A Survey and Analysis of the Data Requirements for Stakeholders in African Agriculture

An initiative of the CGIAR Program on Water, Land & Ecosystems

Commissioned by the UK Department for International Development (DFID)

9 August 2013

CGIAR Research Program on Water, Land and Ecosystems

World Agroforestry Centre
TRANSFORMING LIVES AND LANDSCAPES

Hubbard Decision Research

A Survey and Analysis of the Data Requirements for Stakeholders in African Agriculture

Aaron Clapp, Nathaniel DauSchmidt, Matthew Millar, and Douglas Hubbard

Hubbard Decision Research, Glen Ellyn, IL USA

Keith Shepherd

World Agroforestry Centre (ICRAF), Nairobi, Kenya

Acknowledgements

We acknowledge financial support from DFID and helpful review comments from Professor Margaret Gill (DFID). We thank Grace Muinga for her dedicated support in web research, communications with stakeholders, maintaining stakeholder databases, and arranging interview schedules.

Summary

As part of DFID's contribution to the G8 initiative on Open Data for Agriculture, a broad survey of key stakeholders in sustainable African agriculture was conducted to assess current and emerging trends related to data collection, processing, and dissemination. Stakeholders that promote and support sustainable intensification of agriculture in Africa require access to useful data upon which to base their decisions and evaluate current and future interventions amid limited resources. Information has value for supporting a decision only if it reduces the chance of being wrong and a cost of being wrong. Research across many fields has shown that quantitative decision analysis methods overwhelmingly outperform expert judgment in identifying the economic value of information and improving decisions. Therefore a key focus of the study was to assess the alignment of stakeholders' perceived data needs with areas of decision uncertainty.

A total of 281 stakeholders were contacted by email among 11 organizational categories. Of this list, there were 110 respondents to the online survey with 58 individuals further contributing to in-depth conversations. Results from the online survey were compiled along with live interview responses into a centralized database for analysis. In addition to searching for overall keyword trends and tendencies within groups, we examined motivations for data and use, and whether these data were informing specific decisions.

Less than half of respondents (46%) could specify a decision of any kind in relation to their perceived data needs. Only 36% of respondents stated data needs that were consistent with their stated uncertainties and only 15% showed that perceived needs, uncertainties, and data gathering efforts are aligned. There was broad alignment among effort, perceived needs, and uncertainties for soil data. In other words, soil was the most frequently cited uncertainty, the most frequently stated perceived need, and the most frequently stated focus of current effort. Market data showed similar high priority and alignment. Climate data is frequently cited as both needed and satisfying an uncertainty, but is less frequently cited as a focus of current effort. Biodiversity and poverty data are frequently cited as a focus of effort but infrequently cited as a perceived need or uncertainty. Consistent with other case studies, overall there is evidence of a "measurement inversion" where decisions are either poorly defined or data priorities are disconnected from the decisions they potentially inform.

Based on the survey results and analysis, we provide recommendations for improving the collection and use of data in African agriculture. Comprehensive, centralized web-enabled GIS databases could have large impact on improving decisions if efforts are prioritized by information values. Databases and information gathering requirements aimed at development impact should be based on information values identified by quantitative modeling of key decisions rather than by routine, intuition, or purely subjective means. Initiatives to develop awareness of the key decisions and what data is needed to support them should be widespread and routine. These will inform decision makers and researchers on how to spend their limited resources measuring only the most important information or variables. Researchers should also make greater use of decision analysis techniques that explicitly handle uncertain or incomplete data. This may help optimize the use of existing data for improving specific development decisions.

Introduction

In view of the ever-increasing demands for food production worldwide, there is a critical need for beneficial and sustainable changes in agricultural practices that result in improved yields while minimizing negative environmental impacts. Stakeholders that promote and support sustainable intensification of agriculture in Africa therefore require access to useful data upon which to base their decisions and evaluate current and future interventions amid limited resources.

To better understand and characterize the current state of data management practices and identify future needs, the UK Department for International Development (DFID) commissioned the World Agroforestry Centre (ICRAF) and Hubbard Decision Research (HDR) to survey relevant stakeholders in African agriculture with the intent of identifying areas where future policies and projects can have beneficial impact on food security, especially in developing countries in Africa. The initial findings of the survey were presented to at the G8 International Conference on Open Data for Agriculture, held in Washington, DC from 29th to 30th April 2013, as part of an initiative that aims to "*develop options for the establishment of a global platform to make reliable agricultural and related information available to African farmers, researchers and policymakers, taking into account existing agricultural data systems*." The study aims to inform those taking forward further work within DFID and by the G8 countries post-conference.

In this paper, we report the compiled results of the stakeholder survey, conducted in two phases (online and interview), analysis of results, and recommendations based on lessons from Applied Information Economics (AIE) which prioritizes data needs based on their value of informing specific decisions (Hubbard 2010). From this perspective, we can identify disconnects between current data priorities held by stakeholders and a modified approach that leads to improved decisions and subsequent outcomes in African agriculture.

Objectives

The specific objectives of this work are to:

1. Identify and survey categories of potential stakeholders in Africa (e.g. researchers through to decision-makers within government and including the private sector) who might be interested in having access to data relevant to enhancing food security
2. Consult a feasible number of users through electronic means where possible (and alternative approaches as necessary) on the types of data which they would find useful, the main purposes for which the data (e.g. the decisions the data would actually inform), would be used and the format in which the data would be most helpful
3. Analyze these findings and compare the perceived data requirements with data requirements consistent with uncertainties related to key decisions
4. Provide specific recommendations for data prioritization and dissemination practices which improve decision making
5. Inform those taking forward further work post-conference

Background: Science-based decision making

Ambitious initiatives have been recently undertaken to promote data sharing and access in agriculture and ecosystems (Besemer et al 2012). The wide availability of vast amounts of data should be easier now than ever before in any field of research. This data availability should greatly improve intervention and policy decisions. Yet, the actual experience of many researchers falls short of this. A study published in 2012 states that "Available agricultural data and indicators are often inadequate for decision makers to formulate efficient and equitable public and private sector investments." (Pica-Ciamarra, et al 2012). Another recent DFID-commissioned global review of monitoring initiatives found scant evidence for impact on real world decisions suggesting that either the wrong things are often being measured, or there is no substantial link between these measurements and decisions that they might be expected to support (Shepherd, et al 2013). If technology alone were the only constraint, these would no longer be issues.

Different stakeholder groups are routinely making sets of critical decisions related to agriculture. For example, donors and research managers decide on which projects to fund and how to monitor impacts of investments made; research scientists decide on what to research, which research design to use and what to measure; regional organizations decide on which programs to support and promote; governments decide on which policies, programs and projects to implement, and what to measure to track their performance; universities decide on how to allocate resources among departments and what material to include in their curricula; the private sector decide on which R&D projects to invest in and which products to promote; development NGOs decide on which practices to extend where; farmers decide on which crops to plant, when and how much area to plant, and what inputs to buy. All of these examples are decisions characterized by the need to allocate scarce resources among alternative actions under conditions of large uncertainty in outcomes. Applied research should support these decisions.

Critical agricultural and ecosystem management decisions are made with a variety of decision analysis methods. But researchers and decision makers alike should be aware there are significant variations in the performance of these methods. The only way to properly identify what decision analysis methods work is to measure their performance as opposed to rely on perceptions of their performance (Raitzer et al 2012). The use of ineffective methods will lead to additional data gathering and analysis efforts that provide little or no benefit for actual decisions.

The Need for Quantitative Decision Analysis to Guide Data Requirements

Expert judgment, even for scientifically trained professionals, is rife with a variety of errors including overconfidence, flawed recall, influence by irrelevant factors, inconsistency, and flawed subjective inferences. This likely explains why even relatively naïve statistical models outperform expert judgment in a surprising breadth of areas. In a meta-analysis of 90 studies comparing expert judgment to statistical models in such areas as clinical diagnosis, disease prognosis, the failure of small businesses, and the outcomes of sporting events, the statistical models outperformed the human experts in almost all cases (Meehl 1954). Since Meehl's seminal research, further research has only corroborated his findings. Another more recent study used methods that assessed the performance of quantitative methods more

harshly but still found that quantitative methods overwhelmingly outperformed expert judgment (Grove 2000). Further, this more recent meta-analysis concluded quantitative methods' outperformance is unrelated to the source of study (journal vs. other), publication date, as well as across any field considered (Grove 2000). Numerous other studies over a 50 year period have been summarized in meta-studies, which urge the conclusion that statistical methods simply outperform human experts in most areas of judgment test (Trout, Bishop 2002; Ægisdóttir et al 2006).

It has also been shown in industry and space exploration that building quantitative models, including historical regressions and Monte Carlo simulations, improves estimates and decisions (Bailey et al 2000; Simpson et al 2000; Freaner et al 2008). However, the type of quantitative analysis matters. Merely applying additional effort to analysis is no guarantee of improvements. In fact, there appears to be a type of "placebo effect" with analysis – applying more effort in what seems like a structured analysis procedure or in more data gathering improves confidence even when measured performance is not improved or even degraded (Tsai, et al 2008; Heath, et al 1995; Andreassen 1990; Williams et al 2007).

Therefore, we should directly measure the performance of decision analysis methods. The performance of the method (i.e. observable reduction in decision errors, reduced estimation errors, etc.) should be measured, just as the performance of quantitative measures was done in many of the studies. Gathering additional data without measuring their effect on reducing decision error may provide little more than an illusory benefit.

Measuring What Matters

If quantitatively-based decision analysis methods are to be used, they need to be populated with relevant measurements. HDR has observed in many organizations a significant misalignment between current measurement and monitoring efforts and those measures are likely to affect practical decisions[1]. An informative way to determine how aligned measurements are to decisions is to analyze the economic value of information for every uncertain variable in a decision model. When the value of information is computed we may find that even marginal reductions in uncertainty have a significant value.

In short, information has value if the decisions it supports have a chance of being wrong and a cost of being wrong. If a variable is uncertain and has a significant reduction to the chance of a decision error (e.g. investing in a project when the benefits weren't high enough or failing to invest in a project when one should have), then it will tend to have a high information value

In a variety of industries and government agencies, HDR has observed that information which is more likely to inform highly uncertain but important decisions is not sought as often as information which has little chance of improving the final decision. We call this the "measurement inversion." In over 75 major quantitative decision models with well over

[1] It is important to emphasize that we are referring here to applied research where the objective is to improve intervention decisions and provide practical outcomes, as opposed to fundamental research where the objective is purely scientific discovery. However decision-focused applied research may have no less chance of providing a surprise discovery that has important practical use than fundamental research.

5000 variables among them, HDR has observed the measurement inversion in issues as diverse as insurance, environmental policy, logistics, manufacturing, engineering, information security, and pharmaceuticals.

It appears that decision makers and data analysts of all types would measure very different things if they were to compute the information values of uncertain variables. Instead, they appear to measure what they know how to measure with little regard to whether a measurement is likely to reduce uncertainty about which of several decision choices is the best.

Based on initial findings of decision models being constructed for the Intervention Decision Model (IDM) project run by the CGIAR Program on Water, Land and Ecosystems[2], the measurement inversion persists in agricultural and environmental interventions. One intervention being modeled is related to "payment for environmental services" (PES) to reduce sediment run-off to the Sasumua Dam in Kenya. The model contains 78 individual variables related to reservoir capacity, rainfall, adoption of PES by the targeted population, erosion hot spots, water quality, goodwill, and many more.

For each of these variables, an *a priori* probability distribution was estimated and the sensitivity of the decision to that variable was computed. Based on this information, an "Expected Value of Perfect Information" (EVPI) was computed for each variable using standard decision theory methods. The EVPI is the theoretical maximum a measurement would be worth by improving the chance of making a better decision. While perfect information is rarely feasible, the EVPI is an informative way to determine which variables are worth a significant measurement effort.

We discovered that the only variables with significant information values for ascertaining whether the PES scheme is a good investment alternative are the costs of sediment removal and the current average sediment runoff per year. It is highly unlikely that the scientists or decision makers would have focused measurement resources first on these two variables unless we had performed this calculation.

Further explanation of eliciting *a priori* probability distributions and computing the EVPI is widely available in decision theory literature. The particular application of these methods in the Applied Information Economics method is also available in publically available cases. [3]

In short, it is possible – even likely given observations of the measurement inversion – that when stakeholders are asked for their data requirements, they may be unaware of the data that would most influence key policy decisions. Any survey that asks stakeholders for perceived data requirements should employ some test to see if perceived and actual data needs of stakeholders are aligned.

---

[2] http://wle.cgiar.org/

[3] A case study applying the same method to a different problem (IT security at the Department of Veterans Affairs in the US) can be found on the DFID website here: http://r4d.dfid.gov.uk/pdf/outputs/misc_susag/192446_Appendix5Supplement_AppliedInformationEconomicsExample.pdf

Methodology

Survey Population

Prospective survey participants were selected from African agriculture stakeholder groups by senior staff at the CGIAR. These groups varied by organizational size, organizational mission, sources of funding, and geographic location. In order to sample a representative diversity of stakeholder types, the initial contact list was comprised of 11 broad classification groups including: donors, CGIAR management, CGIAR scientists, advanced research institutes, regional organizations, government ministries, national agricultural research systems (NARS), local governments, development non-government organizations (NGOs), African universities, private sector organizations, and farmer associations. Within these groups, the specific role of the survey recipient varied significantly and included research scientists, educators, program managers, vendors, consultants, and analysts. It was assumed that the survey responses would not be significantly influenced by under or over sampling different groups.

Survey Method

Initial email requests to complete the online survey were sent to the entire contact list of 281 individuals; the actual email text is provided in Appendix A. To ensure a reasonable response rate, reminder email requests were sent to recipients who had not completed the survey within about 10 days. A nine question online survey was created using the website SurveyMonkey (http://www.surveymonkey.com) to collect relevant contact information, stakeholder role, and details of how respondents work with data within their respective organizations. The survey questions and format are shown in Appendix B.

Results of the survey were recorded online with progress monitored by CGIAR staff to assess completion rates and indicate the need for reminders. For respondents who affirmed their availability for the follow-up interview (Appendix C), CGIAR staff scheduled a 30 minute conversation with a member of the Hubbard Decision Research (HDR) staff. These interviews were conducted using Skype voice over IP (VOIP) service (http://www.skype.com) and served to clarify, modify, and/or augment the online survey responses in addition to prompting further discussion about data needs related to their work in African agriculture. The interview findings were summarized by HDR staff as text documents with the discussion recorded using Pamela recording software (http://www.pamela.biz) for optional review after the interview was over.

Assessing Responses

Once all of the surveys and interviews were completed, the online responses and phone interview results were compiled into a central spreadsheet for analysis. A text frequency analysis was performed to find common responses within individual questions and to guide further analysis of relevant trends. Results were also sorted by group to identify variations by stakeholder type. We also assessed whether perceived data needs are aligned with actual decisions. In other words, we were investigating the existence and extent of the previously-

described measurement inversion among these subjects. Lacking a fully developed stochastic model for each of the respondents in the survey, we were unable to compute information values and compare them to perceived data requirements. However, we did ask questions that subjectively assessed alignment between perception of needs and actual requirements. We asked stakeholders what data they spent their time gathering (Q2), if they could identify specific decisions that their data support (Q5), what they believe their data needs to be (Q6), and where they felt they had significant uncertainties (Q7*i*). We used these responses as indicators of what we call decision awareness, needs/uncertainty/effort alignment and, finally decision alignment – defined as follows:

- *Decision Awareness*: Each response was judged whether it indicated an awareness of actual decisions being supported by the research. These assessments were made generously and any indication that the data supported something that sounded like a decision was judged a "yes." If they said the data was being used to inform a policy on project approval, allocation of resources to farmers, or anything else that sounded like a decision, it was judged a "yes." Some stakeholders may not necessarily be involved in policy or intervention decisions so this criterion only applies to those that are.
- *Needs/Uncertainty/Effort Alignment*: Each response was also judged as to whether there was any alignment between the perceived, self-reported data needs and which variables they believed had the most uncertainty (highly uncertain variables are more likely to have high information values). Again, the assessment was made generously and took any similarities between stated needs and stated uncertainties as a "yes." Also, we attempted to determine if the current data gathering efforts were consistent with either their perceived uncertainties or their perceived data needs.
- *Decision Alignment*: For stakeholders who are involved in or at least supporting intervention and policy decisions, each respondent that was judged a "yes" on the previous two items was considered to be a candidate for having perceived data needs that are well aligned with decisions. Since this doesn't assess how they are using data to ultimately make decisions, a "no" is a strong indicator of a lack of decision alignment but a "yes" is not proof that there is decision alignment.

Results and Analysis

Response Statistics

An initial email requesting participation in the online survey was sent to 281 individuals with 110 completing and submitting their answers (39% response rate). Of the online survey respondents, 58 were interviewed through Skype VOIP calls (53% of online participants). Comprehensive survey response statistics, organized by survey phase and stakeholder group, are recorded in Table 1 below.

Table 1: Comprehensive survey response statistics

Stakeholder Group	Initial Email	Online Survey		Phone Interview	
	Number Sent	Number of Respondents	Response Rate	Number of Respondents	Response Rate
Donors	30	14	46.7%	9	64.3%
CGIAR Management	16	4	25.0%	1	25.0%
CGIAR Scientists	25	11	44.0%	6	54.5%
Advanced Research Institutes	25	9	36.0%	7	77.8%
Regional Organizations	29	7	24.1%	4	57.1%
Government Ministries	16	9	56.3%	4	44.4%
Government NARS	22	12	54.5%	6	50.0%
Governments Local	9	3	33.3%	1	33.3%
Development NGOs	22	12	54.5%	5	41.7%
African Universities	47	12	25.5%	4	33.3%
Private Sector	23	11	47.8%	8	72.7%
Farmer Associations	17	6	35.3%	3	50.0%
Totals:	**281**	**110**	**39.2%**	**58**	**52.7%**

The online survey response rate ranged between 24.1 and 56.3% ensuring a reasonable diversity of participants consistent with the initial list. Over half of the online participants were contacted for the phone interview with at least one person from each of the 11 groups interviewed.

Initial Text Frequency Analysis

Online survey responses to each question were analyzed using text frequency analysis which provided an efficient summary of critical subject areas. This method, while rapid and powerful, can also lead to biased results due to incomplete context, a multitude of potential unique responses, and variations in wording. Nonetheless, it can still provide important indications of data priorities common among broad groups of stakeholders and guide further analysis.

This particular analysis is more appropriate for certain types of questions where the number of possible responses is limited to a finite or closed set. We can further improve the result by manually processing the responses to remove superfluous words (articles, prepositions, "data," "agriculture," etc.) and enforce consistency in wording (e.g. singular vs. plural). This approach was particularly attractive for analyzing the results from **Question Q2** regarding the subject areas of highest priority. The results of this analysis are shown in Appendix Table D1 for the 20 most common keywords.

Soil data appear to be the leading priority from the text frequency analysis, however we also observe the importance of market data, especially if we conservatively aggregate related responses. For example, adding counts from "market," "input" (price/purchase), and "consumer" we have an aggregated count of 77, which outpaces the sum of "soil" and "land"

at 69. There is obviously some subjectivity in this approach, but unmistakable data priority trends persist irrespective of how the results are consolidated. These two aggregated categories reveal stakeholder interest in the most fundamental agricultural resource (soil) and arguably the primary indicator of agricultural sustainability (markets).

Question Q3 further inquired about the types of data that stakeholders commonly collect in support of improved practices in African agriculture. This is the first opportunity to examine consistency between subject areas stated as most important to stakeholders (Q2) and their actual practice of collecting data (Q3). Again, there was a wide variety of specific responses and accompanying challenges in distilling these replies into appropriate categories. However, in this case we are looking for simple, observable trends and the expectation that the responses from Questions Q2 and Q3 should be largely consistent. Results for data collection priorities are shown in Appendix Table D2.

Notably, "soil" was the dominant concern as both a broadly defined subject area and data type, and as such topped both lists. Overall, there was good correlation between the two lists, especially if we allow for some expected variation in response wording. For example, "production" and "productivity" collectively were nearly invariant with 25 total counts in Q2 and 26 in Q3. Additionally, the term "biodiversity" is prominent in response to Q2, but garnered only one mention from all respondents in Q3. This disconnect could arise from the fact that "biodiversity" is used as a collective term indicating a broad subject area that is quantified through a number of more data-specific terms such as "yield," "crop" and "fertility" which all appear prominently in responses to Q3. Similarly, we see corroborated interest in "market" trends (indicated in Q3) as shown by high ranking data collection in "yield," "crop", and "price" in addition to the expected "market."

Interestingly, there is an apparent disconnect between the broadly stated interest in "poverty" as a subject area in Q2 and the responses collected in Q3. Indeed, "poverty" only tallied four mentions in the latter, and it is difficult to identify terms that immediately relate specifically to poverty concerns from the responses to Q3, with the possible exceptions of "income" and "economic." However, these two terms could easily extend to other areas beyond poverty. Based on the text frequency analysis, this is probably the most obvious discrepancy between the two lists, which are otherwise highly consistent with each other indicating a harmony between stated areas of interest and data collection habits.

In **Question Q4**, stakeholders were asked to identify their most important sources of data, which provides some insight into how data are currently being sought, shared, and consumed. It can also identify existing sources that are highly regarded by the stakeholder community and potentially indicate areas where improvements could be made. As with other questions, there is likely some response bias due to the provided examples of data sources. With that in mind, we can examine the results first using two word phrase frequencies to pick out some of the recommended options as shown in Appendix Table D3.

The example responses provided for Q4 were: "direct measurement," "literature or web searches of documents," "on-line databases," and "organization's own database." We see a clear trend of mimicking these examples in the actual replies, although the frequency of each differed. There is a strong preference for "direct measurement" using the two word frequency analysis, perhaps motivated by the need for research-specific data that is not currently collected, available, and/or reliable enough to adequately inform stakeholders. The

prevalence of "own databases" and "data collection" further highlights a tendency of data collection and dissemination to be contained within particular organizations. However, there is also a trend of accessing online tools to obtain necessary data as indicated by "on-line databases" and "web searches" which is consistent with outcomes seen in Question Q8 related to preferred data formats and channels.

Individual keyword analysis of the same responses, summarized in Appendix Table D4, revealed other trends. For example, a surprisingly common response indicated that the Food and Agriculture Organization (FAO), and the FAOSTAT database in particular, is an invaluable resource for obtaining data critical to African agricultural policies and decisions. The relatively high frequency of this response is especially interesting because it was not mentioned in the survey question text and yet was a popular response. The utility of FAOSTAT was also corroborated during the phone discussions with stakeholders as well as the general need for expanding other similar centralized databases that store and share reliable information across many countries in Africa.

Other highly cited terms from the list include "field," "survey," and "national," all of which were original responses (falling outside of the suggested list) by stakeholders. In particular, "field" and "national" reflect issues related to geographic scale. In some cases, fine details of production, soil condition, prices, and climate are necessary to understand how policies and related interventions might impact a specific region. Likewise, other policies and interventions may require more aggregated forms of information on the level of one or several countries. In the latter case, data provided on a national level (or by governments themselves) would be more beneficial. Overall, stakeholders desired some flexibility in how data was aggregated depending on their specific area or focus. This explains the seemingly discordant responses related to data resolution and scale. This trend was especially evident during the interview phase where the desired level of data granularity varied widely among stakeholders.

Because data should ultimately be used to inform specific decisions that have a positive effect on sustainable agricultural practices in Africa, we asked stakeholders in **Question Q5** what professional decisions they make using available data. We found an interesting set of responses to this question, many of which indicated that many researchers, rightly or wrongly, did not perceive themselves as supporting intervention decisions. This was especially true with researchers less directly connected to farming practices.

To more efficiently analyze results from Q5 (decisions), Q6 (data needs), and Q7 (uncertainty), we further aggregated stakeholder type (from the original 11 to 6) and identified representative response categories. The stakeholder types were aggregated and defined as follows:

- CGIAR/Advanced Institutes: CGIAR Management, CGIAR Scientists, and Advanced Research Institutes
- National and Regional: Regional Organizations with Government Ministries and Government Local
- Grassroots/Local: Development NGOs and Farmer Associations
- Local Researchers: African Universities and Government NARS
- Donors: Donors (unchanged)
- Private Sector: Private Sector (unchanged)

Responses to Q5 were classified into five response categories: research, analysis and publication; funding decisions and prioritizing research; policy recommendations; and evaluating interventions and programs. While some stakeholders responded with numerous answers, with few exceptions we were able to group them into these five categories which highlighted the main decisions they address. We do note, however, that in some cases the responses stretched the usual definition of a "decision," but we were at least able to sort their answers among this list.

Of the responses we cataloged, those that fell under the first category of "research, analysis, and publication" were not surprisingly the least connected to direct decision making. Obviously there is a critical role for research activities and high-level analysis, however these stakeholders highlighted the research itself is the primary decision they support. The remaining four categories seem to have a stronger link to actual decisions, especially in the areas of farming recommendations and funding decisions. Examination of the responses by stakeholder type helps to clarify why identifying decisions can be difficult for some. Depending on their role, some stakeholders hold positions that are more closely tied to decisions. However, we also note that data collection, measurements and analysis should always be motived by improving understanding (i.e. reducing uncertainty in a critical area), which improves the likelihood of making a better decision at some level.

Table 2: Individual text frequency analysis for Question Q5 (professional decisions made):

Aggregated Stakeholder Type	Research, Analysis, Publication	Funding Decisions, Prioritizing Research	Farming Recommendations	Policy Recommendations	Evaluating Interventions and Programs	Total Stakeholders in Group
Private Sector	10.0%	20.0%	40.0%	20.0%	10.0%	**10**
CGIAR/Advanced Institutes	37.5%	20.8%	0.0%	16.7%	25.0%	**24**
National and Regional	23.5%	17.6%	11.8%	29.4%	17.6%	**17**
Grassroots/Local	37.5%	18.8%	18.8%	12.5%	12.5%	**16**
Donors	6.7%	40.0%	0.0%	13.3%	40.0%	**15**
Local Researchers	26.1%	17.4%	39.1%	13.0%	4.3%	**23**
Total Responses	**27**	**23**	**18**	**18**	**19**	**105**

Some stakeholders stated explicitly in their responses that their position doesn't necessarily involve "decisions" *per se* as they are largely facilitating others' ability to make better decisions through advocacy, resource allocation, research, analysis, or data collection and sharing. Since many of the stakeholders interviewed work in the area of policy and research, they may be somewhat removed from decisions that can be immediately tied to farming practices in the field. Nonetheless, we still found that a surprisingly large number of respondents had a difficult time identifying or describing explicit decisions that they or their

organization routinely makes. In some cases, respondents provided decisions that were not related to interventions (e.g., decisions about research priorities, getting published, etc.). This was especially the case among researchers who are used to compiling information and analysis in the form of databases and manuscripts. Those stakeholders closer to guiding policy, funding projects, and defining budgets seemed to have an easier time identifying data-driven decisions associated with their roles.

If we instead distinguish responses to Q5 as either reasonably supporting decisions or not, we find several interesting trends, including some that vary by stakeholder type:

- Overall, slightly less than half of respondents (46%) could specify a decision of any kind. This included decisions purely related to research.
- Development NGOs and Farmer Associations seemed to have the most difficult time articulating a decision (72%), followed closely by Donors (64%).
- African Universities and Government NARS fared best among their peers but about one third (33%) failed to specify a decision of any kind.

In **Question Q6** regarding stakeholders' perceived data needs, we used the same procedure of aggregating stakeholder types and sorting responses among common categories as was used for Q5. The responses to Q6 were grouped into the following six areas: increased reliability, access to data, more current data, different data, higher resolution, and lower resolution. "Increased reliability" generally refers to responses which mentioned that data was currently accessible but unreliable or inaccurate. "Access to data" means that the data is available somewhere but inaccessible to the stakeholder. "More current data" indicates stakeholder data requirements where timely data is deemed valuable. "Different data" refers to data outside of what the stakeholder primarily collected or utilized, or that which they believed did not exist in any reliable format or accuracy. "Higher resolution" refers to a need for higher frequency, location specific, or disaggregation of some current data series. "Lower resolution" refers to data that the stakeholder desired on a higher political organizational level (i.e. national) but that currently existed at least partially on a smaller scale. Additionally, there were eight people who either declined to answer or were non-respondents who were omitted from Table 3 below. Lastly, only two stakeholders (1.8% of our sample) indicated their data requirements are currently met. They too were omitted.

The most obvious finding from the aggregated analysis was that almost every stakeholder who responded required additional or more accurate data. While the requirements varied significantly, there were a few noteworthy observations. The most frequently identified stated data requirement was for "different data" and many of these (19 out of 45) mentioned requirements related to socio-economic or market data. Another heavily identified category, "higher resolution," further held requirements for data at the farm level (11 out of 34).

Table 3: Individual text frequency analysis for Question Q6 (perceived data needs):

Aggregated Stakeholder Groups	Increased Reliability	Access to Data	More Current Data	Different Data	Higher Resolution	Lower Resolution	Number of Stated Data Requirements	Number of Stakeholders
Private Sector	4	3	1	2	2	1	**13**	**11**
CGIAR/Advanced Institutes	7	5	2	13	15	1	**43**	**23**
National and Regional	3	4	6	9	2	2	**26**	**16**
Grassroots/Local	4	5	2	7	4	0	**22**	**14**
Donors	8	1	1	7	7	3	**27**	**14**
Local Researchers	5	11	4	7	4	2	**33**	**22**
Total	**31**	**29**	**16**	**45**	**34**	**9**	**164**	**100**

All data carry some level of uncertainty, and it is often data with large uncertainty that merit further measurement when considering a decision. To better characterize this effect, stakeholders were asked in **Question Q7** to identify data they perceived to be (*i*) most uncertain and (*ii*) least uncertain. Using the aggregated stakeholder groups, we organized the most uncertain responses (from Q7*i*) into broad categories including: water, land, and ecosystems (physical data); socio-economic and market factors; farms and farming practices (agro-inputs, productivity, and seeds); policy; climate; and research methodology (data collection practices). The results are given in Table 4.

Collectively, the data perceived to be most uncertain by the stakeholders were related to farming and markets. These are also the data categories that are arguably the most important for agricultural sustainability. Although we were not able to determine in a simple survey the value of information for these data, we can say that highly uncertain variables are more likely to warrant greater attention in relation to improving decision outcomes. While most stakeholder types accumulated similar totals in listing uncertain data, CGIAR/Advanced Institutes respondents were much more critical of data, including the area of water, land, and ecosystems. Perhaps surprising given recent attention and unpredictable reputation, climate data was not considered as a top three uncertainty from this list.

We further used the results from Q7*i* in subsequent analysis to determine the degree of alignment among uncertainty, gathering efforts, and stated needs.

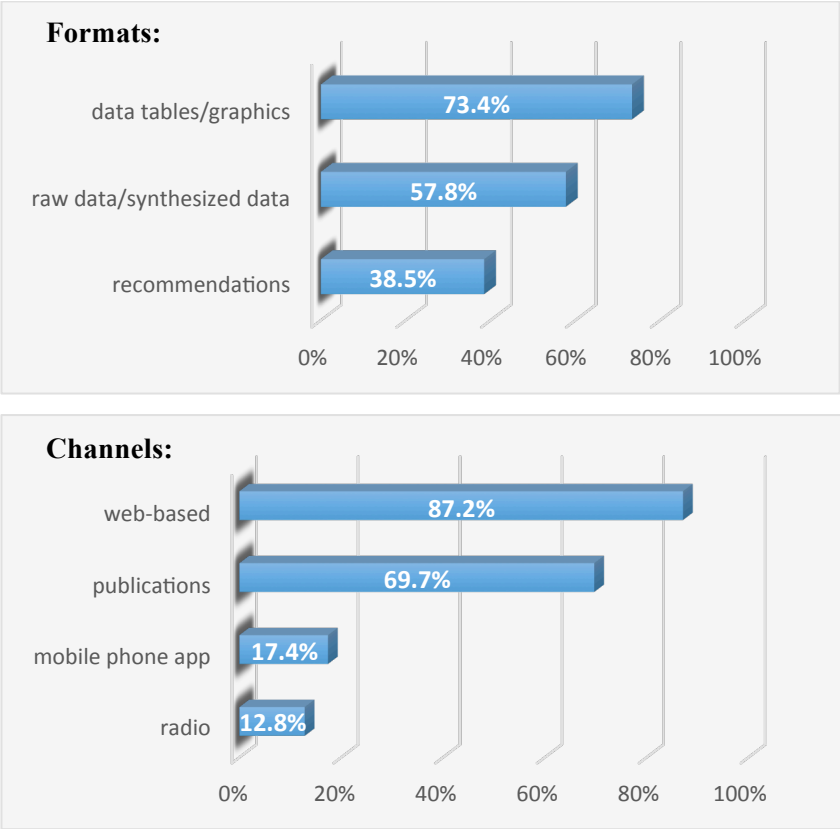
Table 4: Individual text frequency analysis for Question Q7*i* (most uncertain data):

Aggregated Stakeholder Groups	Water, Land, Ecosystems	Socio-economic, Market Factors	Farms and Farming Practices	Policy	Climate	Research Methodology	Number of Stakeholders in Group
Private Sector	3	5	3	2	2	2	**11**
CGIAR/Advanced Institutes	9	10	10	1	3	2	**24**
National and Regional	2	5	4	5	4	3	**19**
Grassroots/Local	4	6	7	2	1	1	**18**
Donors	3	8	9	0	2	1	**14**
Local Researchers	4	3	6	4	3	2	**24**
Total	**25**	**37**	**39**	**14**	**15**	**11**	**110**

Question Q8 asked stakeholders which formats and channels they would prefer data to be made available from a closed list of 8 possibilities (4 channels, 3 formats, and "other"), however they could (and did) indicate specific options under the "other" option. In this case, the analysis is more straightforward and shown in Figure 1 as percentages of respondents that selected each. In this case, respondents could select as many options as they liked.

Unsurprisingly, web-based data channels were deemed important by over 87% of respondents consistent with the ubiquity of internet data sharing worldwide. In fact, most discussions of dissemination and data sharing involved internet-based options, even at the smallholder level in remote areas. However, web-based formats were not exempt from criticism as recorded in the phone interview phase. There were many recommendations for how to format, compile, store, and process data that are currently available on the web. In many cases stakeholders are actually satisfied (or even overwhelmed) with the quantity of data available, but displeased with the current format or organization. The high percentage of respondents preferring data tables and graphics as a format suggests that presentation matters; our in-depth discussions with individuals support that finding. There is a critical need for standardized data processing and data management practices that transform raw data into readily useable formats.

Figure 1: Results from survey Question Q8 (preferred data formats and channels)



The four response options with selection rates below 50% are likely more relevant to farmer and field-level personnel who would benefit most from farming recommendations, mobile phone apps, and radio. Responses of "other" (14.7% of respondents) varied widely and included suggestions such as TV documentaries, "field days," low-cost publications, centralized databases, satellite/geospatial data, artistic productions (drama, song), and statements of reliability.

Assessing Alignment of Data Needs with Decisions

Additional analysis was conducted to investigate alignment among actual data collection effort, perceived data needs, and data uncertainties. Table 5 (below) is an extended version of Appendix Table D1 which further considers how frequently particular types of data were mentioned as a current focus of effort and as a source of uncertainty (Questions Q6, Q2, and Q7*i*, respectively). Agreement or disagreement regarding how often data types are mentioned is used as an initial indicator of alignment while considering that a keyword frequency analysis is a relatively blunt approach ignoring contextual cues and is alone an insufficient indication of a real measurement inversion (which requires quantitative models to compute information values). This information merely directed further investigation of survey responses and phone interview notes.

Table 5: Evaluating consistency among data gathering, needs and uncertainties

	Data Gathering Effort (Q2)		Perceived Needs (Q6)		Perceived Uncertainties (Q7*i*)	
Keyword	Count	Rank	Count	Rank	Count	Rank
soil	45	1	16	1	12	1
market	40	2	11	3	11	2
poverty	33	3	3	12	1	16
biodiversity	27	4	2	16	1	16
water	27	4	8	8	3	12
input	25	6	5	9	4	8
livestock	23	7	3	12	4	8
climate	19	8	11	3	11	2
crop	16	9	12	2	9	4
land	14	10	10	6	7	6
production	13	11	11	3	5	7
productivity	12	12	1	17	4	8
consumer	12	12	0	20	0	20
management	11	14	3	12	1	16
policy	10	15	3	12	4	8
food	9	16	1	17	2	15
forest	9	16	5	9	3	12
price	8	18	9	7	9	4
household	8	18	4	11	3	12
technology	8	18	1	17	1	16

The text frequency analysis led to further scrutiny of each response to assess whether data effort, stated needs, and uncertainties were aligned. Each response was individually evaluated where we attempted to determine the level of alignment among effort, needs and uncertainties. This was combined with the initial text analysis leading to the following key observations:

- Of 110 responses, 40 indicated an alignment between perceived data needs (Q6) and stated uncertainties (Q7*i*) – 36% of respondents stated data needs that were consistent with their stated uncertainties.
- Even when uncertainties and perceived needs are aligned, they may actually be spending current effort in a different way. Only 15% showed that perceived needs, uncertainties, and data gathering efforts are aligned.
- There was broad alignment among effort, perceived needs, and uncertainties for soil data. In other words, soil was the most frequently cited uncertainty, the most frequently stated perceived need, and the most frequently stated focus of current effort. Market data showed similar high priority and alignment.
- Climate data is frequently cited as both needed and satisfying an uncertainty, but relative to other data types, is less frequently cited as a focus of current effort. To a lesser degree, this applies to price data as well.
- Biodiversity and poverty data are frequently cited as a focus of effort but infrequently cited as a perceived need or uncertainty.

Lack of alignment among effort, perceived needs, and perceived uncertainties could be an indicator of a measurement inversion, however there are other possible explanations as well. For example, it is possible that respondents were aware that an uncertain variable may have little impact on a decision and then would correctly not list those among data needs. It is also possible that current data gathering efforts do not focus on stated uncertainties due to other constraints such as data access and resources.

However, our experience with other decision modeling problems suggests that the above possibilities are unlikely to explain these results. Based upon previous findings, we know that, in general, researchers tend to exclude highly uncertain variables from their analysis that would actually tend to inform decisions. Further, they tend not to spend time gathering information which they believe to be less valuable than alternatives. In that case, some potentially important data may be prematurely excluded. Our experience with the prevalence of the measurement inversion in all types of decisions shows that, lacking explicit calculations of information values, choosing the right measurements by chance is highly unlikely. In addition, a recent DFID-commissioned global review of monitoring initiatives also found little evidence for impact on real world decisions (Shepherd et al 2013).

Given the incomplete context of the survey responses, any similarities between effort, uncertainties, and data needs were judged to be evidence of alignment in this study. We did not attempt to determine whether the stated data needs or efforts had any relation to stated decisions, rather we simply noted that efforts, needs, and uncertainties had at least some observed overlap. Had these areas been judged more strictly, the result would have shown very few people having alignment among data needs, uncertainties, efforts, and an awareness of the decisions being supported.

Insights

The survey was revealing in many ways, especially when viewed in context with related work completed in the recent IDM project and past work by HDR.

- Soil and markets were subjects identified both as areas where time is spent gathering the data and where there is a perceived data need. There was moderately good agreement on most subjects between time spent and perceived data needs. In other words, if there was a perceived need, that is where effort is focused. However we noticed a mismatch between time spent and perceived data needs in topics related to poverty. There appears to be a lot of time spent gathering poverty data and little perceived uncertainty about it.
- Despite the prominence of web-based data sharing resources, stakeholders tend to pursue and use their own, newly collected data first. This may indicate inadequate data sharing practices within the larger community or significant gaps in data that meet stakeholders' needs.
- There is a strong desire for comprehensive centralized databases that have reliable and useable data. The FAOSTAT database was cited by many as a much-used source

of information, however a significant opportunity exists to improve data management and sharing practices broadly. Often, data from various projects are lost once funding ceases.

- Stakeholders generally desired improvements in data (more reliable, more current, etc.) while fewer indicated the need for "different data" from what they access currently. Less aggregated data ("higher resolution") is clearly preferred over more aggregated ("lower resolution"), but in some cases there was a desire for both suggesting that databases should allow flexibility in how data is organized and processed.
- Among only the stakeholders who support policy and intervention decisions, we found that a surprising number of stakeholders had difficulty identifying decisions they commonly support through existing data. There was a moderate level of seeking data merely for the sake of having it. Indeed, the experience with IDM has revealed that even in a facilitated workshop, researchers usually took more than one day to identify a specific decision to model.
- Prior research has shown the benefits of constructing quantitative decision models. Even relatively naïve quantitative models outperform expert intuition in a surprising variety of tasks. These models require accurate data measurements as inputs, however only a few such parameters are likely to have significant information value.
- A disconnect between perceptions of data needs, efforts in gathering data, and stated uncertainties are indictors of a measurement inversion among the respondents to the survey, just as there is among analysts in other fields. This presents a significant opportunity to improve current practices in agricultural decisions.
- There may be measurement inversions (i.e. relative effort vs. value of information) related to poverty, biodiversity, climate, prices, water use, and productivity. And, although this keyword analysis would not show this, it is possible that the highest value measurements include variables not mentioned by any of the respondents.

Recommendations

Given the findings of this survey, we make the following recommendations:

- In general, there needs to be better stewardship of data. This includes improved collection methods, formatting, and sharing practices. These improvements should be a requirement tied to project funding to ensure that resulting data is adequately utilized and stored.
- There is a persistent appeal for centralized data, readily available on the web, which can be organized geographically in a GIS system. This may be a significant effort, but if the efforts are prioritized by information values, then it will likely have high impact.
- Awareness of the key decisions and what data is needed to support them should be widespread and routine. Clearly, some researchers will be more concerned with fundamental research than applied research and perhaps there are cases where sponsors recognize that. But in cases where researchers are interested in improving development outcomes – or where their sponsors believe they should be – the researchers should be consistently apprised of what data they should be gathering and how it informs decisions. The habit of collecting data for the sake of having data is a practice that should be discouraged in view of limited resources.
- Key decisions should be modeled quantitatively and information gathering requirements should be based on information values rather than routine, intuition, or purely subjective means. This will inform decision makers how to spend their limited resources measuring only the most important information or variables.
- At least some researchers or at least some or perhaps most of their donors see research as supporting outcomes. At least in these cases, researchers should rely less on intuition and judgment in regards to determining research priorities and, instead, should make such decisions based on computed information values in support of intervention decisions. In order to do this, they will need to make greater use of decision analysis techniques that allow them to best utilize uncertain or incomplete data (e.g. Monte Carlo simulation, Bayesian methods). This may help optimize the use of existing data for improving specific development decisions.

Conclusions

Data gathering should be easier and more economical than it has ever been, but developments in information technology have not been matched by impacts on improving policy decisions. It is possible that there are important differences between what researchers say they need to collect or use and what is most likely to impact decisions. Stakeholders expressed a strong desire for accurate, timely, and comprehensive information related to the status of African agriculture and usually had a good idea of where they wanted to see improvements. While the collective volume of data is already substantial in this area, the quality and completeness of the data is often lacking. Further, data may be difficult to locate and not be readily useable.

This presents a persistent challenge for stakeholders at all levels to make proper assessments and ultimately execute decisions that have the highest impact for improving food security on the continent. Lacking a massive influx of resources, this problem needs to be addressed by incrementally improving data management practices and optimizing decisions in the face of incomplete and imperfect data. This survey shows that stakeholders are aware of these limitations and are supportive of improvements, but may not be cognizant of ways in which they can improve current practices given access to data they already have. For example, a better connection should be made between data measurement practices and decisions these data actually support. Moreover, by applying relatively basic principles of decision analysis, stakeholders can focus their attention on collecting data that has the most impact on measurably improving decisions.

The evidence of a disconnect between measurement efforts and the value of information to decision makers is not conclusive from a survey of this type. Nonetheless, the results are certainly consistent with prior observations in other areas of a misalignment between measurement efforts and what would actually improve decisions. Indeed, given the observed prevalence of the measurement inversion in other areas, it would be surprising if agriculture were somehow immune to this effect; nothing observed in this study would indicate otherwise.

The most important measurement is the value of measurements. Computing the economic value of a measurement tells us whether to measure it and how much to invest in measuring it. Likewise, the most important broad decision for stakeholders to make is how they are going to execute decisions based on the available data and evidence. Deliberately-constructed, quantitative decision models based on optimal measurements is the most reliable way to improve decisions. As long as researchers are faced with challenges that greatly exceed their resources, moving toward a decision-focused method of prioritizing measurements should be a continuing goal.

References

Ægisdóttir et al "The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction" *THE COUNSELING PSYCHOLOGIST*, Vol. 34 No. 3, pp 341-382, May 2006

Andreassen, P. "Judgmental extrapolation and market overreaction: On the use and disuse of news" *Journal of Behavioral Decision Making*, vol. 3 iss. 3, pp 153-174, Jul/Sep (1990)

Bailey W., Couet B., Lamb F., Simpson G., Rose P., "Taking Calculated Risks" *Oilfield Review*, Autumn (2000).

Besemer et al " Review of existing agricultural data management practices, lifecycles and workflows". A Report for INFRA-2011-1.2.2. Data infrastructures for e-Science, 3 Oct (2012)

Freaner, C.W., Bitten, R.E., Bearden, D.A. Emmons, D.L. "An Assessment of the Inherent Optimism in Early Conceptual Designs and Its Effect on Cost and Schedule Growth." Paper presented at the Space Systems Cost Analysis Group/Cost Analysis and Forecasting/European Aerospace Cost Engineering Working Group 2008 Joint International Conference, European Space Research and Technology Centre, Noordwijk, The Netherlands, May 15-16. European Space Agency, Paris, France. (2008)Grove, W. M et al. "Clinical versus mechanical prediction: a meta-analysis."*Psychological Assessment* vol. 12, no. 1 (2000), 19-30.

Heath C., Gonzalez R. "Interaction with Others Increases Decision Confidence but Not Decision Quality: Evidence against Information Collection Views of Interactive Decision Making" *Organizational Behavior and Human Decision Processes,* vol. 61, no. 3, (1995) 305-326

Hubbard D.W. *How to measure anything: finding the value of "intangibles" in business.* 2nd Edition. (2010) Hoboken, New Jersey: John Wiley & Sons, Inc.

Meehl, Paul E. *Clinical Versus Statistical Prediction; A Theoretical Analysis and a Review of the Evidence*. Minneapolis: University of Minnesota Press (1954).

Pica-Ciamarra, U.; Baker, D.; Morgan, N."Core livestock data and indicators: results of a stakeholder survey" A paper of the World Bank, FAO, ILRI Livestock Data Innovation in Africa Project, Washington D.C., May (2012)

Raitzer, D.A., Maredia, M.K., "Analysis of Agricultural Research Investment Priorities for Sustainable Poverty Reduction in Southeast Asia" *Food Policy* vol. 37 (2012) 412-426.

Shepherd, K.D., Farrow, A., Ringler, C., Gassner, A., Jarvis, D., "Review of the Evidence on Indicators, Metrics and Monitoring Systems" World Agroforestry Centre (ICRAF), Nairobi, Kenya (2013).

Simpson G.S., Lamb F.E., Finch J.H., Dinnie N.C. "The Application of Probabilistic and Qualitative Methods to Asset Management Decision Making" *Society of Petroleum Engineers* (2000).

Trout, J. D. and Michael A. Bishop. "50 Years of Successful Predictive Modeling Should Be Enough: Lessons for Philosophy of Science." *Philosophy of Science (Proceedings)* 68 (2002): S197-S208.

Tsai C., Klayman J., Hastie R. "Effects of amount of information on judgment accuracy and confidence" *Org. Behavior and Human Decision Processes,* vol. 107, no. 2 (2008), 97-105.

Williams M. Dennis A., Stam A., Aronson J. "The impact of DSS use and information load on errors and decision quality" *European Journal of Operational Research,* vol. 176, no. 1 (2007) pp 468-81.

Appendix A: Email request to participate in the online survey

The initial request to participate in the survey was sent by email with the following text:

> Dear Sir/Madam,
>
> I am writing to invite you to participate in a stakeholder consultation (survey) on which subsets of reliable agricultural and related data would be of priority in terms of facilitating accessibility to African farmers, researchers and policymakers, with the aim of enhancing food security.
>
> The World Agroforestry Centre (ICRAF) has been commissioned by the UK Department for International Development (DFID) to conduct the survey, which is being undertaken with assistance from Hubbard Decision Research, a professional decision analysis consulting firm that specialises in the use of Applied Information Economics for highly complex and uncertain decisions.
>
> The results of the survey will be presented at a G8 organised international conference on Open Data to be held in Washington DC from 28th to 30th April 2013. The topline goal for the initiative is to obtain commitment and action from nations and relevant stakeholders to promote policies and invest in projects that open access to publicly funded global agriculturally relevant data streams, making such data readily accessible to users in Africa and world-wide, and ultimately supporting a sustainable increase in food security in developed and developing countries. In addition, the conference also aims to 'create an opportunity for the establishment of a global platform to make reliable agricultural and related information available to African farmers, researchers and policymakers, taking into account existing agricultural data systems.'
>
> The link to the web-based survey, which should take you less than 15 Minutes to complete, is: http://www.surveymonkey.com/s/SQPZKMW
> We are also requesting respondents to participate in a 30-minute follow up call, if possible, to further clarify and enrich your responses.
>
> We do hope that you will contribute to this important initiative and look forward to seeing your responses on line. The deadline for submission is 15 April but we would hope to obtain your responses as soon as possible to allow time for a follow up call.
>
> Finally if you know of anyone else who you think may provide a good contribution to this survey, then I would be grateful if you could email their name, organisation, and email address to Grace Muinga (G.Muinga@cgiar.org).
>
> Thank you very much indeed for your time.

Appendix B: Online survey questions

The online survey was conducted using the online survey website Survey Monkey with the following questions. All responses were short answer save Q8 which had specific choices.

Q1: Please enter your contact information.

Name:
Company:
City/Town:
State/Province:
Country:
Email Address:
Phone Number:

Q2: In your organization's role as a stakeholder in African agriculture, in what broad subject areas do you currently spend the most time collecting, compiling or looking for data? (e.g., soils, water, markets, poverty, consumer preferences, biodiversity, livestock genetics, input purchases, etc).

Q3: What types of data (variables, items of information) do you collect or compile?

Q4: What are your current main sources of data or information? (e.g., direct measurement, literature or web searches of documents, on-line databases, organization's own databases, etc.)

Q5: What professional decisions do you make explicitly based on these data? These can include decisions that are outside of those your organization already supports. (Please provide at least three)

Q6: What data would you like to have but which is not currently available or accessible that could help most improve those decisions?

Q7: Thinking about decisions you support, which variables do you think: (i) have the most uncertainty (or that you are least sure about), and (ii) which are you most certain about?

Q8: Through what channels and what formats would you like data to be made available?

Answer Choices:

channel: web-based
channel: mobile phone app
channel: publications
channel: radio
format: raw data/synthesized data
format: data tables/graphics
format: recommendations
Other (please specify)

Q9: Please verify your availability for the follow-up discussion interview (about 30 minutes). This is a necessary component of the interview process. Someone will contact you to set up an agreeable time.

Appendix C: Email request for a follow-up interview

If the survey participant indicated interest in a follow-up discussion (from their response to Q9 in the online survey) they received the following email:

> Thank you very much for responding to the questionnaire.
>
> Kindly provide us with your availability in terms of date and time for the 30 minute interview.
> It will be helpful if you could also provide us with a Skype name if we can contact you by Skype, otherwise please confirm the phone number you will be reachable on at the appointment time.
>
> You will be contacted by either Aaron Clapp or Nathan DauSchmidt from Hubbard Decision Research at your preferred appointment time.

Appendix D: Text frequency analysis results

Table D1: Text frequency analysis for survey question Q2 (broad subject areas where the most time is spent collecting/compiling data):

Rank	Keyword	Count
1	soil	45
2	market	40
3	poverty	33
4	biodiversity	27
4	water	27
6	input	25
7	livestock	23
8	climate	19
9	crop	16
10	land	14

Rank	Keyword	Count
11	production	13
12	productivity	12
12	consumer	12
14	management	11
15	policy	10
16	food	9
16	forest	9
18	price	8
18	household	8
18	technology	8

Table D2: Text frequency analysis for survey question Q3 (types of data collected):

Rank	Keyword	Count
1	soil	32
2	yield	29
3	crop	28
4	price	20
5	production	19
6	market	13
6	farm	13
6	land	13
9	livestock	12
10	use	11

Rank	Keyword	Count
11	income	10
11	water	10
13	fertility	9
14	products	8
14	irrigation	8
14	economic	8
14	climate	8
18	development	7
18	input	7
18	productivity	7

Table D3: Two word text frequency analysis for survey Question Q4 (sources of data):

Rank	Keyword Pairs	Count
1	direct measurement	48
2	on-line databases	20
3	web searches	19
4	own databases	14
5	data collection	13

Table D4: Individual text frequency analysis for Question Q4 (sources of data):

Rank	Keywords	Count
1	direct	56
2	database	52
2	measurement	52
4	literature	42
5	web	30
6	own	28
7	FAO/FAOSTAT	18
8	field	12
8	survey	12
10	national	9