framework for evaluation, and then to test the value of that framework in a genuine case study. For the latter, it was decided that the SSHiE project would make a more tractable case study than the considerably larger and more complex CSISA III project.

The action research methodology of iteratively developing the research project following cycles of action and reflection has proved to be extremely valuable. At the inception of the project the intention was to undertake a traditional economic analysis of the value of data governance in a CABI/ODI funded project. However, what quickly became apparent was that the data needed to undertake any formal evaluation was not available and was extremely challenging to generate ex-post. As a result of the initial cycles of action and reflection, culminating in the first exploratory interviews with members of the CSISA project team, a number of key findings were identified:

- In order to be able to evaluate the value of data governance, data governance should be a specific task of the individual project/programme itself – this project has shown that it is difficult to detangle data governance from other objectives ex post.
- All parties involved with the project should have a shared understanding of the meaning of data governance and how it is to be measured in relation to the project.
- The evaluation of data governance needs to be planned ex-ante so that formal mechanisms to collect the data required for any formal economic analysis are built in to the programme.
- Valuing of data governance is in its infancy and therefore the development of a formal framework should encourage the systematic collection and comparison of the value of data governance across projects and programmes.

As a result of these challenges, a further review was undertaken. After discussions initially between CABI and the Gates Foundation, then subsequently between CABI and UWE, a further revision to the project was agreed.

The primary outcome of this study would be to develop and then assess the applicability of the Five Safes framework with the aim of providing CABI and its partners with a standardised approach for developing and evaluating any future investments in data governance. It was also expected that the project being evaluated would directly benefit through gaining a greater understanding of the value of data governance within the project and how this is created, while also potentially highlighting specific areas where further investments could increase such value. The expectation was that the initial case study would also generate specific insights from the project which will have wider applications in demonstrating the value of investing in improving data sharing practices.

The results of this case study have been reported above; however the following findings relate to the research process itself.

- The Five Safes framework is a useful framework which can be consistently applied in the evaluation and reporting of the value of data governance
- Mapping the Five Safes framework against the activities, outputs and outcomes of each work package is helpful to identify data requirements
- The work package/Five Safes matrix is not intuitive for interviewees to understand and therefore this should not be shared with interviewees; instead specific questions should be drafted and shared ahead of the interview.
- The interview process itself, regardless of any findings reported, acted as a catalyst to stimulate the project team to reflect and consider new approaches. For example, one interviewee commented:

“It’s nice talking to you [the research team] and I think a number of new ideas and things came up, so it’s always good to discuss about these with some outsiders also to get some fresh ideas” (15)
- Given the potential cultural and language barriers in undertaking an international research project, undertaking unstructured or semi-structured interviews can be challenging. As one participant put it:

“It would have been better if you sent your questions in advance. I think we have given you some insight about but it would have been better if you sent us the pitch before” (11)

- In international research projects where language may be an issue, clarity when asking questions is vital. For example, in the earlier interviews, the interviewees were asking multifaceted questions, until one observer noted:
“instant feedback - try to ask one question at a time”
- Developing online videos to support the written guidance was extremely helpful when asking experts to provide complex information which in itself is not easily quantifiable
- It was apparent that there was a mixed understanding of some key, but nebulous, concepts such as ‘value’ and ‘data governance’; further guidance to the interviewees on these key terms would have been helpful.
- The Five Safes framework was generally well understood and was helpful in structuring the collection of data.
- Applying the framework enabled the identification of a number of key transmission mechanisms which were then formally tested using an illustrative quantitative framework.

When exploring issues in relation to the value of nebulous concepts such as data governance and access it is helpful to gain an understanding of how value is generated. However, from a policy and donor community perspective there is also a need for some informative quantification of the value of these activities. Therefore, in order to gain a broader understanding of the value of the transmission mechanisms identified through the qualitative research, a formal quantitative framework was adopted in order to provide an insight of the value.

In respect of the generation of value estimates, we note that:

- It is important to be clear on what is and what is not possible and/or worthwhile to measure and/or estimate quantitatively.
- Working to a standard cost and benefit framework was initially helpful in allowing the research team to focus on identifying what information could be generated for each level of impact across the logic model (i.e. inputs, output, outcomes and impacts).
- However, when it comes to the detail of data collection, frameworks need to respect the nature of the data being sought; the framework developed for this project to evaluate data governance in SSHiE, for example, would be inappropriate to evaluate the soil improvement programme itself.
- Given that very few costs and benefits could be identified with certainty (or even a degree of accuracy), multiple approaches, including qualitative methods, are essential to provide the necessary triangulation for any quantitative results.

In terms of the research process a number of lessons were identified which may be helpful for future research projects encountering similar issues:

- Working with a supportive partner who encourages a learning-by-doing approach has been extremely helpful when working in a novel area where previous research is lacking.
- Working in close cooperation with international partners is extremely important to help overcome any institutional and/or cultural barriers – a good example of this was working with the CABI team to co-develop, target and administer the questionnaire.
- Developing estimates that give a broader view of the potential order of magnitude can be informative when it is challenging or unrealistic to create monetary estimates of value.
- The data collection, identification and evaluation process has value in itself, irrespective of the final ‘number’, by providing insights into structures, procedures and impacts.

Conclusion

This report has described both a framework for the evaluation of data governance, and the results of applying that framework to a real-world project. The development of the framework was an iterative process, with multiple evaluation structures tried and rejected before the final framework was agreed. In application, the framework did appear to be practical and useful.

However, it is noticeable that the framework was most useful in providing a structure for qualitative data collection. For the quantitative data collection, the more traditional 'logic model' (inputs, activities, outputs, outcomes, impact) provided the basis for structuring data collection. Just as the results from the case study show that both qualitative and quantitative approaches are important for identifying value, both the traditional and the new framework were necessary to organise and interpret that data collection.

Turning to the specific results, the qualitative analysis highlighted multiple areas where current expenditure and investment appeared to yield the most significant gains. Two common factors repeatedly underlay these findings: the importance of building and maintaining relationships between individuals and organizations; and respect for the differences between stakeholders.

The quantitative analysis showed that it is possible to generate statistics that have both plausibility and credibility, even where the underlying intervention being measured is initially seen as 'unmeasurable'. The case study results suggested that, on a direct cost measure, the SSHiE data governance activities yielded a negative return; but they also show that relatively few 'indirect' benefits are necessary to reverse that view, at least from the point of the economy more generally.

The estimates generated for this report are exploratory and tentative, being based on a small number of data points, highly subjective answers, and some very broad assumptions about costs and benefits. Nevertheless, we believe they provide a useful illustration of how something as nebulous as the return on investments in data governance may be measured. There is almost no literature on this to date – which means, amongst other problems, that there are no other studies to compare our findings with – but we expect this will open up opportunities for future research in this area.

References

- Abebe, G. and Alemu, A. (2017) Role of Improved Seeds Towards Improving Livelihood and Food Security at Ethiopia. *International Journal of Research*, 5(2).
- Agu, C. and Caliar, A. (2014). Economic diversification and macroeconomic policies: is Africa's growth enough? In *Paper for Consideration in the CSAE Conference*.
- Aker J. (2011) Dial "A" for agriculture: a review of information and communication technologies for agricultural extension in developing countries. *Agricultural Economics* 42:631–647
- Al-Amin, A.A., Akhter, T., Islam, A.H.M.S., Jahan, H., Hossain, M.J., Prodhan, M.M.H., Mainuddin, M. and Kirby, M. (2019). An intra-household analysis of farmers' perceptions of and adaptation to climate change impacts: empirical evidence from drought prone zones of Bangladesh. *Climatic Change*, pp.1-21.
- Aldosari F., Saleh Al Shunaifi M., Ullah M. A., Muddassir M., Noor M.A. (2019) Farmers' perceptions regarding the use of Information and Communication Technology (ICT) in Khyber Pakhtunkhwa, Northern Pakistan. *Journal of the Saudi Society of Agricultural Sciences* 18:211–217
- Anderson, J., Dillon, J. (1992) *Risk analysis in dryland farming systems No 2*. Food and Agriculture Organization (FAO), Rome, Italy
- Baker, J.L. (2000) *Evaluating the impact of development projects on poverty: A handbook for practitioners*. The World Bank.
- Belmana (2018) *Interim evaluation of the UK Research Partnership Investment Fund (UKRPIF) programme*. London
- Bergemann, D., and Bonatti, A., (2018), *Markets for Information: A primer*. CEPR Discussion Paper 13148.
- Basu S. (2004) E-government and developing countries: an overview. *International Review of Law, Computers & Technology*, 18:1, 109-132, DOI:10.1080/13600860410001674779
- Birthal, S. Pratap S, Kumar, Shiv, Negi, S. Digvijay and Roy Devesh (2015). The Impact of Information on Farm Income. Policy Paper 29. October 2015. ICAR – National Institute of Agricultural Economics and Policy Research, New Delhi - 110 012
- Bollo, F., and Ngwenyama, O. (2008). Are ICTs investments paying off in Africa: An analysis of total factor productivity in sis West African countries from 1995-200. *Journal of Information Technology and International Development*. 14, pp. 294-307.
- Carletto C., Jolliffe D. and Banerjee R. (2015) From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. *Journal of Development Studies*, Vol. 51, No. 2, 133–148, <http://dx.doi.org/10.1080/00220388.2014.968140>
- Chaudhury, A., Helfgott, A., Thornton, T., Sova, C. (2014) Participatory adaptation planning and costing. Applications in agricultural adaptation in western Kenya. *Mitigation and adaption strategies for Global Change*. 21(3), pp. 301-322

Chavula, H.K., *et al.* (2017) Monetary and Other Financial Policies for Africa's Structural Transformation. In: Lopes, C., Hamdok, A. and Elhiraika, A. ed., (2017) *Macroeconomic Policy Framework for Africa's Structural Transformation*. Cham: Palgrave Macmillan. Available from <http://doi.10.1007/978-3-319-51947-0>

Clayton, G.W., Brandt, S., Johnson, E.N., O'Donovan, J.T., Harker, K.N., Blackshaw, R.E., Smith, E.G., Kutcher, H.R., Vera, C. and Hartman, M. (2009) Comparison of certified and farm-saved seed on yield and quality characteristics of canola. *Agronomy journal*, 101(6), pp.1581-1588.

Cole, S.A. and Fernando, A.N. (2014) The value of advice: Evidence from the adoption of agricultural practices. *HBS Working Group Paper*, 1(1.3), p.6.

Corrado, C., Hulten, C. and Sichel, D., 2009. Intangible capital and US economic growth. *Review of income and wealth*, 55(3), pp.661-685.

Deloitte (2017) Assessing the value of Transport for London's open data and digital partnership

Dasgupta, A., and D. Pearce. (1972) *Cost-Benefit Analysis. Theory and Practice* UK: Palgrave Macmillan.

Deichmann U., Goyalb A., and Mishrac D. (2016) Will digital technologies transform agriculture in developing countries? *Agricultural Economics* 47 supplement 21–33

Digital India (no date) Ministry of Electronics and Information Technology, Government of India [online] Available from <https://www.digitalindia.gov.in> [Accessed on 16th September]

E-Transform Africa. (2012) *The transformational use of Information and Communication Technologies in Africa. World Bank and African Union*. [online] Available from <http://siteresources.worldbank.org/extinformationandcommunicationandtechnologies/Resources/282822-1346223280837/MainReport.pdf> [Accessed 18th September 2019].

European Commission. (2008) Directorate General Regional Policy. Guide to Cost Benefit Analysis of Investment Projects. Edward Elgar Publishing

European Data Portal, (2017), "The Economic Benefits of Open Data", Analytical Report 9

Eurostat (2016) Self-study material for the users of Eurostat microdata sets. <http://ec.europa.eu/eurostat/web/microdata/overview/self-study-material-for-microdata-users>

Frontier Economics (2019) Evaluating the economic and social returns. Frontier Economics Project Report

GoAP (2015). Primary Sector Development: Status, Strategy and Action Plan. NITI Aayog.

GODAN (2016) Ownership of Open Data: Governance Options for Agriculture and Nutrition. Global Open Data for Agriculture and Nutrition. Oxford.

GODAN (2018) Digital and Data-Driven Agriculture: Harnessing the Power of Data for Smallholders. Global Open Data for Agriculture and Nutrition. Oxford.

Gorla, N. (2008) Hurdles in rural e-government projects in India: Lessons for Developing countries. *Electronic Government, An international Journal*. 5(1), pp. 91-102

- Gotsch, N. and Burger, K. (2001) Dynamic supply response and welfare effects of technological change on perennial crops: The case of cocoa in Malaysia. *American Journal of Agricultural Economics*, 83(2), pp.272-285.
- Gummagolmath, S.K.C., and Sharma, P. (2011) ICT Initiatives in Indian Agriculture-An overview. *Indiana Journal of Agricultural Economics*. 66(3), pp. 1-9.
- Hafner H-P., Lenz R., Ritchie F., and Welpton R. (2015) Evidence-based, context-sensitive, user-centred, risk-managed SDC planning: designing data access solutions for scientific use. In UNECE/Eurostat Worksession on Statistical Data Confidentiality 2015, Helsinki.
- HM Treasury, (2011a). *The green book: appraisal and evaluation in Central Government*. HM Treasury.
- HM Treasury, (2011b). *The magenta book: guidance for evaluation*. HM Treasury.
- HM Treasury, (2018), *The Economic Value of Data: A discussion paper*.
- Kuteesa A. and Kyotalimye M. (2019) Documentation and data handling: how can Africa promote record keeping and investment in data management? *African Journal of Food Agriculture Nutrition and Development* 19(1):14171-14189
- Lateral Economics (2014) *Open For Business: How Open Data Can Help Achieve the G20 Growth Target*. Omidyar Network
- Lemaga, B., Sekitto, I., Nsumba, J., Natabirwa, H., Namutebi, A., Owarwi C. and Tindiwensi, C. (2005) Cost benefit analysis of sweetpotato-based on-farm enterprises in Central Uganda. 34 pp. PRAPACE, (The Regional Network for Improvement of Potato and Sweetpotato in East and Central Africa), Kampala, Uganda.
- Little, I.M. and Mirrlees, J.A. (1974). *Project appraisal and planning for developing countries*. New York.
- Lio M. and Liu M-C (2006) ICT and agricultural productivity: evidence from cross-country data. *Agricultural Economics* 34:221–228
- London Economics (2019) *Study of the economic and social returns from the ODI's innovation programme*. London
- London Economics (2020) *Impact assessment of the ODI's R&D programme*. London
- Maini, P. and Rathore, L.S. (2011) Economic impact assessment of the Agrometeorological Advisory Service of India. *Current Science*, 101(10), pp.1296-1310.
- Makadia, J.J., Mistry, H.H. and Kuthe, S.B. (2017). Impact of Soil Health Card on fertilizer consumption and yield of sugarcane and kharif paddy in Gujarat State. *Economic Affairs*, 62(1), p.61.
- Mall, R.K., Singh, R., Gupta, A., Srinivasan, G. and Rathore, L.S. (2006). Impact of climate change on Indian agriculture: a review. *Climatic Change*, 78(2-4), pp.445-478.
- Manjappa, K. and Yeledalli, S.B. (2013) Validation and assessment of economic impact of agro advisories issued based on medium range weather forecast for Uttara Kannada district of Karnataka. *Karnataka Journal of Agricultural Sciences*, 26(1), pp.36-39.
- Maredia, M., Byerlee, D., Pee, P. (2000) Impacts of food crop improvement research: Evidence from Sub-Saharan Africa. *Food Policy* 25(5), 531–559.

Mascarenhas, O. (2010) Broadening the agenda for ICTs for poverty reduction: Picture-Africa. *Journal of Information Technology and International Development*. 6 (Special Edition), pp. 37-44.

Mawer, C., (2015), "Valuing data is hard," Silicon Valley Data science blog post available at: <https://www.svds.com/valuing-data-is-hard/>

McKinsey, (2013), Open Data: Unlocking innovation and performance with liquid information

Mittal, S. *et al.* (2010) *Socio-economic impact of the mobile phones on Indian Agriculture*. Working Paper, No. 246 Provided in Cooperation with: Indian Council for Research on International Economic Relations (ICRIER). [online] Available from <https://www.econstor.eu/bitstream/10419/176264/1/icrier-wp-246.pdf> [Accessed on 20th September 2019].

Mohan Kumar, S., Suman, S., Kulkarni, U.P. and Siddalingaswamy, N.H. (2019). An Attempt towards Structuring Agricultural Information using WhatsApp as Query redressal Social Media Platform. *J Robot Mech Eng Resr*, 3(1), pp.1-6.

OECD, (2019), "Measuring data and data flows", Chapter 9, Measuring the Digital Transformation: A roadmap for the future

OECD, (2019), "Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies", OECD Publishing, Paris,

Otto, B. (2011). Organizing data governance: Findings from the telecommunications industry and consequences for large service providers. *Communications of the AIS*, 29, 45–66.

Panganiban G. (2019) E-governance in agriculture: digital tools enabling Filipino farmers. *Journal of Asian Public Policy*, 12:1, 51-70, DOI:10.1080/17516234.2018.1499479

Parmar, I.S., Soni, P., Kuwornu, J.K. and Salin, K.R. (2019). Evaluating Farmers' Access to Agricultural Information: Evidence from Semi-Arid Region of Rajasthan State, India. *Agriculture*, 9(3), p.60.

PwC, (2019), "Putting a value on data", available at: <https://www.pwc.co.uk/data-analytics/documents/putting-valueon-data.pdf>

Raj, D.A., Murugesan, A.P., Aditya, V.P.S., Olaganathan, S. and Sasikumar, K. (2011) A crop nutrient management decision support system: India. *Strengthening Rural Livelihoods: The impact of information and communication technologies in Asia*. Rugby and Ottawa: Practical Action Publishing and International Development Research Centre, pp.33-52.

Reddy, A., (2017) Impact Study of Soil Health Card Scheme. *Impact Study of Soil Health Card Scheme, National Institute of Agricultural Extension Management (MANAGE), Hyderabad-500030*, p.210.

Savona, M (2019) The Value of Data: Towards a Framework to Redistribute It. *SPRU Working Paper*

Short, J., and Todd, S., (2017), "What's your data worth?" MIT Sloan Management Review, Spring 2017

Singh, R.P., and Singh S. (2016) Optimizing seed replacement rates in Jharkhand: present scenario, challenges and opportunities. *Jharkhand Journal of Development and Management Studies*. XISS, Ranchi, Vol. 14(2), pp. 6987-7007

Slotin, J., (2018), "What Do We Know About the Value of Data?", Global Partnership for Sustainable Development Data

Snelson, S. Wilson, C. (2019) Evaluating the economic and social returns. Frontier Economics Project Report

Spiekermann, S., et al., (2015), "The challenges of personal data markets and privacy", *Electronic markets* 25.2 (2015): 161-167.

Todaro, M.P. (1994) *Economic Development*. 5th ed. New York: Longman.

Wellcome Trust (2015) *Enabling data linkage to maximise the value of public health research data*. London, March

Wdown, J. and Diepeveen, S. (2020) *The Value of Data*. Bennett Institute for Public Policy. Cambridge

Wiggins, S. (2013). *Agriculture and Growth in Low-Income Countries. Overview of Debates and Links to Current Projects in the DFID-ESRC Growth Research Programme* [online]. London: Department for International Development (DFID). Available from: <https://www.gov.uk/dfid-research-outputs/agriculture-and-growth-in-low-income-countries-overview-of-debates-and-links-to-current-projects-in-the-dfid-esrc-growth-research-programme> [Accessed 19th September 2019].

World Bank (2006b) *Overcoming Drought Adaptation Strategies for Andhra Pradesh, India* Washington [online]. Available from <http://documents.worldbank.org/curated/en/256801468035339156/pdf/372600IN0Overc101OFFICIAL0USE0ONLY1.pdf> [Accessed on 20th September 2019].

World Bank. (2005) *India's Water Economy: Bracing for a Turbulent Future No Place* [online] Available from <http://documents.worldbank.org/curated/en/731801468033688876/pdf/34750.pdf> [Accessed on 16th September 2019].

World Bank (2006a) [online] *India water supply and sanitation: bridging the gap between infrastructure and service*. Washington Available from <http://web.worldbank.org/archive/website01291/WEB/0-co-78.HTML> [Accessed on 16th September 2019]

World Bank Report. (2019) *Understanding Poverty: Agriculture and Food*. Washington [online] Available from; <https://www.worldbank.org/en/topic/agriculture/overview#1> [Accessed on 12/09/2018]

World Bank (no date a) *How does the World Bank classify countries?* [online]. Washington: World Bank. Available from <https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries> [Accessed 21st September 2019].

World Bank (no date b) *World Bank Country and Lending Groups*. Washington: World Bank. [online] Available from <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups> [Accessed 21st September 2019].

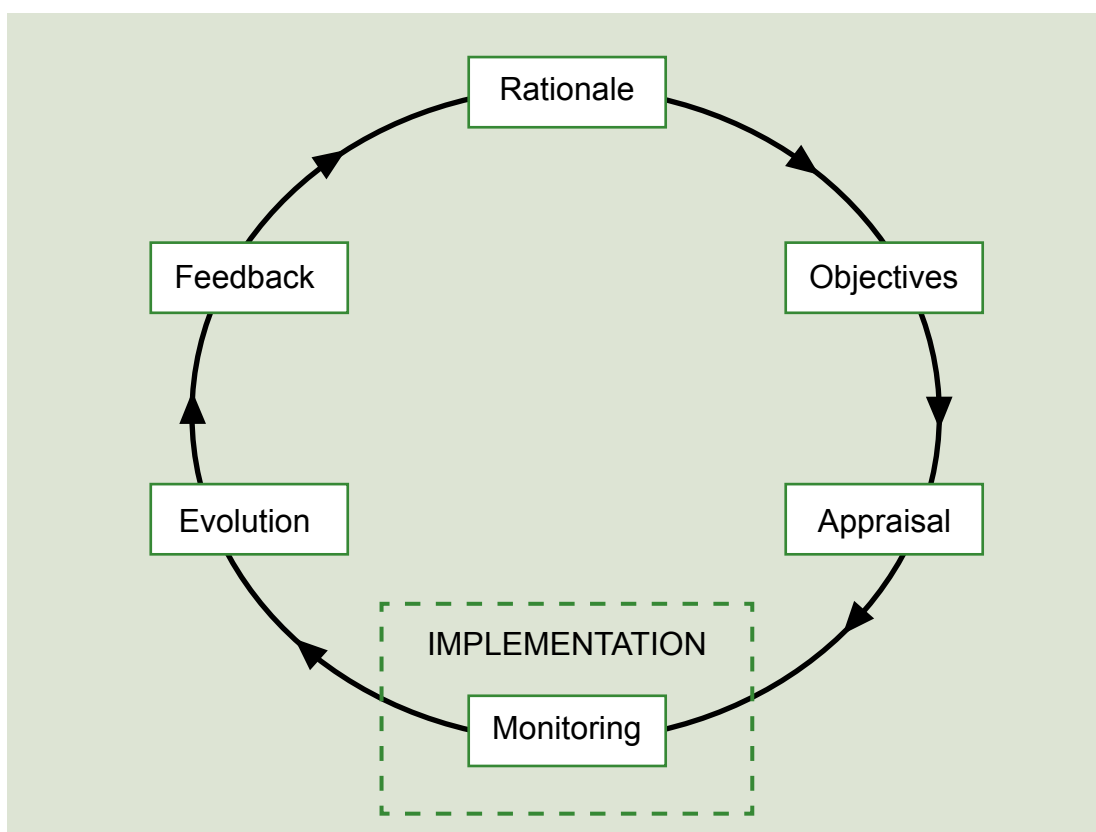
Yamano, T., Khanam, T.S. and Yaguchi, Y. (2017) *Who pays for Agricultural Information on Mobile Phones? Evidence from Three Countries in South Asia* (No. 2196-2019-1281).

Appendix 1: Extended Literature Review – Economic Evaluation

In an attempt to allocate resources more efficiently, policies, programmes and projects are increasingly subject to evaluation both at the appraisal and at completion stages. The role of the evaluation is to assess whether the intervention is worthwhile and has delivered its aims and objectives. The evaluation should then clearly communicate these results (European Commission, 2008).

Evaluations, however, should not just be viewed as an activity that is carried out after completion of an intervention, but should be seen as part of a broader project cycle. This has been formalised in the HM Treasury Green Book - Central Government Guidance on Economic Appraisal and Evaluation (2011a) in their ROAMEF (Rationale, Objectives, Appraisal, Monitoring, Evaluation and Feedback) cycle.

Figure 2: ROAMEF CYCLE



(Source: HM Treasury, 2011a)

At the start of the cycle is the rationale. HM Treasury (2011a) argues that public sector investments should be limited to dealing with market failure and issues of equity, whereas the private sector have no such limitations. The objectives of the intervention should then follow, potentially set out in a hierarchy of activities (what is delivered), outputs (what is produced in the short-term) and outcomes and impacts (medium/long-term effects). Due to practical issues in relation to measurement, valuations often take place at the level of activities and outputs, although ideally the real benefit only becomes apparent when measured at the level of outcome and impacts. The appraisal stage is akin to the evaluation stage; however, the main difference is that the appraisal options evaluated are based on forecast data rather than observations. It is stressed that the final feedback stage is important as this ensures that lessons learned are fed back into the decision making process.

There is widespread agreement that the final evaluation should not be planned after the project is complete, but should be embedded within the intervention's structure at the programmes' formulation stage. This helps to deal with issues around the collection of data for the final evaluation (Filipovici et al. 2014).

Evaluation Frameworks

There is a variety of frameworks using qualitative and advanced quantitative methods from which to approach an evaluation. The Magenta Book (HM Treasury 2011b) stresses the importance of undertaking the right evaluation for the intervention, stating that the evaluation should be tailored to the types of questions it is attempting to answer. In turn these questions should be related to the underlying "logic model", its inputs, activities, outputs, intermediate outcomes and impacts. Each approach has its advantages and challenges. For example, the case study approach uses quantitative and qualitative approaches to assess whether the intervention delivered expected outputs and outcomes. One of the key challenges with this approach is that there is a lower possibility of generalising findings.

Best practice will therefore have a two-step procedure: first determine whether something has happened (outcome-descriptive statistics), and then determine whether the programme was responsible (attribution) (HM Treasury 2011b).

Empirical impact and economic evaluations are useful when it is important to understand whether a particular intervention caused a particular outcome (Asian Development Bank, 2013). Ideally, this has a randomly-selected comparison group. Reitzer (2003) argues that

“Correlations do not establish causation when presented in isolation, so the data gathering process must illustrate that the target innovation is a primary causal factor. To do so requires that mitigating influences be identified, and the role of such in affecting productivity changes be methodically assessed.” (p8)

In order to achieve this Reitzer acknowledges the need for high quality data for both the intervention and a comparison group that is randomly selected:

“The sample size should be sufficient for deriving reasonably (given time and resource constraints) precise and repeatable figures, and the population should be randomly sampled, to prevent bias.” (p8)

Without quality data, he concludes, that “only crude extrapolations may be made”. This approach can yield statistically robust evidence when there is quantitative data available on the intervention. However, it may be that econometric analysis is not feasible; for example, where an intervention was universally implemented, or where data on behaviour change are not available. This is likely to occur in post-factum evaluations. In these cases, alternative approaches such as case studies or qualitative interviewing need to be used to extract useful information.

It is also argued that it is important, where possible, to use multiple-source verification or triangulation for key variables (Baker, 2000). Verification is particularly valuable to create confidence in non-econometric studies.

Economic Evaluation

An economic evaluation is a systematic measurement and valuation of at least two alternative interventions. These can take the form of a Cost Effectiveness Analysis (CEA), Multi-Criteria Analysis (MCA), Economic Impact Assessment (EIA) or CBA (Filipovici et al., 2014).

The Asian Development Bank (2013) reminds us of the importance of taking a social welfare (economic evaluation) as opposed to private welfare approach (financial evaluation) when evaluating public sector interventions. This is because the results of a financial evaluation can be highly misleading as key outputs may not be sold on a market or are sold in a distorted or controlled market. Indeed, in the EU’s model for assessing whether EU financial support should be given to an agribusiness project, it describes the methodical approach needed to take it from a financial analysis to an economic analysis (EU, 2008)

HM Treasury’s guide to evaluation argues that sufficient resources be allocated to deliver a successful, comprehensive but proportionate evaluation. It also states that the earlier the final evaluation is considered within the evaluation cycle, the better the results:

“crucial that the needs of any economic evaluation are considered at the design stage. Otherwise it is very likely that the evaluations will generate information which, although maybe highly interesting and valid itself, is not compatible with a cost-benefit framework, making it very difficult to undertake an economic evaluation.”

(HM Treasury, 2011b, p.20)

This need for long-term planning is echoed in Sain et al.’s (2017) study of Guatemala’s maize-beans production systems:

“To obtain more accurate information to inform future analyses for inclusion in strategic planning and policy development, it is necessary for national research institutions to plan long-term experiments or data collection schemes to establish the effects of CSA practices on farming systems and surrounding ecosystem services given their long response time.”

(p.171)

Completing an economic evaluation is complex and can pose a number of practical difficulties. Regardless, there are a number of clear principles that should be followed in order to produce a successful evaluation. These include:

- An economic evaluation should compare costs and benefits of at least two options
- The full costs should include direct and indirect costs and attributable overheads
- All benefits, both direct and indirect, should be valued unless it is clearly not practical to do – however it remains important to consider valuing the differences between the options
- All costs and benefits should be valued at market price (opportunity costs)
- Where possible, estimates for wider social and environmental costs and benefits, for which there is no market price, should be included
- The value should cover the useful lifetime of the asset encompassed
- Cost and benefits should be expressed in ‘real terms’
- Cost and benefits should be discounted due to private/social time preference
- Cost and benefits of different options should be valued and the net cost and benefit calculated

(European Commission, 2008; HM Treasury, 2011a, HM Treasury, 2011b, Asian Development Bank, 2013).

As described above, one of the key issues in economic evaluation is being able to compare outcomes of different options. The reason behind this is to consider if there are better uses for the resources and better ways to achieve the objectives. Therefore, at the completion stage, the evaluation should produce credible estimates of the counterfactual.

“to ensure methodological rigor, an impact valuation must estimate the counterfactual, that is, what would have happened had the project never taken place or what otherwise would have been”

(Baker, 2000).

A further issue is that of time preference and as such estimating the relevant discount rate. Time preference relates to two issues. First, individuals expect their income to increase and hence their level of consumption to grow in the future and therefore their marginal utility of consumption would fall in the future. Second, pure time preference assumes that individuals prefer to consume now rather than in the future, partially because of impatience and partially because of the risk of not being alive in the future. Social time preference, therefore is the rate at which society is willing to postpone a marginal unit of current consumption for future consumption (Dasgupta and Pearce, 1972)

The discount rate of future benefits and costs allows costs and benefits with different time spans to be compared on a common “present value” basis. There is much debate as to the most appropriate method to calculate the discount rate; however, a common approach is to use the real interest rate on money (Sain et al. 2017).

An example of the wide disparity in the use of discount rates can be seen when comparing the estimates across different countries and sectors. HM Treasury (2011a) suggest using a discount rate of 3.5% when assessing public sector investments in the UK. The Asian Development Bank (2013) report significant variations in public discount rate, with developing countries in general applying a higher social discount rate (SDR) of between 8% and 15%. It argues further that this reflects both perceived differences in social opportunity cost across

countries, and how intergenerational equity is taken into consideration. Further examples of the higher rates used in LMICs include the 12% used in the CBA of Guatemala's agricultural sector (Kometter, 2012; Sain et al., 2017) and the 13% used in the evaluation of sweet potato technology in Uganda (Lemaga et al., 2005).

Non-market Valuation

As discussed, within a cost benefit analysis framework, it is important to move past just a financial evaluation of the project to a full economic evaluation that should include estimates of social and environmental costs and benefits. It is recognised that it may be difficult to directly observe and measure all cost and benefits. However, results of previous studies can be helpful in estimating the economic value of changes stemming from current programme. It must be noted that care must be taken to allow for different circumstances (HM Treasury 2011a).

It is also possible to apply a monetary value to non-market goods by looking at choices that people make. This can be done by either observing related (revealed preference) or hypothetical markets (stated preference). Where the price cannot be determined from market data it is possible to use 'willingness to pay' for a benefit and 'willingness to accept' techniques to estimate the price (HM Treasury, 2011a). However, depending on resource constraints, it may be sufficient to look for lower cost options in data collection such as estimating data points with expert groups (Sain et al, 2017).

Filipovici et al. (2014), however, acknowledges that CBA may not be limited only to monetary considerations, particularly in relation to environmental and social costs and benefits. When it is not possible to value all cost and benefits in monetary terms, qualitative measures should be included, adding to the holistic measure of net present value.

Uncertainty

Other areas of uncertainty within a CBA can be addressed through estimating the probabilities of likelihood of risk and size of outcome (Anderson and Dillon, 1992). Alternatively, where there is uncertainty in the estimates, it may be useful to use maximum and minimum likely outcomes as well as central estimates. Therefore, it is good practice to conduct a sensitivity analysis around the discount rate and other uncertainties to present a range of values estimated under different scenarios and assumptions. A good example of this is Sain et al. (2017), who used probabilistic CBA in an analysis of maize-beans production in Guatemala. They used probabilistic modelling to address field variability and high uncertainty around parameter values. It is clear that limiting discussions of risks and critical externalities will distort understanding and therefore every effort must be made to include these estimates within any CBA (Sain et al: 2017; Chaudhury et al., 2014).

Unintended Consequences

As well as the direct cost and benefits, it is important to consider the wider effects and how these may influence the overall evaluation. For example, it is good practice to take account of any unintended consequences such as displacement (e.g. positive outcome offset by negative outcome elsewhere), substitution (the effects of an intervention are realised at the expense of others), leakage (benefits others outside the target area or group) and deadweight (outcomes which would have occurred anyway) (HM Treasury, 2011a).

Appendix 2: Expert Questionnaire

Section 1: Work Package 0 (inception phase)

This section relates to the inception phase to analyse data needs and bottlenecks, assess gaps for the use cases, and mobilise stakeholders

Section 1 - Activity 1: Direct impacts from the Inception Workshop

This question relates to the initial project workshop

Question	Number	
1. What proportion of attendees for the Inception Workshop were from the following groups?	Type of worker	%
	Government official	
	Other public sector worker	
	Scientist/academic/researcher	
	International agency staff	
	External consultants	
	Other workers	
2. What percentage of the Inception Workshop attendees were directly employed by the SSHiE project?		
3. Please list (if any) publications produced that were delivered or funded through the Inception Workshop that have an impact beyond this work package?		
4. What percentage of these publications were produced by individuals directly employed by the SSHiE project?		

* Q5 below asks you to consider whether an activity led to a decrease (time saving) and/or increase in the amount of time (additional time) it would have taken to complete any subsequent work packages. Q6 then asks you to attribute this decrease and/or increase to particular categories. See 'Estimating time saved or lost' section of the instructions for further detail on how to answer these questions.

If you believe the activity only led to a **time saving** to complete subsequent activities then please just **fill in column 1 of Q5 and Q6**.

If you believe the activity only led to a time increase to complete subsequent activities then please just **fill in column 2 of Q5 and Q6**.

However, if you believe that the activity in part led to a time saving for certain reasons, but a time increase for other reasons, then please **fill in both column 1 and column 2 of Q5 and Q6**.

Q7 is optional but allows you an opportunity to explain in more detail how you believe the activity led to a time saving, time increase or both.

5. As a result of the Inception Workshop please estimate the time saved and/or additional time needed to complete subsequent work packages.	Time	Time saved	Extra time needed
	Months		
	Weeks		
	Days		
6. Of your estimate in Q5 of the time saved and/or additional time needed, please estimate the proportion attributable to the following reasons. <i>* for description of 'Champions' and 'Blockers' see instructions</i>	Reason for change	% share of time saved	% share of extra time
	Increase/decrease in Trust between stakeholders		
	Positive/negative engagement of 'Blockers'		
	Positive/negative Change in Culture in relation to data sharing		
	Increase/decrease in Data Governance Knowledge and Skills of stakeholders		
	Increase/decrease Consistency of Data Capture		
	Other (please specify)		
7. Please provide in more detail how you believe the activity led to a time saving, time increase or both [optional].			

Section 1 - Activity 2: Direct impacts from the sub-grant agreement

This question relates to the sub-grant agreement with International Centre for Tropical Agriculture (CIAT) (Building Capacity Initiatives)

Question	Number	
1. In order to build capacity of national partners in the development and analytics of big data, please estimate the number of training/knowledge transfer events delivered or funded through the CIAT Sub-grant .		
2. On average, how many people would attend the training/knowledge transfer events delivered or funded through the CIAT Sub-grant ?		
3. On average, how much time would individuals spend attending the training/knowledge transfer events delivered or funded through the CIAT Sub-grant ?	Unit	Time
	Months	
	Weeks	
	Days	
4. Of those attending training/knowledge transfer events delivered or funded through the CIAT Sub-grant , what was the proportion from the following groups?	Type of worker	%
	Government official	
	Other public sector worker	
	Scientist/academic/researcher	
	International agency staff	
	Other workers	
5. What percentage of those attending training/knowledge transfer events were directly employed by the SSHiE project?		
6. Please list (if any) publications produced that were delivered or funded through the CIAT Sub-grant that have an impact beyond this work package?		
7. What percentage of these publications were produced by individuals directly employed by the SSHiE project?		

8. As a result of the CIAT Sub-grant please estimate the time saved and/or additional time needed to complete subsequent work packages.	Time	Time saved	Extra time needed
	Months		
	Weeks		
	Days		
9. Of your estimate in Q8 of the time saved and/or additional time needed, please estimate the proportion attributable to the following reasons. <i>* for description of 'Champions' and 'Blockers' see instructions</i>	Reason for change	% share of time saved	% share of extra time
	Increase/decrease in Trust between stakeholders		
	Increase/decrease engagement of high-level ' Champions '		
	Positive/negative engagement of ' Blockers '		
	Positive/negative Change in Culture in relation to data sharing		
	Increase/decrease in Data Governance Knowledge and Skills of stakeholders		
	Increase/decrease Consistency of Data Capture		
	Other (please specify)		
10. Please provide in more detail how you believe the activity led to a time saving, time increase or both [optional].			

Section 1: Assessment of accuracy of your answers to Section 1

On a scale of 0-10 , how would you rate your level of confidence in the accuracy of your answers to Section 1	
---	--

Section 2: Work Package 1 (stakeholder policy)

This section relates to the work to craft a policy for stakeholders across organizations to share data and collaboratively develop solutions to easing Ethiopian soil system bottlenecks

Section 2: Activity 1: Direct impacts from the Open Data Sensitization Workshop

Question	Number		
1. What proportion of attendees for the Open Data Sensitization Workshop were from the following groups?	Type of worker		%
	Government official		
	Other public sector worker		
	Scientist/academic/researcher		
	International agency staff		
	External consultants		
	Other workers		
2. What percentage of the Open Data Sensitization Workshop attendees were directly employed by the SSHiE project?			
3. Please list (if any) publications produced that were delivered or funded through the Open Data Sensitization Workshop that have an impact beyond this work package?			
4. What percentage of these publications were produced by individuals directly employed by the SSHiE project?			
5. As a result of the Open Data Sensitization Workshop please estimate the time saved and/or additional time needed to complete subsequent work packages.	Time		Time saved
	Months		
	Weeks		
	Days		
		Extra time needed	

<p>6. Of your estimate in Q5 of the time saved and/or additional time needed, please estimate the proportion attributable to the following reasons.</p> <p><i>* for description of 'Champions' and 'Blockers' see instructions</i></p>	Reason for change	% share of time saved	% share of extra time
	Increase/decrease in Trust between stakeholders		
	Increase/decrease engagement of high-level 'Champions'		
	Positive/negative engagement of 'Blockers'		
	Positive/negative Change in Culture in relation to data sharing		
	Increase/decrease in Data Governance Knowledge and Skills of stakeholders		
	Increase/decrease Consistency of Data Capture		
	Other (please specify)		
7. Please provide in more detail how you believe the activity led to a time saving, time increase or both [optional].			

Section 2 - Activity 2: Direct impacts from the 'Coalition of the Willing'

Question	Number		
1. What was the total number of individuals that joined the Coalition of the Willing (CoW) ?			
2. Of those, what percentage were directly employed by the SSHiE project?			
3. On average , each year how much time in total would individuals invest working as part of the CoW?	Time	Time saved	Extra time needed
	Months		
	Weeks		
	Days		
4. What was the proportion of the CoW who were from the following groups?	Type of worker		%
	Government official		
	Other public sector worker		
	Scientist/academic/researcher		
	International agency staff		
	External consultants		
	Other workers		

5. Please list (if any) publications produced that were delivered or funded through the CoW that have an impact beyond this work package?	
6. What percentage of these publications were produced by individuals directly employed by the SSHiE project?	

7. As a result of the CoW please estimate the time saved and/or additional time needed to complete subsequent work packages.	Time	Time saved	Extra time needed
	Months		
	Weeks		
	Days		
8. Of your estimate in Q7 of the time saved and/or additional time needed, please estimate the proportion attributable to the following reasons. <i>* for description of 'Champions' and 'Blockers' see instructions</i>	Reason for change	% share of time saved	% share of extra time
	Increase/decrease in Trust between stakeholders		
	Increase/decrease engagement of high-level ' Champions '		
	Positive/negative engagement of ' Blockers '		
	Positive/negative Change in Culture in relation to data sharing		
	Increase/decrease in Data Governance Knowledge and Skills of stakeholders		
	Increase/decrease Consistency of Data Capture		
	Other (please specify)		
9. Please provide in more detail how you believe the activity led to a time saving, time increase or both [optional].			

Section 2 - Activity 3: Direct Impacts from Consultation Workshop on Data Access and Sharing

Question	Number		
1. What was the PROPORTION of attendees for the Consultation Workshop on Data Access and Sharing (CWoDAaS) event were from the following groups?	Type of worker	%	
	Government official		
	Other public sector worker		
	Scientist/academic/researcher		
	International agency staff		
	External consultants		
Other workers			
2. What PERCENTAGE of the attendees for the CWoDAaS event were directly employed by the SSHiE project?			
3. Please list (if any) publications produced that were delivered or funded through the CWoDAaS event that have an impact beyond this work package?			
4. What percentage of these publications were produced by individuals directly employed by the SSHiE project?			
5. As a result of the CWoDAaS event please estimate the time saved and/or additional time needed to complete subsequent work packages.	Time	Time saved	Extra time needed
	Months		
	Weeks		
	Days		

<p>6. Of your estimate in Q5 of the time saved and/or additional time needed, please estimate the proportion attributable to the following reasons.</p> <p><i>* for description of 'Champions' and 'Blockers' see instructions</i></p>	Reason for change	% share of time saved	% share of extra time
	Increase/decrease in Trust between stakeholders		
	Increase/decrease engagement of high-level 'Champions'		
	Positive/negative engagement of 'Blockers'		
	Positive/negative Change in Culture in relation to data sharing		
	Increase/decrease in Data Governance Knowledge and Skills of stakeholders		
	Increase/decrease Consistency of Data Capture		
	Other (please specify)		
7. Please provide in more detail how you believe the activity led to a time saving, time increase or both [optional].			

Section 2 - Activity 4: Direct impacts from the Data Sharing Task Force

Question	Number		
1. How many individuals contributed to the data sharing task force (DSTF) set up facilitate data mining approaches using locally generated data?			
2. What was the percentage of people who were directly funded to be part of the DSTF ?			
3. On average , over a year how much time would individual contribute to the DSTF?	Unit	Time	
	Months		
	Weeks		
4. Of those on the DSTF , what was the proportion from the following groups?	Type of worker	%	
	Government official		
	Other public sector worker		
	Scientist/academic/researcher		
	International agency staff		
	External consultants		
5. Please list (if any) publications produced that were delivered or funded through the DSTF that have an impact beyond this work package?			
6. What percentage of these publications were produced by individuals directly employed by the SSHiE project?			
7. As a result of the DSTF please estimate the time saved and/or additional time needed to complete subsequent work packages.	Time	Time saved	Extra time needed
	Months		
	Weeks		
	Days		

<p>8. Of your estimate in Q7 of the time saved and/or additional time needed, please estimate the proportion attributable to the following reasons.</p> <p><i>* for description of 'Champions' and 'Blockers' see instructions</i></p>	Reason for change	% share of time saved	% share of extra time
	Increase/decrease in Trust between stakeholders		
	Increase/decrease engagement of high-level 'Champions'		
	Positive/negative engagement of 'Blockers'		
	Positive/negative Change in Culture in relation to data sharing		
	Increase/decrease in Data Governance Knowledge and Skills of stakeholders		
	Increase/decrease Consistency of Data Capture		
	Other (please specify)		
9. Please provide in more detail how you believe the activity led to a time saving, time increase or both [optional].			

Section 2 – Assessment of accuracy of your answers to Section 2

1. On a scale of 0-10 , how would you rate your level of confidence in the accuracy of your answers to Section 2	
--	--

Section 3: Work Package 2 (technical and operational support)

This section relates to the work to provide technical and operational support for a national IT host for a centralized data repository.

Section 3 - Activity 1: Direct impacts from development of the Stakeholder Consultation Workshop

Question	Number	
1. What was the total number of attendees for the Stakeholder Consultation Workshop (SCW) ?		
2. How long in terms of time was the SCW ?	Unit	Time
	Days	
	Hours	
3. Of those attending the SCW , what was the proportion from the following groups?	Type of worker	%
	Government official	
	Other public sector worker	
	Scientist/academic/researcher	
	International agency staff	
	External consultants	
	Other workers	
4. What percentage of the SCW attendees were directly employed by the SSHiE project?		
5. Please list (if any) publications produced that were delivered or funded through the SCW that have an impact beyond this work package?		
6. What percentage of these publications were produced by individuals directly employed by the SSHiE project?		

7. As a result of the SCW please estimate the time saved and/or additional time needed to complete subsequent work packages.	Time	Time saved	Extra time needed
	Months		
	Weeks		
	Days		
8. Of your estimate in Q7 of the time saved and/or additional time needed, please estimate the proportion attributable to the following reasons. <i>* for description of 'Champions' and 'Blockers' see instructions</i>	Reason for change	% share of time saved	% share of extra time
	Increase/decrease in Trust between stakeholders		
	Increase/decrease engagement of high-level ' Champions '		
	Positive/negative engagement of ' Blockers '		
	Positive/negative Change in Culture in relation to data sharing		
	Increase/decrease in Data Governance Knowledge and Skills of stakeholders		
	Increase/decrease Consistency of Data Capture		
	Other (please specify)		
9. Please provide in more detail how you believe the activity led to a time saving, time increase or both [optional].			

Section 3 - Activity 2: Direct Impacts for the need identification study

This question relates to the work to identify Human Capacity Development and IT Support needs

Question	Number		
1. Please estimate the number of individuals who benefitted increasing the human capital as a result of the identification of development needs, particularly in relation to the migration of ETHioSIS data to ESRI's geo-spatial database?			
2. As a result of this early identification of development needs, on average , how much additional time was spent on training these individuals?	Unit	Time	
	Months		
	Weeks		
	Days		
3. As a result of this upskilling and increase in human capital please estimate the time saved and/or additional time needed to complete subsequent work packages.	Time	Time saved	Extra time needed
	Months		
	Weeks		
	Days		
4. Of your estimate in Q3 of the time saved and/or additional time needed, please estimate the proportion attributable to the following reasons. <i>* for description of 'Champions' and 'Blockers' see instructions</i>	Reason for change	% share of time saved	% share of extra time
	Increase/decrease in Trust between stakeholders		
	Increase/decrease engagement of high-level ' Champions '		
	Positive/negative engagement of ' Blockers '		
	Positive/negative Change in Culture in relation to data sharing		
	Increase/decrease in Data Governance Knowledge and Skills of stakeholders		
	Increase/decrease Consistency of Data Capture		
	Other (please specify)		
5. Please provide in more detail how you believe the activity led to a time saving, time increase or both [optional].			

Section 3 - Activity 3: Direct impacts from international exposure visits

This question relates to the impact of international visits such as that to the Bangladesh Soil Resource Development Institute.

Question	Number		
1. What was the total number of participants who went on international exposure visits (e.g. to Bangladesh's Soil Resource Centre (BSRC))			
1. On average for each individual, how much time was spent on knowledge exchange activities on these international exposure visits ?	Unit	Time	
	Months		
	Weeks		
1. What was the proportion of the participants on these international exposure visits who were from the following groups?	Type of worker	(%)	
	Government official		
	Other public sector worker		
	Scientist/academic/researcher		
	International agency staff		
	External consultants		
1. As a result of the international exposure visits please estimate the time saved and/or additional time needed to complete subsequent work packages.	Time	Time saved	Extra time needed
	Months		
	Weeks		
	Days		

<p>1. Of your estimate in Q4 of the time saved and/or additional time needed, please estimate the proportion attributable to the following reasons.</p> <p><i>* for description of 'Champions' and 'Blockers' see instructions</i></p>	Reason for change	% share of time saved	% share of extra time
	Increase/decrease in Trust between stakeholders		
	Increase/decrease engagement of high-level 'Champions'		
	Positive/negative engagement of 'Blockers'		
	Positive/negative Change in Culture in relation to data sharing		
	Increase/decrease in Data Governance Knowledge and Skills of stakeholders		
	Increase/decrease Consistency of Data Capture		
	Other (please specify)		
1. Please provide in more detail how you believe the activity led to a time saving, time increase or both [optional].			

Section 3 - Assessment of accuracy of your answers to questions in Section 3

1. On a scale of 0-10 , how would you rate your level of confidence in the accuracy of your answers to Section 3	
--	--

Section 4: Assessment of longer-term outcomes

In this section we are interested in assessing longer-term outcomes due to improvements in data governance brought about by the SSHiE project.

The initial round of interviews revealed that the final potential outcomes of the SSHiE project could be broad-based and wide ranging. These are 'spill-over effects': they 'spill out' from the project and have a wider impact on society. It is very challenging to estimate the extent of these impacts, so in this section we are aiming to estimate the potential impact and likelihood of such outcomes.

In this section we will list a number of potential scenarios and we would like you to rate on a scale of 1-5, the likelihood of the impact happening and the potential impact such an outcome would produce. The likelihood and impact should be judged using the following criteria.

5 Point Likelihood Score	
1	Not likely to occur
2	Possible but in rare circumstances
3	Probable to occur
4	Very likely to occur
5	Almost certain to occur

5 Point Impact Score	
1	No impact
2	Minor impact on a limited region or sub-group of people
3	Major impact on a limited region or sub-group of people
4	Minor impact nationally
5	Major impact nationally

Please estimate the likelihood of the impact happening and the potential impact such an outcome would produce from the following scenarios.		
The SSHiE project will either directly or indirectly through spill-over effects lead to...	Likelihood Score	Impact Score
improved level of trust between stakeholders in...		
the soil and agronomy sector		
the agricultural sector		
other sectors (e.g. manufacturing/services)		
change of culture towards data sharing and a more open data framework for...		
the soil and agronomy sector		
the agricultural sector		
other sectors		
improved levels of human capital in data governance and data access for...		
the soil and agronomy sector		
the agricultural sector		
other sectors.		
improved infrastructure and technology for...		
the soil and agronomy sector		
the agricultural sector		
other sectors		
improved data sharing and access policies for...		
the soil and agronomy sector		
the agricultural sector		
other sectors		
improved economic policies for...		
the soil and agronomy sector		
the agricultural sector		
other sectors		
improved regional and national resource allocation for...		
the soil and agronomy sector		
the agricultural sector		
other sectors		
improved productivity for...		
the soil and agronomy sector		
the agricultural sector		
other sectors		

THANK YOU FOR COMPLETING THE SURVEY

Appendix 3: Assumptions made to generate quantitative estimates

1. Internal direct value calculated as follows:

Value of activity/attendance = number of events*number of people*hours*hourly wage rate

2. Training/workshop events estimated to be six hours a day unless specifically asked in questionnaire

3. When answers to questions should equal 100%, where this did not occur for an underestimation the residual would be allocated to 'other' while any overestimation would be taken off equally in absolute terms across the board.

4. Weighted averages are conditioned on answer to confidence level questions; where no response has been given to confidence level questions they have been assigned a value of 0.5% (neither confident nor unconfident).

5. Estimates of time were converted as follows: 1 month = 4 weeks = 20 days = 160 hours

6. Number of publications was derived from using the highest figure for each category where information on the type of publication is given. However, where no information on the types of publications are provide the average figure is used

7. Training in relation to activity 3.1 (Human Capacity and Development) is fully allocated to government officials as the training only relates to upskilling of staff of national IT centralised data repository.

9. Value of time saved is calculated as follows:

= amount of time multiplied by average wage rate

10. Average salary rates for the six sectors were provided by CABI's Ethiopian staff. They were verified by checking against Salary Explorer.

[http://www.salaryexplorer.com/salary-survey.php?loc=69&loctype=1#:~:text=The%20average%20hourly%20wage%20\(pay,ETB%20for%20every%20worked%20hour.](http://www.salaryexplorer.com/salary-survey.php?loc=69&loctype=1#:~:text=The%20average%20hourly%20wage%20(pay,ETB%20for%20every%20worked%20hour.)

Two amendments were made to the salary estimates provided by CABI. For scientists/academics/researchers CABI provide a range of estimate for annual income (1000-5000). In order not to overestimate the impact, the lowest estimate was used, as the highest estimate was attributed to international workers and was significantly greater than comparable estimates on Salary Explorer.

For international agency workers the estimate was 200-1000 per annum. As the spread was narrower and due to the fact that potentially a considerable proportion of the international agency staff would not be permanently be based in Ethiopia, the middle estimate (600) was used.

11. The exchange rate as of 14th January 2021 and as listed on xe.com was used for conversion from Ethiopian Birr. The rates were as follows:

£1 = 44 ETB Br

€1 = 40 ETB Br

\$1 = 35 ETB Br

Appendix 4: CSISA III Methodology

The Macro Approach

This approach will focus on the CSISA III project as a whole.

In line with a case study approach, the first phase will be to undertake an initial discussion with top level project managers in order to find out (a) what specific problems the project was addressing, compared to CSISA I & II, (b) the intended outcomes of the project, (c) whether data governance or sustainable data capacity building were explicit objectives, and (d) how attitudes to data governance and management were changing compared to previous projects. The discussion will help us understand the context of the overall project. The interviews with the Project Managers will be used to point to us where the information and data sources are.

Initially, we need to ask:

1. Why was CSISA III necessary? Why was there a need for CSISA III?
2. Why was CSISA III different from CSISA I and CSISA II? In particular, were there differences in the way data or information was seen as an input to the projects?
3. How was data used/accessed across the programme?
4. Was there a conscious decision to increase investment in data governance/access?
5. What was the impact of data governance and data access on the outcomes of the project?
6. What were the benefits of data governance and data access to the project?
7. Was it useful to invest in data governance and data access; if so, why?

If this amount of money was not spent on data activities what could have been the outcome?

1. We would also like to ask, perhaps via email, more technical questions:
2. What proportion of the project costs was spent on data governance and data access?
3. What proportion of the programme team's time was spent on data governance and data access issues
4. What proportion of the overall benefits were derived from data governance and data access activities?
5. How much time did data governance and data access agreements save the project team?

The Micro Approach

The micro approach can be used to validate the evidence of the macro approach by building the cost benefit analysis up from the level of the sub projects. The starting point for the micro approach is to isolate the elements of data governance and data access within the sub-project.

We propose focusing on the 'geospatial' project within CSISAIII. It is not entirely clear if this is a single project, or an application across various areas of the overall project. Nevertheless, there appear to be specific targets and outcomes (linking a database, providing access to it, creating information about it), which suggest we can focus on this particular investment.

Within the project, we will use a framework to assess whether sub-project investment activities are primarily data governance/data access activities. These elements will form the overall evaluation. If data is available at a higher level but not directly available at lower levels of aggregation (e.g. budget information so far seems very high-level), estimation techniques will be applied in order to estimate costs and benefits.

These estimations will be informed by discussions with experts from the programme. Given the uncertainties surrounding such estimates, UWE's research team will assess the robustness and the generalisability of the findings. Any judgements will be analysed using Maredia and Raitzer's (2010) 'best evidence' approach to assess the credibility of the results produced.

The collection of (cost and benefit) data for the project will come from four potential data sources

- Project documentation
- Interviews
- Questionnaires
- Secondary data sources

We will initially need a discussion with the sub-project, project managers. This will mainly focus on the implementation phase. It aims at exploring the overall objective of the sub-project, and teasing out specific governance/access activities. This might include questions such as:

1. What was the problem that you were trying to resolve?
2. Please explain the flow of data through the project
 - What was pre-existing data?
 - How was it acquired/transformed?
 - How was it used?
 - Where was it held?
 - How was it managed?
 - What happened/will happen to the data at project end?

3. What services were developed from the data?
4. What would have been the effect on the wider project if data or data delivery mechanisms had not been available?
5. What were the main challenges and lessons learned?
 - What were the problems around access to data?
 - What things could have been done differently (with respect to data access; how solid is the ethical foundation i.e. potential risks).

After these initial interviews/questions we will move towards structuring a questionnaire. Therefore the next stage is for the UWE research team to work with the CABI project team to set up the preliminary interviews. For the macro project, the initial interviews will be with the top level project managers. For the micro project, the initial interviews will be with the sub-project managers.

The target date to complete these interviews is by the end of November 2019.

Data Sharing Toolkit



ACKNOWLEDGEMENTS

This document was authored by the Open Data Institute and CABI as part of a Bill & Melinda Gates Foundation funded investment.

The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation or CABI.

datasharingtoolkit.org

cabi.org | theodi.org | gatesfoundation.org

 **CABI** Data Sharing Toolkit



BILL & MELINDA
GATES *foundation*



Except where otherwise noted, content on this site is licensed under a Creative Commons Attribution 4.0 International license.