Review of the Evidence on Indicators, Metrics and Monitoring Systems

Commissioned by the UK Department for International Development (DFID)

Conducted by the CGIAR Program on Water, Land & Ecosystems

Coordinated by the World Agroforestry Centre (ICRAF)

Authors: Keith D Shepherd[1], Andrew Farrow[2], Claudia Ringler[3], Anja Gassner[1], Devra Jarvis[4]

[1] World Agroforestry Centre (ICRAF)

[2] Consultant for World Agroforestry Centre (ICRAF)

[3] International Food Policy Research Institute (IFPRI)

[4] Bioversity International

Contents

Acronyms and abbreviations

AAFC	Agriculture and Agri-Food Canada
AfSIS	African Soil Information Service
ARC	Agricultural Research Corporation Sudan
BISE	Biodiversity Information System for Europe
CAS	Chinese Academy of Sciences
CCW	Countryside Council for Wales
CCAFS	Climate Change, Agriculture and Food Security
CEH	Natural Environment Research Council Centre for Ecology and Hydrology
CEPAL/ECLAC	La Comisión Económica para América Latina / Economic Commission for Latin America and the Caribbean
CI	Conservation International
CIAT	Centro Internacional de Agricultura Tropical / International Center for Tropical Agriculture
CIAT-TBSF	Tropical Soil Biology and Fertility Institute
CIRAD	Centre de Coopération Internationale en Recherche Agronomique pour le Développement
CNR-IBIMET	Portale dell'Istituto di Biometeorologia del Consiglio Nazionale delle Ricerche
CSIR	Council for Scientific and Industrial Research
DEFRA	Department for Environment Food and Rural Affairs
DSA-SUN	Seconda Università di Napoli
EEA	European Environment Agency
EI	The Earth Institute, Columbia University
EP	EcoAgriculture Partners
ESRC	The Economic and Social Research Council
FAO	Food and Agriculture Organization of the United Nations
FAS	Foreign Agricultural Service of the United States Department of Agriculture
GEO	Group on Earth Observations
GEOSS	Global Earth Observation System of Systems
GIEWS	Global Information and Early Warning System
GTOS	Global Terrestrial Observing System
HEA	Household Economy Approach
IAO	Istituto Agronomico per l'Oltremare
ICRAF	World Agroforestry Centre
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics
IFAD	International Fund for Agricultural Development

IFPRI	International Food Policy Research Institute
IFPRI	International Food Policy Research Institute
IIASA	International Institute for Applied Systems Analysis
IRRI	International Rice Research Institute
IRSA	NASA/IPAC Infrared Science Archive
ISRO	Indian Space Research Organisation
JECAM	Joint Experiment for Crop Assessment and Monitoring
JRC	Joint Research Centre of the European Commission
KCL	King's College London
LDSF	Land Degradation Surveillance Framework
LSCE	Laboratoire des Sciences du Climat et de l'Environnement
LSMS-ISA	Living Standards Measurement Study-Integrated Surveys on Agriculture
LTER	Long Term Ecological Research Network
MPIB	Max-Planck-Institut für Bildungsforschung (Max-Planck-Institute of Biogeochemistry)
NASA	National Aeronautics and Space Administration
NEON	National Ecological Observatory Network
NERC	Natural Environment Research Council
NIEA	Northern Ireland Environment Agency
NOAA	National Oceanographic and Atmospheric Administration
NSF	National Science Foundation
OPHI	Oxford Poverty & Human Development Initiative
UCL	Université catholique de Louvain
ULEIC	University of Leicester
ULUND	Lunds universitet
UMD	University of Maryland
UNEP-WCMC	UNEP-World Conservation Monitoring Centre
UR2PI	Unité de Recherche sur la Productivité des Plantations Industrielles
USAID	United States Agency for International Development
USDA	United States Department of Agriculture
USGS	United States Geological Survey
UST	Università degli Studi della Tuscia
VDS	Village Dynamics Studies
WAW	World Agriculture Watch
WDPA	World Database on Protected Areas
WLE	CGIAR Program on Water, Land and Ecosystems

Glossary of terms and definitions

Term	Definition
Agro-ecosystems	Terrestrial ecosystems in which agriculture is a significant activity
Belief	Perception of truth. Degree of belief in the truth of an event.
Driver	Any natural or human-induced factor that directly or indirectly causes a change in a system
Fast variable	A variable that fluctuates rapidly with time in response to external factors and requires frequent measurement (e.g. hourly or daily) to track
Health	Level of functional efficiency or capacity, in humans or ecosystems
Indicator	A succinct measure that aims to describe as much about a system as possible in as few points as possible
Intervention	Any promotive, preventive, curative, or rehabilitative activity where the primary intent is to improve conditions
Measurement	The assignment of numbers to objects or events; all measurements consist of three parts: magnitude, dimensions (units) and uncertainty
Metric	A measurement that quantifies results
Risk	A probability of an adverse outcome, or a factor that raises that probability
Slow variable	A variable that fluctuates only slowly with time in response to external factors and requires only infrequent measurement (e.g. every several years) to track
Trade-off	A situation where a gain in one quality or aspect results in losses in another quality or aspect.
Uncertainty	The lack of complete certainty or the existence of more than one possibility. The “true” outcome/state/result/value is not known (Hubbard, 2010)
Value	Relative worth, utility, or importance
Value of Information	Reduction in expected opportunity loss (i.e. the cost of being wrong x the chance of being wrong) as a result of further information
Variable	A quantity that may assume any given value

Preface

Global challenges associated with food insecurity, poverty, nutrition and health, ecosystem damage, and climate change are intensifying in an increasingly crowded planet. Many donors, NGOs, the private sector, governments and research organizations are responding by challenging the way they do business to accelerate improvements in outcomes and increase impacts. For example the CGIAR is reorganizing itself around new CGIAR Research Programs, guided by a new Strategy and Results Framework, with a strong emphasis on producing results towards well-defined development outcomes. DFID and other donors are also seeking ways to increase the impact of their programmes and investments and increase cost-effectiveness.

Following the adage "if you cannot measure it, you cannot manage it", this review critically examines what we can learn from monitoring initiatives over the past 20 years in terms of how we should be prioritising investments in research for development and monitoring research outcomes and impacts. It challenges, whether we are actually doing the right thing (i.e. becoming more effective) as opposed to getting better at doing things right (becoming more efficient) but at the wrong thing. As the management systems scientist Russell Ackoff advised: more effort should be put on designing what we do want as opposed to getting rid of what we don't want.

We hope that the findings of this review will assist in stimulating new thinking on how to approach measurement and monitoring initiatives and new ways of doing business in agricultural research and development. The CGIAR Program on Water, Land and Ecosystems is currently investing in new research based on the recommendations of this report.

Keith Shepherd

Project Leader

World Agroforestry Centre (ICRAF) & Information Systems Research for Water, Land and Ecosystems

Andrew Noble

Interim Director

CGIAR Program on Water, Land and Ecosystems

Executive summary

Research and development stakeholders working on sustainable intensification of agro-ecosystems are striving to become more effective in achieving development outcomes. Key questions facing them are:

- How to evaluate alternative research and development strategies in terms of their potential impact on productivity, environmental services and welfare goals, including trade-offs among these goals?
- How to cost-effectively measure and monitor actual effectiveness of interventions and general progress towards achieving sustainable development objectives?

The purpose of this review was to identify lessons and opportunities for the derivation and use of data from monitoring initiatives in the sustainable intensification of agriculture. The ultimate goal is to provide decision-makers with tools that they can use to explore trade-offs between food security, environmental and socio-economic goals. The analysis is intended to inform the development of any future DFID research investments and engagement with stakeholders in this area.

The first step was to identify key initiatives in data monitoring systems relating to agriculture, paying particular attention to those that also acknowledge the impact on ecosystem health, and/or poverty and well-being. A total of 103 monitoring initiatives were identified. The second step was to review the identified initiatives with respect to their degree of achievement in meeting a set of 34 criteria that had been established from a general literature review. All initiatives were evaluated with respect to their conceptual framework and a subset of 24 initiatives was screened against the full set of criteria. Pertinent additional findings from the literature on monitoring of data in other fields were also considered. Based on this information a gap-analysis of the systems, indicators and metrics was conducted identifying strengths and weaknesses in methodology and use. Insights, lessons and recommendations were then drawn. Preliminary findings were shared with a group of experts and stakeholders identified in consultation with DFID and their feedback incorporated.

Common weaknesses were identified in a number of areas. Many initiatives lack a clear conceptual framework that could demonstrate an understanding of the system under study. In particular theories of change on how the monitoring results would affect behaviours and explicit linkage to specific decisions are lacking in most initiatives. There are problems in most initiatives in defining the target inference space (geography, population) and how that is being sampled. Very few initiatives have come to terms with how to integrate biophysical and socio-economic indicators and sampling frames. A common problem is the lack of well-defined sample units or strata and a lack of consistent measurement protocols. Use of statistically sound study designs to allow attribution of outcomes to interventions is still rare. A lack of consideration of uncertainties both conceptually and in communicating results is pervasive. Similarly few initiatives have tackled trade-offs among objectives. Data sharing agreements were found to be wanting. Evaluation of monitoring initiatives themselves is lacking and there is virtually no information on cost-effectiveness of the measurements or the impact of the initiatives. Few initiatives have been sustained over the long term pointing to inadequate consideration of institutional sustainability.

An over-riding lesson is the surprising lack of evidence for the impact of monitoring initiatives on decision-making and management. Useful insights were drawn from public health surveillance, systems thinking in industry and public services, and decision sciences. A set of recommendations for good practice in monitoring initiatives is given and opportunities for new thinking in monitoring design are identified, including a decision analytic conceptual framework.

Introduction

In order to meet the food requirements of a growing, wealthier population in the face of natural resource constraints and climate change, there is wide recognition of the need for sustainable intensification of agriculture in which yields are increased without adverse environmental impacts. Research and development stakeholders, including donor agencies, investment banks, governments, development NGOs, the private sector, and research organizations such as the CGIAR, are all asking the same questions:

- How to evaluate alternative research and development strategies in terms of their potential impact on productivity, environmental services and welfare goals, including trade-offs among these goals?
- How to cost-effectively measure and monitor actual effectiveness of interventions and general progress towards achieving sustainable development objectives?

Of particular importance here is the increased need and demand for monitoring initiatives that allow to link biophysical and socioeconomic indicators along with the increased demand for assessing environmental impacts of human development focused interventions. This demand is also reflected in one of the outcomes of Rio+20, which called for development of Sustainable Development Goals (SDGs), which would complement and further develop the MDGs with a focus on sustainability.

Governments need to be able to target, design and evaluate agricultural programmes based on scientifically sound data on the potential of land and likely productivity-environment-welfare outcomes from local to national scales.

The private sector is increasingly interested in demonstrating good practices with respect to environment-welfare indicators and there is growing interest in certification schemes that demonstrate good practices. An example is the UN Global Compact that has developed a new Sustainable Agriculture Initiative, which is bringing together 10-15 of the world's largest agricultural corporations to develop a new approach to business principles.

The newly reformed CGIAR is increasingly being challenged, and challenging itself, to better target its research towards the system outcomes of reducing rural poverty, improving food security, improving nutrition and health, and sustainable management of natural resources, and to facilitate uptake and use of research results.

Donors, including DFID, USAID[1] and many others are increasingly developing adoption targets for agricultural technologies that would need to be both pre-assessed and monitored to ensure that targets can be achieved and that potentially unforeseen adverse environmental impacts can be avoided.

The ultimate goal of this work is to provide decision-makers with tools that they can use to explore trade-offs between food security, environmental and socio-economic goals. Many of these are lacking at the agricultural production system level, but to make progress in informing decision-makers, we need to learn lessons from other sectors of the economy on what is feasible.

The purpose of this review, therefore, is to identify lessons and opportunities for the derivation and use of data from monitoring systems in the sustainable intensification of agriculture. The analysis is intended to inform the development of any future DFID research investments and engagement with stakeholders in this area. The results are also relevant to other funders of research, particularly those investing explicitly in sustainable intensification

[1] Speech by Raj Shah, USAID administrator, at IFPRI on Dec 7 --http://www.ifpri.org/event/harnessing-agricultural-innovation-transformative-impact

(e.g. BMGF, USAID amongst others), as well as the CGIAR itself. Specifically it has the potential to inform:

- The development of systems to measure ex-ante and ex-post the impact of CGIAR investments (of relevance to DFID as a significant funder) at the level of the 4 system outcomes (http://consortium.cgiar.org/our-strategic-research-framework/) and to link outcomes at CRP level with higher level impacts.

- Mechanisms to analyse the impacts and trade-offs associated with sustainable intensification at different scales (sub-national, national, regional).

- Future research on value for money metrics for measuring agriculture, ecosystem and poverty and nutritional outcomes.

Objectives and scope

Specific objectives of the review are to:

1. Make an inventory of key initiatives in data monitoring systems relating to agriculture, paying particular attention to those which also acknowledge the impact on ecosystem health, and/or poverty and well-being; including their objectives, indicators and metrics used.

2. Summarise any pertinent findings from monitoring of data related to ecosystem health and poverty/well-being not directly connected to agricultural systems.

3. Classify the indicators identified in relation to their primary intended users (e.g. scientists vs decision-makers).

4. Undertake a gap-analysis of systems, indicators and metrics, identifying strengths and weaknesses in methodology and use.

5. Undertake an evidence-based synthesis of lessons based on a number of criteria including application/use at different levels, standardisation of indicators, composite metrics, coverage and replicability, validity, cost of indicator systems, level of disaggregation (age, gender, wealth etc).

6. Synthesise lessons learned, guided by an expert panel, drawing on the evidence collated in the inventory and gap analysis, with particular emphasis on the indicators which would be of most use in helping to inform decision-makers on the evaluation of trade-offs.

The scope includes to review:

- The evidence on monitoring systems and metrics of relevance to low- and middle-income countries undertaken in the last 20 years, but to also review systems in developed countries in order to derive evidence-based lessons.
- Data collection and metrics developed by the private as well as public sector bodies for analysing key ecological processes and biodiversity under conditions of agricultural intensification.
- Evidence in relation to a number of criteria including the validity and costs of such systems.

Approach and methods

The steps taken were:

1. Identify key initiatives in data monitoring systems relating to agriculture, paying particular attention to those that also acknowledge the impact on ecosystem health, and/or poverty and well-being. Annex A of the original call (Appendix 1 DFID call) provided a starting point for this inventory. The principal method of identifying additional monitoring systems was a search of websites using the Google search engine.

2. Review the identified initiatives with respect to their degree of achievement in meeting or considering the criteria laid out in Table 1 below, expressed as a series of questions. This list of questions constitutes the main conceptual framework for the review. The framework was applied in two phases due to the large number of combinations of initiatives and criteria. Phase 1 entailed a rapid scoping study of the initiatives. The performance of all the identified initiatives was assessed against a few key criteria to identify a subset of initiatives for detailed review. Phase 2 was a more detailed review that evaluated the subset of initiatives against the full list of criteria.

3. Summarise any pertinent additional findings from the literature on monitoring of data related to ecosystem health and poverty/well-being not directly connected to agricultural systems.

4. Based on the above information, undertake a gap-analysis of the systems, indicators and metrics, using the criteria in 2 above, identifying strengths and weaknesses in methodology and use.

5. Synthesise lessons learned, drawing on the evidence collated in the inventory and gap analysis, with particular emphasis on the indicators that would be of most use in helping to inform decision-makers on the evaluation of trade-offs.

6. Share progress and findings with a group of experts and stakeholders identified in consultation with DFID and incorporate feedback.

Table 1. Questions on key aspects to consider in design and evaluation of monitoring initiatives

1. Does the monitoring system have a conceptual model?
2. Does the system have clearly defined monitoring objectives or questions?
3. Has the system been designed to support specific decisions?
4. What is the scope of inference of the monitoring – both temporally and spatially?
5. Can the problems being addressed be decomposed into components that are directly observable?
6. Does the system allow the relationships between research outputs and outcomes to be explored via a means-ends framework?
7. Is it possible to assess the trade-off between the value of the analysis of the information collected and the cost of measurement?
8. Who are the actors in the system and what is their engagement with the system?
9. Which stakeholders were involved in the design and implementation of the system?

10. Have the target and sample populations or units been defined?
11. Has the scale hierarchy of the natural and/or human system been considered and reflected in the sampling design or the analysis of the measurements?
12. Has consideration been given to conditioning variables (such as gender, economic status) that influence outcomes, for example via stratification?
13. Is the system able to detect change in the indicators or measures proposed within the temporal and spatial measuring frames?
14. Has the sampling frame been communicated?
15. Is the sampling frame representative and appropriate for the questions that need to be answered?
16. Is the sampling frame or sampling process biased?
17. Is the sampling process practical?
18. Are proxy indicators sought for 'fast' variables that are difficult or costly to measure?
19. Are there published benchmark or critical values with which to judge the measured variables?
20. Do the indicators allow disaggregation according to farm types, resource endowment, age, gender etc.?
21. Does the monitoring system use composite metrics or indices, and if so are these replicable?
22. Is there any evidence that measurement protocols have been developed and used?
23. Have the trade-offs between the cost of measurement (frequency, access, equipment etc.) and the errors (of omission and commission) been assessed?
24. Does the monitoring system employ statistical analysis methods?
25. Does the monitoring system link data from different sources, and if so what are the analytical tools used to harmonise these data?
26. What are the data sharing principles of the monitoring system?
27. How does the monitoring system deal with confidentiality issues and other ethical considerations (e.g. intentional reduction of precision of GPS coordinates to avoid identification of individuals)?
28. How are the results of the monitoring interpreted – is there any trade-off analysis or assessment of dose-response relationships?
29. How are the results of the monitoring communicated – what is the audience and what channels are used?
30. Are the results of the monitoring evaluated, and if so is there feedback into the conceptual framework or choice of indicators?
31. Is there any link between system monitoring and the evaluation of interventions?
32. Can the monitoring results be used for cost benefit analysis (either ex ante or ex post) of agricultural interventions?
33. How does the monitoring framework consider sustainability issues (e.g. discount rates, valuation of ecosystems services)?
34. Is there an explicit treatment of trade-offs among productivity, environmental and well-being objectives?

Inventory

The call for proposals from DFID included a provisional list of 21 initiatives for monitoring agriculture and ecosystems and 8 initiatives for monitoring food security, poverty and livelihoods. A number of these initiatives were in fact the names of institutions that have a mandate that includes monitoring, rather than specific initiatives. The top 100 Google web search results (Table 2) were examined and were found to refer to initiatives that had already been identified by DFID, while others were not relevant or referred to the same initiative. We recognise a potential bias in the search procedure towards initiatives in Anglophone countries, but we believe that most global and continental scale initiatives were identified. The WLE review team provided further information on initiatives, and this was supplemented by information from Manuel Winograd (formerly senior advisor of social and ecological vulnerability at the European Environment Agency), from Matt Walpole, Head of Programme for Ecosystem Assessment at UNEP-WCMC, and from Barbara Herren at FAO.

Table 2. Results of website and webpage search in the Google search engine (1 – 15 September 2012)

Search keywords	Number of results
"farming systems" AND "monitoring"	606,000
"ecosystems monitoring"	101,000
"poverty monitoring"	96,000
"agricultural monitoring"	50,000
"monitoring poverty"	44,000
"monitoring agriculture"	21,500
"livelihoods monitoring"	8,050
"monitoring livelihoods"	3,080

In addition to the original DFID list, we identified a further 57 initiatives under the heading of agriculture and ecosystems, resulting in 78 initiatives. Thirty of these initiatives were global in extent, 21 were continental in their scope, and 15 were limited to a particular nation state, while the remaining initiatives were for groups of countries or case studies. Under the heading of poverty and livelihoods, we identified 17 more initiatives to give a total of 25. Of these 14 were global in their reach, five initiatives spanned at least one continent, the focus of two initiatives was just on one particular country, while the remaining four concentrated on a few selected countries.

Scoping study

The first three initiatives in the list were assessed using all 34 criteria/questions and comments and reflections entered in a spreadsheet (Appendix 2 Scoring of Initiatives). Despite the brevity of the comments, this exercise resulted in a fairly comprehensive idea of the gaps, and the pros and cons of each of the three initiatives. This was, however, an extremely time consuming exercise and we decided to seek the criterion/question that best explained the answers to the others as a means of reducing the number of initiatives to a manageable number. The best guide appeared to be whether the initiative had a conceptual framework, since this gave a good indication of whether the initiative actually carried out monitoring activities and whether these were relevant for agricultural or farming initiatives.

The conceptual framework or model (examples of which can be seen in Appendix 3 Narrative of detailed reviews), appear to be particularly important for initiatives that seek to link issues of agriculture, ecosystems and environment, and poverty and nutrition. It is possible that more focussed or specialised monitoring initiatives may not need such an organising framework, but the existence of a conceptual framework is also a guide for clear documentation of the initiative – which is a fundamental for an accurate assessment of the initiative.

The websites and relevant documentation (where available) of all 103 initiatives were searched for information on conceptual frameworks. The results (Table 3) show that the majority (66) of initiatives have a conceptual framework in either graphical or narrative format. Nevertheless, a significant minority of initiatives – particularly those that deal with agriculture or ecosystems – did not have a framework or are not really monitoring initiatives. Instead a number of these initiatives are platforms or networks of researchers that allow experiences to be shared or methods to be discussed.

The number of initiatives with a graphical or narrative framework was still too large to assess in detail given the time available so we decided to reduce the number further based on the resolution of the initiative. An important consideration was whether the initiative could capture the dynamics of specific agricultural production systems, especially any links between agricultural production, livelihood outcomes and the impacts on natural resources. Another consideration was the potential to compare initiatives that were superficially similar – such as those that rely on remotely sensed data for crop monitoring, or which monitor ecosystems – but supplied by different types of agencies[2]. In addition a deliberate choice was made to include at least one national level monitoring initiative from a developing country. The shortlist of initiatives selected for a detailed review (Table 4) using all criteria included 17 that concentrate on agriculture and ecosystems, and 7 that focus on poverty and livelihoods. The selection process is likely to result in the initiatives not surveyed in detail having a higher frequency of weaknesses than those that were reviewed in detail.

Table 3. Results of scoping study of potential monitoring initiatives

Does the initiative have a conceptual framework?	Agriculture and Ecosystems	Poverty and Livelihoods
Yes	28	12
Narrative of a framework	16	10
No information available	14	2
Framework available but not really a monitoring initiative	16	1
No framework and not really a monitoring initiative	5	0
TOTAL	**79***	**25**

* Total number of agriculture and ecosystem monitoring initiatives was 78 due to duplication of one initiative

[2] Indicators were not classified by types of users as a selection criterion since it would have involved in-depth analysis of each of the 103 initiatives, however, the users of the information provided by the initiatives were identified in the in-depth analysis.

Table 4. List of initiatives selected for in-depth review

Agriculture and Ecosystems				
	Initiative name	**Organisation(s)**	**Extent**	**Shortlist?**
1	UK National Ecosystem Assessment	UNEP-WCMC, CCW, DEFRA, ESRC, NERC, NIEA, Scottish Government, Welsh Assembly Government	UK	Yes - due to ecosystem service framework, and reviewed in detail to assess key criteria
2	CarboAfrica	UST, MPIB, ULUND, GTOS, FAO , CIRAD , CEH, CNR-IBIMET, IAO, DSA-SUN, CSIR, UR2PI, ARC, LCSE, KCL, ULEICS	SSA	No, but reviewed in detail to assess key criteria
3	Land Degradation Surveillance Framework of the African Soil Information Service	EI, CIAT-TBSF, ICRAF	SSA	Yes - due to sampling strategy, and reviewed in detail to assess key criteria
4	Integrated Monitoring System for African Landscapes (Vital Signs)	CI, EI, CSIR	Africa / SSA?	Yes, due to explicit links between agriculture, livelihoods and ecosystems
5	Millennium Ecosystem Assessment	Multiple partners	Global	Yes - due to ecosystem service framework, and global scope
6	Landscapes for People, Food and Nature	EP	Global - EcoAgriculture Case studies?	Yes, due to explicit links between agriculture, livelihoods and ecosystems
7	GEOSHARE	Purdue University, McGill University, Stanford University, Universität Bonn, IRRI, IFPRI, CIAT	Global, regional nodes for LAC, Africa and Asia	Yes, for potential and realisation of need to integrate
8	GEO GLAM	UMD, JRC, FAS, IRSA, CAS, ISRO, GEO Secretariat, AAFC, UCL, IIASA, USGS	Global	Yes, because although concentrating on remotely sensed imagery there are links to ground data
9	Committee on sustainability assessment (COSA)	Non-profit global consortium of institutions (http://sustainablecommodities.org/partners)	Global	Considers sustainability, good links to interventions (such as fair trade) limited to case studies but relevant for DFID
10	National Agri-Environmental Health Analysis and Reporting Program	AAFC	Canada	Explicit link between agricultural metrics and NRM outcomes
11	Joint Experiment for Crop Assessment and Monitoring	GEO Agriculture Monitoring Community of Practice	North America - Canada South Nation Watershed, Canadian Red River Watershed,	This brings together various crop monitoring initiatives with a coherent science plan - although the links to livelihoods and ecosystems

			Mexico, USA. South America - Argentina, Paraguay. Europe - Belgium/France, Russia, Ukraine. Africa - Mali, South Africa. Asia - China Anhui, China Guangdong, China Heilongjiang, China Jiangsu, China Shandong.	are weak
12	Statistics and Ecological Monitoring Directorate	Nature Conservation General Directorate (DGCN), Government of Burkina Faso	Burkina Faso	Has strong links to conservation issues, and includes data collection at multiple scales and from multiple instruments - worthy example of developing country initiative
13	Long Term Ecological Research Network	Multiple - funded by NSF	USA	Well-designed monitoring initiative but direct interaction with agriculture limited to one site - still worth reviewing in depth
14	National Ecological Observatory Network (NEON)	NEOn Inc. - funded by NSF	USA	Appears similar to LTER and worth reviewing to see what are the complementarities - offers a good example of potential for duplication
15	Biodiversity Information System for Europe	JRC, Eurostat, EEA	Europe	Good example of continental scale combination of different monitoring initiatives
16	World Database on Protected Areas	UNEP-WCMC	Global	Good example of monitoring a response indicator
17	World Agriculture Watch	FAO, IFAD, Government of France and CIRAD	Global	Yes, possibly few activities have taken place but the conceptual framework is very promising

Poverty and Livelihoods

	Initiative name	Organisation(s)	Extent	Shortlist?
18	Women's Empowerment index	OPHI/IFPRI/USAID	Guatemala, Bangladesh, and Uganda	Spatial scope is presently not extensive but does provide new gender indicators, and implementation is expanding rapidly
19	Global Information and Early Warning System	FAO	Global	Makes specific use of RS imagery and links to food security prospects, worth looking at in greater detail to compare with JECAM

				and GLOBCAST
20	FEWSNET - Famine Early Warning System	USAID, Chemonics International Inc., USGS, NASA, NOAA, USDA	Sub-Saharan Africa, Central America, Haiti, and Afghanistan	Like FIVIMS an important and necessary addition to the RS based crop growth monitoring initiatives which only capture (national) food availability, but has more of a monitoring aspect
21	Living Standards Measurement Study-Integrated Surveys on Agriculture	World Bank	Mali, Niger, Nigeria, Ethiopia, Uganda, Kenya, Tanzania, Malawi	Explicit link between agricultural metrics and outcome metrics at the household level - but limited to certain countries
22	Household Economy Approach	FEG Consulting and Save the Children - UK	Sub-Saharan Africa, Central America, the Balkans and Asia	Yes, because of explicit links to livelihood zones which have a strong link to agricultural activities
23	Basic needs mapping	CEPAL / ECLAC	LAC	Little direct relation to agriculture but offers monitoring of development outcomes for disaggregated spatial units
24	Village Dynamics Studies	ICRISAT, IRRI	India and Bangladesh	Explicit links between livelihood outcomes, agricultural interventions and environmental drivers

Review of evidence

Inference space

Unit of analysis

The landscape, household and plot scales were the most common units of analysis, which includes those initiatives that rely on data from sensors mounted on satellite platforms (Figure 1). Units of analysis at the household level were more prevalent among those initiatives that sought to measure some component of poverty, food security or livelihoods. Initiatives that rely on remotely sensed data alone are unlikely to be able to capture data that can be used for monitoring of the impacts of interventions that promote sustainable intensification. In remotely sensed initiatives, ground truthing efforts and validation also need to be conducted, requiring ground measurements.

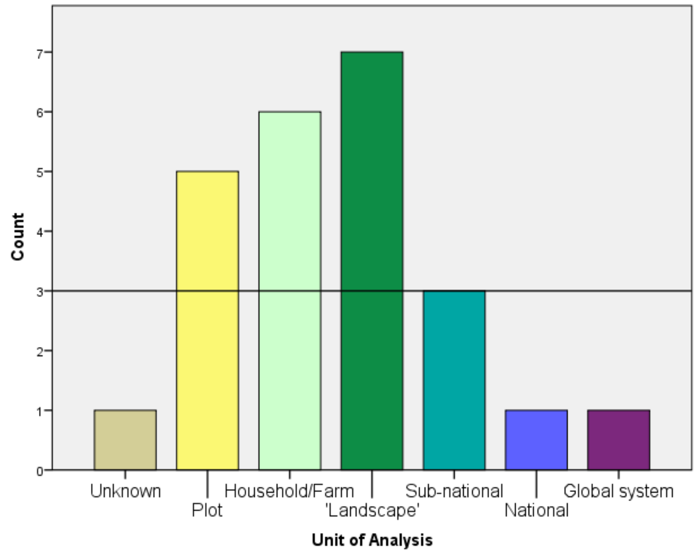

Figure 1. Unit of analysis for each monitoring initiative reviewed

Spatial extent of monitoring

In some initiatives the spatial extent is defined by the availability of data (for example the GEO-GLAM and JECAM initiatives that originate from the Earth Observation community); in other cases the defining factor is the location of cases where a particular intervention is being carried out (examples include the EP landscapes initiative and the COSA coffee farms), or the participation of partners from specific localities or with a specific geographical mandate. Other restrictions on the spatial extent are the costs of establishing a monitoring initiative, and as a consequence a preference for pilot studies to prove the concept and methods.

However, in most cases there was no specific inference geography defined which was then sampled. The exception to this is the Africa Soil Information System, which takes a random sample of sites from Sub-Saharan Africa stratified on climate zones.

Frequency of monitoring

Data availability and practicality define the unit of analysis and the temporal scope in many cases. Exceptions include specialised monitoring initiatives, such as CarboAfrica that measures CO_2 fluxes constantly. One common problem was a lack of information (Figure 2) about the frequency of monitoring particularly among the agriculture and ecosystem category. Many of the initiatives which appeared to be monitoring initiatives have actually only produced a baseline assessment. These were more common among the initiatives that concentrate on agriculture and ecosystem themes such as the ecosystem assessments. This raises a flag on the sustainability of monitoring initiatives in the long-term.

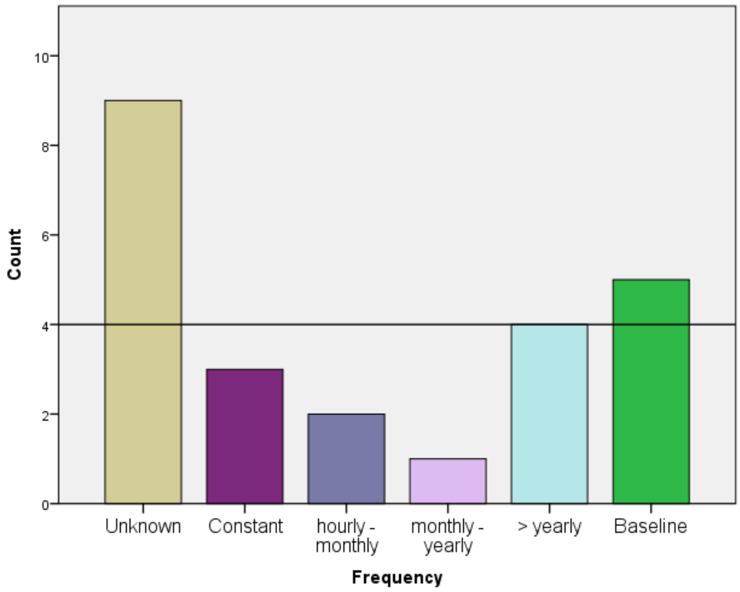

Figure 2. Frequency of monitoring in each initiative reviewed

Monitoring initiatives that relied on household data from surveys or population censuses could often only be updated on a 5 or 10 yearly basis. This frequency can make it difficult to attribute change to a specific technological intervention but is more appropriate for policy interventions, and attribution is also limited to the disaggregation of agricultural production information and environmental indicators collected in panel surveys. Panel surveys that return more frequently to the same households are more limited in their spatial extent.

Objectives and decision support

Two thirds of the initiatives reviewed have a clear set of objectives; those initiatives that are less clear in communicating objectives tended to be global in their extent and/or under development. Five complementary types of objectives (Figure 3) were identified: these ranged from the description of some phenomenon often using indicators of the state of an agricultural system, through attempts to explain phenomena with indicators on pressures or drivers, to some kind of prediction or projection of future impacts and potential responses. At the same time numerous initiatives had other objectives relating to the construction or maintenance of institutional (and personal) networks that enable monitoring, and the development of methods that can be applied for the monitoring of agricultural systems.

The most common type of objective was the description of systems or environments, with all except four initiatives (GEO-GLAM, JECAM, LSMS-ISA and BNI) stating this as one of their goals. The four initiatives that did not aim to describe phenomena were instead concerned with developing methods and preferred to devolve actual measuring responsibilities to other institutions. Fourteen initiatives sought to explain some aspect of the system they were monitoring, and the goal often was to look forward and create scenarios, predictions or recommendations for future systems.

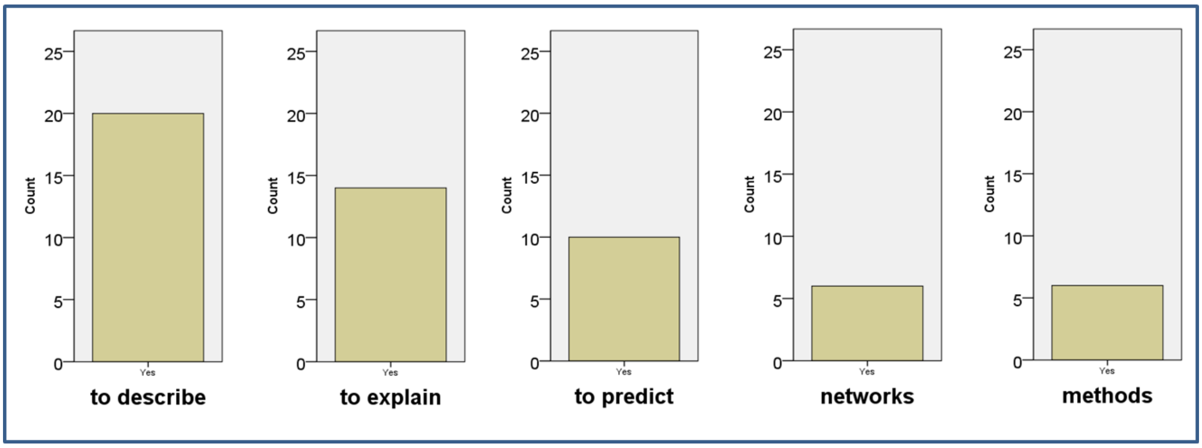

Figure 3. Objectives of the initiatives reviewed

All initiatives showed evidence of engagement with stakeholders in the system, but only six of the twenty-four initiatives have been designed to support decisions, of which only some decisions are specific, e.g. investment decisions supported by COSA, farm-level decisions on nutrient management supported by AAFC's NAHARP, or supporting the mitigation of local food insecurity by FEWSNET. Initiatives whose results were used mainly by researchers had not been designed to support specific decisions.

Other initiatives assist more general decisions, such as the UK National Ecosystem Assessment, which aims to establish a "comprehensive evidence base to support decision-making by different actors and at different scales". What is notable is that the initiatives that seek to support decisions tend to be those that have a more ecological focus.

Only 5 out of 17 of the agro-ecosystems initiatives and none of the 7 livelihoods initiatives had clearly defined decisions that the initiatives were designed to address. Of all the 24 initiatives, only the Millennium Ecosystem Assessment had an explicit objective to identify key uncertainties that hinder effective decision-making. Several initiatives had the objective to provide "up-to-date and accurate information", but without regard to the value of reducing uncertainty.

There was not a clear example of a means-objectives framework illustrating a theory of change of how monitoring results would achieve development or decision outcomes.

Sampling Strategy

Statistical design considerations are a necessary component of monitoring initiatives in order to make reliable inferences while reducing the sampled population to improve cost effectiveness. Attention also needs to be given to key stratification variables or covariates that are known to influence outcomes (e.g. gender, resource endowment).

We were able to find evidence for a sampling strategy and the sample frame in just over half of the initiatives reviewed. There was no significant difference in this respect between the monitoring initiatives dealing with agriculture and ecosystems and those specialising in livelihoods, poverty and food security.

A common problem was the lack of well-defined sample units or strata. Examples of statements are "representative land cover types", and "geographical distribution of sites or ecosystem type never became active criteria in the selection process". Often countries or case studies or villages within countries are used to define geographic scope of an initiative but without reference to whether they represent a sample of a larger area. In contrast LTER and AfSIS are examples of initiatives where sample units are well defined.

One third of the initiatives had sampling frames that potentially allowed the attribution of the impacts of agricultural interventions, and less than one third considered conditioning factors explicitly via stratification in their sampling frame (Figure 4). Initiatives that were categorised as addressing livelihoods, poverty, and food security were more likely to stratify according to conditioning factors. Only one initiative (Commission on Sustainability assessment – COSA) included different treatments and conventional farms (as counterfactuals) in their sampling frame.

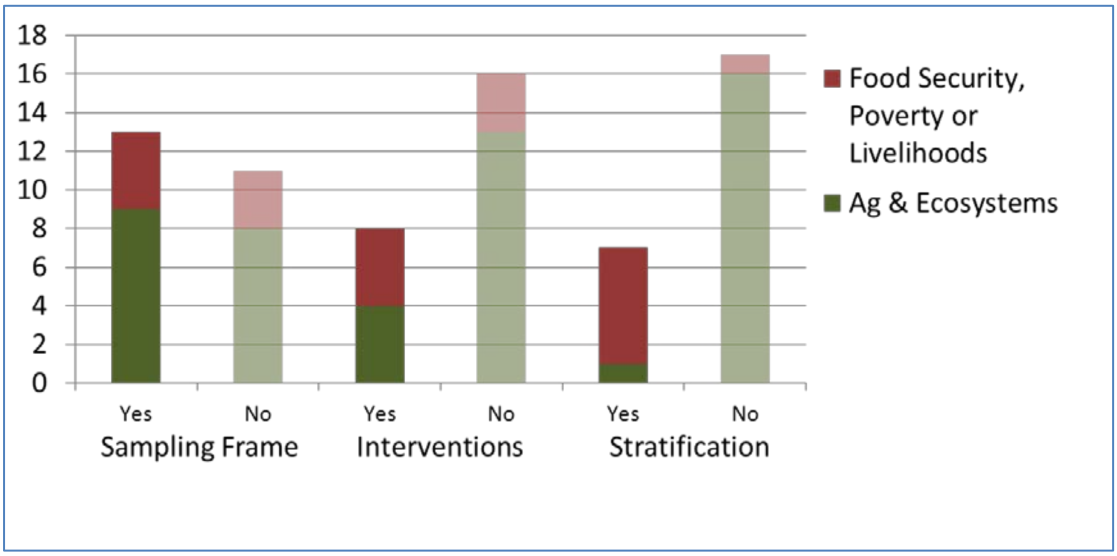

Figure 4. Number of monitoring initiatives that (i) had a well-documented sampling frame, (ii) allowed impact of interventions to be measured and attributed (iii) stratified on conditioning factors.

Fewer than half of the monitoring initiatives reviewed considered the hierarchy of the human or natural system in their sampling frames (Figure 5). Notable exceptions were the Landscapes for People, Food and Nature initiative which considered scales from the plot to landscape, ICRISAT's Village Dynamics Studies which measured indicators at the plot, farm, village and district levels, and the World Agriculture Watch which was novel in considering the whole value chain in addition to production units, and 'territories'.

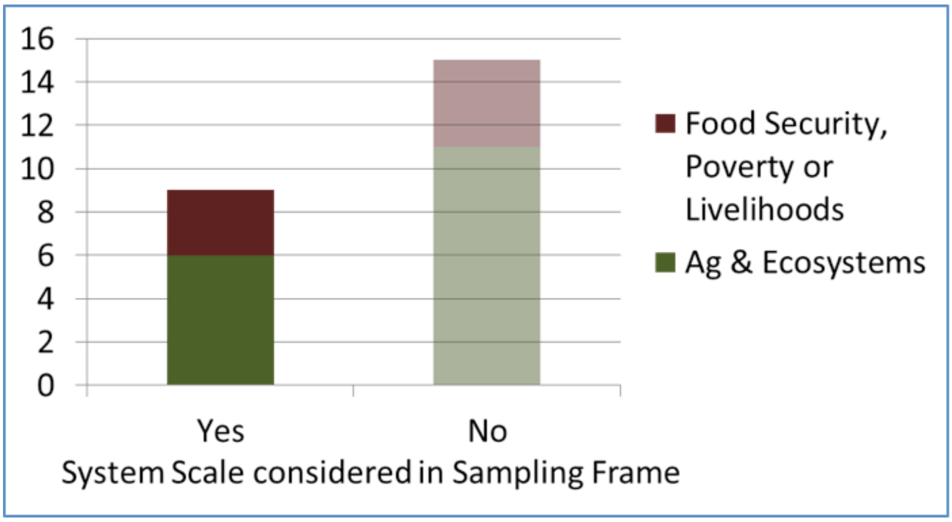

Figure 5. System scale or hierarchy considered in the sampling frame.

Indicators

There was evidence for the use of proxy indicators for 'fast' variables that are difficult or costly to measure in 5 out of 17 agro-ecosystem initiatives and 1 out of the 7 livelihood initiatives, while we found that published or critical values were used to judge the measured variables in only 3 of the agro-ecosystem initiatives but in 6 of the livelihood initiatives. This is a reflection of a lack of direct linkage of the ecological monitoring initiatives to decision-making.

There was evidence for indicators being analysed at different levels of conditioning variables (i.e. disaggregated) in only 2 of the agro-ecosystem initiatives and 4 out of the livelihood initiatives. This is an indication of a lack of further analysis of monitoring results, which perhaps indicates a lack of demand for their further use in decision processes.

There was evidence of clearly defined measurement protocols actually used in only 2 of the agro-ecosystem initiatives and 4 out of the livelihood initiatives. This indicates that more indicators are proposed than actually measured in practice.

There was a lack of clarity on data sharing arrangements in all but one agro-ecosystem initiative (COSA) and one livelihood initiative (LSMS-ISA), which both make their data publically available.

Sustainability and trade-offs among productivity, environmental and well-being objectives

The ultimate goal of a monitoring initiative for sustainable intensification of agriculture should be to provide decision-makers with tools that they could use to explore trade-offs among food security, environmental and socio-economic goals. These kinds of trade-offs were considered explicitly in only five of the 24 initiatives reviewed.

One of the approaches for dealing with trade-offs is via economic valuation of ecosystem goods and services and the use of a common (monetary) metric. Alternative approaches are based on ethical considerations, with respect to net losses of landscape functions at particular scales, as proposed by the LMRC; it is not clear, however, whether reference is made to critical values in this initiative. Potential trade-offs need to be considered from the beginning in the choice of indicators and the development of (evidence-based) scenarios can be used to show the impact of trade-offs. There was no evidence of the use of Multiple Goal Linear Programming, multi-attribute decision models or use of utility curves.

Cost effectiveness and the value of information

Very few of the initiatives assessed considered or reported the cost-effectiveness of the monitoring initiative or used any form of cost-benefit analysis. This is due mainly to poor quantification of the value of the monitoring; even so very few initiatives have released information about the costs. Exceptions are Vital Signs, which estimates that the cost of measuring and managing the data would amount to 1-2% of existing investment in agriculture, and GEOGLAM whose total costs are estimated at US$7.5M per year, but in neither case is there is an estimate of the monetary benefits that would accrue as a result of implementation or a social rate of return.

For other initiatives the imperative to conduct monitoring is expressed in terms other than cost-effectiveness, such as the Women's Empowerment in Agriculture Index, where measurement and analysis are necessary to evaluate interventions that support a specific target group. This is also the case where monitoring initiatives are a response to a regulatory framework, be it national in the case of the Canadian NAHARP initiative, international treaties such as the Convention on Biological Diversity, or regional agreements such as those addressed by the Biodiversity Information System for Europe. Documentation for this latter initiative suggests that the costs of implementing a well-designed indicator system would be less than the societal cost of inaction or the development of poor policies, but does not provide figures of opportunity costs to back up this assertion. Similar statements were encountered in the 30 year review of the LTER, for example "we feel that the existing network is of incredible value and represents one of the highest scientific values for the science dollar", but there is no mention of the methods used to determine the value-for-money.

Evaluation and feedback

Most initiatives use web-based dissemination of results but there was limited evidence for more active mechanisms for dissemination of results to target audiences. There was also limited evidence for active evaluation and feedback mechanisms: only three of the agro-ecological initiatives and two of the ecological indicators showed any evidence for modification of the initiative based on evaluation and feedback.

Insights from other fields

Public health surveillance

Many of the complex problems involved in managing agro-ecosystem health are similar to the types of problems that occur in public health and therefore valuable lessons may be gleaned from science approaches used in public health surveillance, which has a long history (Appendix 4 Insights from public health surveillance). For example, problems are associated with a range of physical, biological, social and economic determinants (or risk factors), both at individual (e.g. farm field) and population (e.g. landscape or national) levels. Risks are often inter-related and act together to cause a health problem; they range from proximal risks acting directly to cause the problem, to distal risks that are further back in the causal chain. In addition, risks are often separated from outcomes in time, sometimes by many decades, making it difficult to establish causality. In both public health and agro-ecosystem health, risks are generally greater for the disadvantaged in our societies, and poverty is a major risk factor. Evaluating the cost-effectiveness of alternative preventive and rehabilitation interventions is complex and must consider many factors.

A key concept in public health surveillance is a focus on the health of *populations* rather than individuals, and sampling designs are used to make inferences from the sample to the population. Interventions are designed based on a sound knowledge of the frequency of health problems in populations, such as prevalence and incidence measured using standardized protocols. Interventions are often hinged on a few important risk factors associated with a health problem. For example surveillance data established that three-quarters of cardiovascular disease – the world's leading cause of death – results from only three risk factors: tobacco use, high blood pressure, and cholesterol (WHO, 2002). Interventions and monitoring can thus be focused on those factors.

Operational surveillance systems are built into everyday health policy and practice, for both design and evaluation. The surveillance approach presumes that health status and risks are continuous processes, which cannot be approached with a discrete project-based approach. This has encouraged the use of monitoring indicators of health status and cost-efficacy over-time which are independent of individual interventions, donor or implementing agencies, or specific initiatives. As a result, health-policies and programs can more readily be compared. A significant example is the "disability-adjusted life year" (DALY), developed by the World Health Organization in 1990 to assess the total global disease-burden consistently across different diseases, risk factors and regions. This comparability enables donors and policy-makers to develop empirically informed priorities when facing investment choices to address issues as disparate as malnourishment, HIV/AIDS, road traffic safety, and lung-cancer. The diversity of stakeholders that have adopted such standardized measures means that baseline data is more readily available at the global, national and sub-national level.

Interventions are rigorously evaluated in the real world using study designs that guard against internal and external threats to validity, for example using controls, replication and counterfactuals. Recommendations on interventions are based on systematic review of evidence, mostly from randomized control trials.

Agro-ecosystem science could greatly benefit from applying the scientific principles and rigour used in public health surveillance. Recent efforts in this direction include land health surveillance (UNEP, 2012) and the Africa Soil Information Service (www.africsoils.net), which has taken the first ever population-based sample of soil health in sub-Saharan Africa.

Cinderella science

Euan Nisbet (2007), reflecting on Charles Keeling's autobiography (Rewards and penalties of monitoring the Earth), coins monitoring as science's Cinderella, unloved and poorly paid. He warns "On-the-ground monitoring is unglamorous work, seldom rewarded by funding agencies or the science community. But we neglect it at our peril." Keeling's work made us aware of rising CO_2 levels in the atmosphere and was a result of painstaking years of effort and innovation. Keeling's autobiography documents his continuous struggle to justify and fund the meticulous measurements required, but which resulted in what is without doubt the most significant finding for the survival of mankind.

Nisbet stresses that global understanding is underpinned by rigorous *in situ* (measured on the surface) long-term data series and that while satellites are providing unprecedented data on earth systems, there is a need to pair these measurements with the much more accurate *in situ* monitoring and validation. Nisbet provides the example that the ozone hole was originally missed by satellite monitoring but found by careful ground measurements.

However, the reality is that funds are finite, and funding agencies have to set priorities and make decisions on resource allocation. Therefore we argue that there is a need to link monitoring initiatives more directly to the decisions that are of high priority for human development, and demonstrate which variables are key and/or particularly uncertain toward making those decisions.

Decision analysis

Modelling decisions first

A recurrent problem with many measurement initiatives is that they are designed with the assumption that the data generated will be transformed into information, or that which results in change in behaviour. However often there is little evidence, including from this review, that this assumption is actually met. In addition, it has been observed that organizations often frequently spend 10 times the value of information on surveys and trials (Ron Howard in Savage, 2009). Experience with land resource monitoring in Australia is that a focus on monitoring has never provided the right information for decision-making (Neil McKenzie, personal communication).

The increasing need to allocate finite resources between competing interventions, programmes and projects is leading to increasing use of economic evaluation for decision-making in a number of fields (Briggs et al., 2006; Hubbard, 2010). Of particular relevance is the use of Bayesian decision theory and Value of Information Analysis. The principle behind this approach is to model the uncertainties in our existing knowledge in terms of the full range of costs and benefits and identify the value of obtaining further information towards improving investment decisions.

As one example, the UK National Institute for Health and Clinical Excellence (NICE) uses this approach to inform decisions about medical devices, diagnostics technologies and surgical procedures (Briggs et al., 2006). Up until now randomized control trials are often the sole source of evidence for intervention decisions in public health, but it is well known that trials have weaknesses and a significant number of economic evaluations are based on data from a single trial. One of the arguments for adopting decision analytic modelling is that all relevant evidence and all appropriate options can be included.

Applied Information Economics

Probabilistic decision modelling has been developed into a field called Applied Information Economics—a universal approach to measurement (Hubbard, 2010). This is a decision-orientated approach that applies to any sort of measurement problem and has been developed by combining validated approaches from several fields of decision and related sciences.

Experience from applying Applied Information Economics in many different fields over 17 years has revealed a number of insights (Table 5). One of the most striking features is the measurement inversion – the variables that have most value for improving a decision are typically not being measured and most current measurement effort is being spent on variables that have little or no information value. Further information on the approach and its application is given in Appendix 5 Applied Information Economics.

Table 5. Insights from applying Applied Information Economics in many fields (based on Hubbard 2010).

- We are not as clear as we think on the decisions we are trying to influence
- Expressing uncertainty dissolves assumptions and allows all benefits, costs and risks to be included, however intangible (especially environment!)
- Interventions can be quantitatively linked to development outcomes
- We need calibrating to reliably estimate probability distributions
- There are usually only a few variables with high information value
- We are often measuring the variables that have least economic value
- And completely missing the ones that do have value (e.g. we tend to measure costs but ignore benefits, which are typically uncertain).
- Measurement is about uncertainty reduction, not attaining a gold standard
- We often need different data than we think
- We often need less data than we think
- Even small reductions in uncertainty can have considerable value

Indicators and decision-making

The term ‘indicators’ can be applied in many different ways. In this review the definitions given are within the context of monitoring agro-ecosystems and livelihoods. Indicators are symbolic representations designed to communicate a property or trend in a complex system or entity (Moldan and Dahl, 2007). Environmental indicators have taken on such importance because they provide ‘‘a sign or signal that relays a complex message, potentially from numerous sources, in a simplified and useful manner’’ (Jackson et al., 2000). The United Kingdom’s (UK) National Health Service defines indicators (Pencheon, 2008) more simply, as: “succinct measures that aim to describe as much about a system as possible in as few points as possible”. Indicators help us understand a system, compare it and improve it.

Environmental indicators are designed to provide an important source of information for policy makers and help to guide decision-making as well as monitoring and evaluation.

There is an extensive literature on indicators in the environment, development and public health sectors, including even a scientific journal dedicated to Ecological Indicators (Elsevier). However the large volume of literature on indicators is disproportionate to their effectiveness in influencing actual policy and practices, which often remains limited. Remarkable quantities of resources have been used to develop sustainable development strategies and indicators at international, national, and local levels over the past two decades (Rinne et al., 2012). For example the International Institute for Sustainable Development found as many as 894 indicator initiatives for the monitoring of sustainable development (IISD, 2010). However, in many cases indicators are strongly technical in focus, with no close link to management decisions (Pannell and Glenn, 2000).

There is uncertainty over the performance of indicators in practice. Many indicator sets for sustainable development have been assembled and countries have started their own indicator programmes at national level, as called for by the Commission for Sustainable Development, but none has been widely implemented (Moldan and Dahl, 2007). Based on a review of sustainable development indicators, Parris and Kates (2003) concluded that no indicator sets are universally accepted, backed by compelling theory, rigorous data collection and analysis, and influential in policy. This was attributed to the ambiguity of sustainable development, the plurality of purpose in characterizing and measuring sustainable development, and the confusion of terminology, data, and methods of measurement. Rinne et al. (2012) also concluded that the development of universally agreed, balanced, and comprehensive indicators, indices, or indicator sets is challenged by changing natural and social conditions, scientific discoveries opening new questions, and changes in public and policy concerns. These results suggest that indicators need to be designed for a specific audience, purpose and context. Indeed, Pannell and Glenn (2000) conclude: "Uncertainty is subjective and personal. For this reason, the value of a sustainability indicator is necessarily subjective and personal".

Rickard et al. (2007) concluded that indicator developers state an objective of informing decision-making but make only weak efforts to ensure that indicators are designed to achieve policy impact. In the public health sector, although users generally accept that indicators have been helpful (with respect to the objectives of advocacy, accountability, system management, quality improvement, and research) evidence or well-developed evaluation of indicators is often lacking (Etches et al., 2006). In a review of the use of sustainable development indicators at national and EU level Rinne et al. (2012) noted that studies of whether and how the indicators are actually used and what intentional or unintended influences they may have are still few in number.

Rapport and Hildén (2013) consider that indicators for monitoring the state and trends of ecosystems and the consequences of anthropogenic pressures may not lead to necessary action unless they are coupled with identification and monitoring of drivers[3]. They consider that monitoring of drivers will form a basis for evaluating the effectiveness of policy responses, thus providing information that is actionable by policy makers and the public. This practice is well established in public health surveillance, where interventions and monitoring are focused on the risk factors associated with health problems (e.g. behavioural risk factor surveillance) once those factors have been identified, e.g. monitoring people's exercise habits as a risk factor for heart disease).

Rickard et al. (2007) identified three main groups of users for indicators, those requiring: (i) simple, structured information (e.g. voters, non-specialist media and decision makers); (ii) an intermediate level of detail (e.g. local government, policy implementers, NGOs, research

[3] Any natural or human-induced factor that directly or indirectly causes a change in a system

funding bodies, and industry); and (iii) technical information (e.g. policy makers, academics, and some NGOs). Since indicators are primarily used as a communication[4] tool, participatory development of indicators is needed to ensure credibility, legitimacy and salience (Cash et al., 2003; Parris and Kates, 2003).

Since indicators are designed to influence decisions (whether made by a researcher, decision maker, manager, or the general public) decision analysis approaches are particularly relevant to indicator development and use. Indeed a major lesson from the above reviews is the need for indicators to be specific to a given decision situation. Considerations of scale, aggregation, critical limits, and thresholds, etc will also be situation dependent, and situations are constantly changing. In many cases, the value of continuing to monitor would fall over time as uncertainty is reduced, and the value of observing a sustainability indicator may be dramatically reduced after a small number of observations and further monitoring then has little or no additional value (Pannell and Glenn, 2000).

Pannell and Glenn (2000) further reinforce the need to view indicators from a decision entry point:

> "The value arises purely from changing a decision maker's management choices. If an indicator does not have the potential to change a management choice, it has no value, economic, social or environmental, other than perhaps its intrinsic-interest value. If there is no management option which can economically address the sustainability problem when it reaches a bad enough level, then there is no prospect of changing management as a result of monitoring. Consequently, in this situation there is no value to the manager of a sustainability indicator for this problem. The change in management, if it occurs, is the result of a reduction in uncertainty about the impacts of different management strategies."

A classic example of this situation is the frequent advocacy of soil organic carbon levels as a soil health indicator, even though no one can say for a given location or soil type what is a good or bad level of carbon, and what action should be taken as a result of knowing its value.

Examples of how indicators are used in decision analysis are given here. Keeney (1992) advocated a process of (i) focus first on values, (ii) identify and structure objectives, and (iii) design attributes that measure the degree of achievement of the objectives. Pannell and Glenn (2000) proposed an indicator framework based on Bayesian decision theory, particularly its use to calculate the value of information under conditions of uncertainty. Hubbard (2010) uses this approach in his Applied Information Economics approach, which advocates a process that starts with identifying and clarifying the specific decision, and then identifying measurements that have value for reducing uncertainty in the decision (see Applied Information Economics). With this approach there is a strong focus on finding ways to represent any variable that decision makers consider important to a decision, no matter how seemingly intangible, expressing existing knowledge on its degree of uncertainty, and then evaluating whether their further measurement is warranted or not. Experience using the approach in many sectors has shown that indicators that people are currently using have no information value, and indicators that have high value are frequently not being measured (Hubbard, 2010). Representing the uncertainty in indicators is thus key to being able to establish whether further measurement is warranted, but is seldom done.

Gregory et al. (2012) distinguish between monitoring indicators and performance measures. They define performance indicators as being used to assess predicted performance of an alternative, whereas monitoring indicators are used to keep track of something after an action has been taken. However, we have argued that monitoring indicators that are

[4] Communication has only actually occurred when there is a response in terms of changed action.

disconnected from decisions have limited value and one should aim to validate predicted performance through monitoring, especially of variables over which there is large uncertainty.

The crucial importance of properly considering uncertainty in indicators was illustrated by Jacobs et al. (2007). They show an example of how composite scores for performance of hospitals were indistinguishable when the uncertainty in the scores was represented. However, after taking account of random variation in the underlying performance indicators, it was possible to estimate genuine performance variations. They further emphasize that different ways of aggregating underlying performance data can have a large impact on the ratings and show how composites are unstable when methods change from year to year. They conclude that composite indicators need to be published with indications of uncertainty to communicate the sensitivity of the reported measure. Further information on methodological issues is given in Appendix 6 Methodological issues of composite indices.

In summary, focusing on the decision first automatically ensures that indicators are relevant to the objectives and particular situation under consideration, while directly connecting to desired decision outcomes. Conducting the decision modelling process in a participatory way maximizes the probability of decision impacts. Representing uncertainty in indicators is critical for their correct interpretation and for determining whether their further measurement is warranted or not. It is the consequences of uncertainty that matter, rather than the precision with which quantities can be estimated (Briggs et al., 2006).

Systems thinking in public services and industry

In public service and industry, monitoring against targets has become pervasive, and yet the systems thinking literature (e.g. Deming 2000ab; Seddon 2008) concludes that performance is almost universally made worse by the specifications, regulations and targets by which organizations have been obliged to comply. Targets, for example, create unintended but predictable perverse consequences leading to diversion of resources away from the real work in order to ‘game’ the targets. Monitoring of conditions itself does not lead to better decisions or performance – for example, collecting statistics on accidents does not tell you how to reduce the frequency of accidents. The test of a good measure is: can it help in understanding and improving performance i.e. making better decisions?

An example of how current practice and thinking diverges from the systems thinking concepts is given here. A recent article published in Science (Pereira et al., 2013) presented a case for increased investment in monitoring of essential biodiversity variables (EBVs). The article begins with the context of the 2020 targets of the Convention of Biological Diversity CBD), whose international goals are to reduce the rate of biodiversity loss and avert dangerous biodiversity change. A key problem presented is the lack of consensus about what to monitor: nearly 100 variables have been proposed for the 2020 CBD targets. These variables were scored for importance to arrive at a reduced set of indictors. The article concludes “We hope that EBVs will catalyse investment in biodiversity observations, as ECVs have done for climate.” Thus the current efforts and thinking are centred on monitoring against targets, as opposed to defining what variables have economic value towards improving decisions on the design and implementation of specific intervention alternatives. An economic justification for making the measurements in terms of improved outcomes is lacking.

The role of evidence in policymaking—SciDev.Net Global Review

To better understand how to mainstream science evidence and technology innovation for development and poverty reduction, SciDev.Net conducted a series of research projects that created a robust picture of policymaking and science journalism in a number of contexts (Ramos, 2012). About 3,000 responses were submitted as part of a global survey, and an additional 44 specialist stakeholders participated in a series of focus groups in the South East Asia and the Pacific region.

Key findings included:

- Analysis of research results helps improve the uptake of science and technology information, and is more valuable when provided by a trustworthy, authoritative and accurate source.
- Nearly 70% of respondents used evidence for personal knowledge building, rather than only for delivering specific activities. This correlates to research carried out by IDS2 as well as Weiss's seven models of research utilisation, where information is used when it has been internalised and in most cases repackaged to serve a specific action or purpose. The importance for science communicators or knowledge brokers is the ability to facilitate this internalisation and conversion of information.
- Lack of human or financial resources was widely cited as an impediment to uptake, indicating that more resources should be allocated and capacity building efforts put in place to help the uptake of evidence, especially at organisational and national levels.
- Lack of analysis of the economic implications of research was identified as the next most important impediment to uptake, followed by a lack of social analysis (i.e. the impact of research on certain groups).
- Over 60% of policymakers resorted to personal knowledge and experience when sourcing evidence for policy development. Around half of all the public sector respondents used media organisations and the public perception as common sources of information for policy development. In addition, policy stakeholders seem to also rely heavily on government sources of information, which reminds us of the need to ensure that S&T is not politicized.
- In the majority of the regions it was found that evidence is integrated or used rather late in the policymaking process — i.e. 'once a decision is made, as supportive documentation and justification' and 'post-implementation, to assess impacts and outcomes of policy'. In addition, about 40% of policy respondents also felt that 'significance of views held by policy participants and other stakeholders are generally more influential than findings presented in papers for policymaking. Further, more than 60% of respondents involved in policymaking and lobbying felt that 'economic stakeholders dominate or influence more successfully', that there is a 'lack of formal avenues for involvement in policymaking and development', and that the 'nature of policymaking does not favour uptake'.

These findings imply much decision-making is currently based on perceptions or current knowledge of decision makers and therefore is prone to bias (Kahneman, 2011) and there is relatively little systematic use of evidence in decision-making processes. High priority should therefore be accorded to capacity building and facilitation in the area of evidenced-based and structured decision-making (e.g. Gregory et al., 2012), most especially for decision makers at institutional and government level.

Monitoring biodiversity in agricultural ecosystems

The biodiversity of crops, farm animals, aquatic organisms, micro-organisms and invertebrates –thousands of species and their genetic variability – make up the web of biodiversity in agricultural ecosystems that underpins the livelihood strategies of small scale farmers and continues to be fundamental in trying to achieve global food security. Global information monitoring initiatives and databases are for crop, livestock and aquatic genetic resources were not included in the inventory synthesis but are provided in Appendix 7 Information systems for biodiversity in agro-ecosystem. The FAO Commission on Genetic

Resources for Food and Agriculture oversees and guides the preparation of global assessments of genetic resources for food and agriculture. These assessments are based on information and databases from FAO's member states. Data and indicators focus not only on monitoring the extent and distribution of genetic resources in agricultural production systems but also the links of this diversity to support global food security and sustainable development, benefit sharing protocols and mapping this information to provide informed policy recommendations.

Gaps and opportunities

Combining the evidence from the review of monitoring initiatives in agro-ecosystems and livelihoods with insights from other related fields led to identification of a number of gaps and opportunities, summarized below.

Conceptual model

The quality of conceptual models varied considerably among initiatives. Donors and program leaders must develop clear conceptual frameworks that define the frequency of monitoring and how the set of monitoring indicators address the relationships between objectives. Such a framework must also define thresholds and responses to indicator data so that implementers translate monitoring into responses.

Conceptual models capture the current state of understanding of the system being monitored and can reveal critical leverage points where improved monitoring can help improve performance. For example, these insights can help guard against design problems (Olsen et al., 1999) such as lack of measurement selectivity (e.g. a measurement made at the individual level may not well represent an ecosystem property of interest that is at a higher level of scale). A number of reviews identified having a good conceptual model as a prerequisite for development of meaningful indicators (e.g. sustainable development indicators). For example, Lindenmayer and Likens (2011) emphasize the need for appropriate conceptual models of the target agricultural ecosystems, key ecological processes (including populations of particular elements of the biota), and regionally relevant agricultural practices.

Objectives, audiences and decision-making

Defining clear objectives seems an obvious step but decomposing objectives into constituent parts that are directly observable, less uncertain or easier to measure usually requires further attention than is often given (Hubbard, 2010). When relating research outputs to outcomes, a chain of means-ends objectives are usually involved, ranging from proximal to more distal factors, and indicators of the degree of achievement of objectives need to be pitched at a level in the hierarchy that matches the decision context (see Keeney, 1992, Gregory et al., 2012). This is particularly pertinent to the link between research outputs, research outcomes, system level outcomes, and development impacts (e.g. CGIAR, 2011).

Despite investment of massive amounts of resources, experience in sustainable development over several decades shows that there is considerable uncertainty on the impact of indicators and monitoring initiatives on decision-making or management, except when initiatives have been explicitly designed to inform specific decisions (e.g. regulation).

Monitoring initiatives are often designed in "passive mode" in the expectation that the collected information will be useful for informing a number of decisions, as opposed to an "active mode" where the measurements are specifically designed to directly inform specific decisions. For example a number of initiatives aimed to learn about the status of ecosystems or livelihoods, but with little clarity on how that information would actually be used to change decisions or behaviours, beyond creating general awareness raising among stakeholders (e.g. UK National Ecosystem Assessment).

Wider use of structured decision-making and decision analysis approaches (beyond simply holding stakeholder consultations on priorities) could play an important role in focusing monitoring efforts and increasing cost effectiveness. There is a strong case therefore for putting decisions before measurements. Experience shows that without formal analysis of uncertainties in decisions, the intuition of decision makers or researchers will be insufficient to identify indicators that have high value for improving decisions (Hubbard, 2010). Experience in the public sector (Seddon, 2008) has revealed that much current monitoring actually hampers delivery of services, rather than help to give insights into how to improve performance.

A clear idea on the role of different actors in terms of intended audience for the results, who are the direct users, and who will implement the initiative has important implications for the design of the initiative, including features such as the level of complexity, skill levels required, communication strategy, and sustainability pathway. Much of the data collected in monitoring systems is not used—it is thus important to identify the value/stakeholders/users early on to ensure that they have a stake and own the data that is being collected both to improve performance of the monitoring system and to ensure that the data is used. Program designers must ask the question “How will this monitoring data be utilized?” to ensure that monitoring systems are not an unnecessary cost. Indicator systems can prompt implementers to focus on outcomes at all stages of their work.

The procedures used to consult stakeholders on the design of monitoring initiatives is often of poor quality and there is much scope for drawing on structured elicitation approaches commonly used by decision analysts to improve the quality of inputs (e.g. structuring objectives).

Inference space, sampling designs & statistical analysis

Defining the scope of inference, both in terms of spatial and temporal extents, is critical but not often adequately considered. As a consequence it is not clear to which populations or extents that inferences made from monitoring results can be applied. For example if the target is resource-poor land users in low income developing countries, then the monitoring framework needs to ensure that the scope of the design permits inference at this level i.e. things need to be sampled at that scale. Many monitoring initiatives have floundered on these first conceptualization steps. For example is the desired inference space resource-poor land users in developing country tropics, or farming households in a district in Kenya? Extrapolating results beyond the sampled inference space is dangerous as both conditions and relationships between variables are likely to differ when moving into new areas.

Statistical design considerations are often the weakest of all components of monitoring initiatives and early involvement of statisticians is vital. A drive towards participatory monitoring schemes has often led to statistical design being neglected. A failure to link the sampled population to the scope of inference is a frequent problem. Even where purposive sampling (e.g. use of sentinel sites or units) is used, as opposed to random sampling of a defined population, there still needs to be a good quantification of what the units represent relative to the conditions associated with the wider scope of inference.

A lack of clear definition of sample units has plagued many studies, especially in relation to monitoring of ecosystem health and farm system attributes. For example is the sample unit a fixed area of land, a farm field, a farm, a catchment, a person, a household, a village, or a country? Inadequate attention has also been given to key stratification variables or covariates that are known to influence outcomes (e.g. gender, resource endowment). Consulting standard statistical texts on sampling techniques (e.g. Cochran 1977; Webster and Lark, 2013) is recommended.

Incorporating scale hierarchy is particularly critical for sustainable agricultural intensification and needs to be explicitly built into the sampling design and statistical analysis methods. Tools for doing this, for example through multilevel sampling, and use of mixed effects statistical models, have recently become more easily accessible.

There is need to separate objectives of monitoring initiatives that are designed for tracking trends in states of agro-ecosystems and their drivers from types of monitoring studies aimed at intervention evaluation. For the latter objective there has been rather limited use of study designs that deploy controls and counterfactuals, even though this is standard practice in the public health sector. Much intervention testing in agro-ecosystems research has centred on disparate case studies and a lack of standardization of methods and reporting that has precluded meta-analysis and the drawing of generalizable conclusions across tropical agro-ecosystems.

Measurements and indicators

The choice of variables and design of indicators has too often been divorced from consideration of objectives, value of additional information, sample units, and measurement methods and yet these are inter-related, requiring an iterative approach among the design steps in Table 1. For example, a greenhouse gas flux, such as nitrous oxide emission, is a fast variable that fluctuates rapidly over very short periods of time and short distance in space; it may need to be monitored continuously in time steps of minutes using sophisticated equipment over several years to obtain reliable data, and is thus an extremely expensive measurement. The number of sample units that can be cost-effectively monitored is extremely limited and this severely limits the ability to make wider inferences. A Value of Information Analysis might reveal that the uncertainty in this variable for a given decision context could be reduced through a much simpler proxy measurement or estimation and obviate the need for flux measurements. For example the potential for nitrous oxide emissions may be limited by soil nitrogen availability in the systems under consideration. Many ecosystem services are relatively fast variables and there is often a need to identify slower functional properties that generate the services to be able to reliably track time trends.

Monitoring indicators from different disciplines such as genetic resources, greenhouse gas emissions, and socioeconomic livelihood indicators need to be analysed in an integrated manner. This means that the conceptual frameworks utilizing such data to monitor specified outcomes must address how such indicators can be related to one another for decision-making purposes. Additionally, the routine collection and analysis of multiple indicators must be designed both to efficiently collect all necessary data and to maintain understanding of the conceptual relationships between indicators at all stages of collection, analysis, and response.

Scientists designing monitoring initiatives should critically evaluate the very purpose of why measurements are being made in the first place (Hubbard, 2010). This implies that we must have some prior knowledge or belief on what is the current level of uncertainty – perhaps we do not need more certainty to make a rational decision. It would thus seem critical to represent our prior uncertainty before embarking on new measurements, and to evaluate what will be the gain from further reducing uncertainty. This is the essence behind Bayesian decision theory and Value of Information Analysis. The principle is that it is the consequences of uncertainty that matter, rather than the precision with which quantities can be estimated (Briggs et al., 2006).

Guidelines for interpreting indicators are often lacking. For example soil organic carbon is often cited as an indicator of soil health but it is difficult to define what is a ‘good’ or ‘bad’ level of this indicator at a given location or for a given soil type, let alone critical thresholds linked to management recommendations. Again this is due to a lack of defining the decision in the first place.

One of the outstanding failures of international agricultural research and natural resource management has been the lack of development and widespread adoption of standardized measurement protocols. This has prevented aggregation and meta-analysis of results and has severely hampered the development of generalizable knowledge and provision of a reliable picture of state and trends. This failure is particularly evident in areas such as land degradation assessment but is general to environmental condition. There are important lessons to be learned from the public health sector, where meta-analysis is the foundation for informing public policy and practice. However this situation is partly due to research-led activities inevitably focusing on something new. Wide scale monitoring is much less exciting and more difficult to fund than taking the first measurements of something, or doing the strategic measurement necessary to discover principles for wider application.

Linking biophysical with socioeconomic variables

Most monitoring initiatives focus either on biophysical variables, such as soil health, or socioeconomic variables, such as poverty, calorie intake or health indicators. What remain under-studied, under-monitored and under-assessed are meaningful linkages and causal relationships between biophysical variables and socioeconomic variables. Efforts to improve this situation are visible in the global information initiatives linked to the FAO Commission on Genetic Resources that are used to prepare the State of the World Reports on plant, animal, aquatic genetic resources for food and agriculture. These initiatives monitor not only information provided by countries on the amount and distribution of genetic resource but also social and economic variables related to the role of this diversity for sustainable livelihoods.

An additional effort to improve this situation is made through the CCAFS sentinel sites that try to identify which climate smart agricultural interventions work best for the specific biophysical and socioeconomic environments. However, it has proven to be extremely challenging to link biophysical monitoring initiatives (that might be deployed on a grid of 10 km X 10 km) with socioeconomic variables and human outcomes as the population density might be too low for meaningful statistical analysis or because other stratifications would be more insightful for population-nutrition-gender-health aspects. For example, for nutrition-related interventions, it is important to sample children in the household and particularly young children, as interventions are less likely to be effective in older children. However, it is laudable that first efforts are undertaken in this area. A different, but similar linkage effort is being undertaken by HarvestChoice (http://harvestchoice.org/). Moreover, human behaviour is part of what makes the “bridge” between biophysical processes and socioeconomic outcomes. Thus, indicators need to get at the factors that affect that behaviour, for example, gender.

While remote sensing has done much towards reducing the costs of environmental monitoring, ground truthing and validation remain costly, and linkage with socioeconomic variables remains spotty. Similarly, LSMS, DHS and other panel household data sets have provided valuable information on how people’s wellbeing is changing, but the information is seldom collected in a way to inform direct policy processes in agriculture and it remains challenging to link environmental outcomes to these otherwise rich datasets.

It is important to understand how objectives across the agriculture-environment-human outcome sphere are interlinked when identifying indicators. A main lesson is to make sure from the start that indicators from different disciplines are meaningfully integrated and can be analysed together.

It is clear that over the last 10-15 years, the number and depth of efforts toward improving monitoring of both biophysical and socioeconomic indicators has substantially increased. Despite this mushrooming of new initiatives to monitor the health and wellbeing of both ecosystems and the people who depend on these systems, we have a long way to go toward affordable, effective and efficient monitoring initiatives that are satisfactory on both the biophysical/environmental and socioeconomic front.

Representing uncertainty

Consideration of uncertainty is key to problems of agricultural intensification and poverty alleviation (e.g. food insecurity, yield variability, extreme weather events, pest and disease outbreaks, income stability) and its consideration (Lindley, 2007) is also critical for reliable detection of change in indicators (with time or levels of another variable or treatment). The common failure to consider and adequately represent uncertainty is one of the most significant shortcomings of most monitoring initiatives to date. An analysis of the value of information at the outset can focus monitoring on a few, economically justified measurements and save much cost and time.

Trade-offs

The lack of adequate consideration of how to deal with trade-offs is a major flaw in most initiatives and yet is critical to considerations of agricultural intensification and sustainable development. There is need to distinguish material trade-offs (e.g. crop productivity at the expense of biodiversity) from preference trade-offs (e.g. how to weight one objective versus another, time and risk preferences). Preference trade-offs are implicit in any decision and quantifying the preferences of different stakeholders is key to improving the quality and transparency of multi-stakeholder decisions (Gregory et al., 2012). For example, in some cases parties believe they have competing goals when, in fact, they actually have contrary expectations of observable outcomes. For example downstream water users may start out with a perception that investments in agriculture upstream may reduce downstream water availability, but exposure to research findings that suggest the opposite effect may bring preferences of upstream and downstream users into harmony. Quantitative models can bring clarity to this process.

While statistical tools for temporal and spatial analysis and for explicitly representing uncertainties are becoming more accessible and widely used, consideration of trade-offs among multiple objectives requires development of some type of value model. In fact only maximization of expected utility leads to sensible decision-making (Lindley, 1985). Trained decision analysts are needed to elicit information about value judgements from relevant stakeholder groups in a careful and reasoned way, however in practice value models are often constructed in a superficial way and then give arbitrary results. In summary, monitoring initiatives must tackle the modelling preference trade-offs if they are to connect with decision-making.

Data sharing

Widespread access to computing and low-cost connectivity is transforming the way science for development is conducted (Ballantyne et al., 2010). Advances in web services, applications programming interfaces, cloud computing and automated workflows are enabling researchers to explore massive datasets and cooperate in new ways. Meanwhile, rapid developments in digital platforms and interfaces and open standards that facilitate interoperability across systems are providing new opportunities for universal access to science data products, tools and information. Mobile phone technology is opening up possibilities for two-way data and information flow with resource-poor land and water users in remote areas. A key challenge is to harness these advances for both accelerated scientific progress and effective decision support for stakeholders at different levels. An example of use of social media analytics in evidence-based decision-making is given in Appendix 8 UN Global Pulse.

Despite the opportunities for data sharing, much data and information generated by research and government institutions (e.g. precipitation levels, land use maps, soil maps, household survey data, etc) are effectively inaccessible and/or have substantial costs associated with access. The barriers are diverse and relate to inadequate national and organizational policies, to individual scientists' perceptions and to inadequate technical capacities in the use of digital media. However efforts, for example by the FAO, have been made to make available data more accessible.

While the overall situation remains somewhat bleak, new data sharing and open source policies are continuously being developed. An example is the World Bank Open Data work (http://data.worldbank.org/). There is increasing use of open source licences, such as under Creative Commons licensing (http://creativecommons.org), as scientists and agencies realize that such approaches can actually increase their impacts and lead to revenue streams linked to applications as distinct from provision of raw data. In some cases advances in remote sensing (e.g. of groundwater levels) may leapfrog over the need for extensive ground monitoring networks, but judicious ground calibration or truthing is often still required.

A number of recent international environmental laws emphasize the importance of protecting the rights of traditional knowledge holders, particularly where that knowledge concerns biological resources, and sharing benefits associated with the uses of that knowledge and associated diversity[5]. The international community developed these standards out of recognition that traditional knowledge holders (among them farmers, indigenous peoples, fisher folk, livestock herders, etc) were not being provided with incentives to share knowledge, and that important knowledge related to the conservation and sustainable use of biological resources was being lost. Developing mutually acceptable agreements with traditional knowledge holders, whereby they agree to share important information related to resources, environment, ecosystems, on terms that are acceptable to them, remains a gap in the field of agricultural and ecosystem sciences.

It is important for donors to distinguish between intellectual property concerns such as those mentioned and public-goods such as national survey data funded, collected, and distributed through public institutions. Most institutions still view public-data-sharing as an aspiration or exception rather than the de facto approach to data-collection and distribution. Donor

[5] The Convention on Biological Diversity, Article 8 (j); "Each Contracting Party shall, as far as possible and as appropriate: subject to its national legislation, respect, preserve and maintain knowledge, innovations and practices of indigenous and local communities embodying traditional lifestyles relevant for the conservation and sustainable use of biological diversity and promote their wider application with the approval and involvement of the holders of such knowledge, innovations and practices and encourage the equitable sharing of the benefits arising from the utilization of such knowledge, innovations and practices; and the Nagoya Protocol Article 7. Access to Traditional Knowledge Associated with Genetic Resources which states that: "In accordance with domestic law, each Party shall take measures, as appropriate, with the aim of ensuring that traditional knowledge associated with genetic resources that is held by indigenous and local communities is accessed with the prior and informed consent or approval and involvement of these indigenous and local communities, and that mutually agreed terms have been established."

(a) Community protocols in relation to access to traditional knowledge associated with genetic resources and the fair and equitable sharing of benefits arising out of the utilization of such knowledge;

(b) Minimum requirements for mutually agreed terms to secure the fair and equitable sharing of benefits arising from the utilization of traditional knowledge associated with genetic resources; and

(c) Model contractual clauses for benefit-sharing arising from the utilization of traditional knowledge associated with genetic resources.

policies need to continue the trend towards requiring public sharing of data generated through their grants.

Value of information, cost-effectiveness and sustainability

Data on costs and benefits of monitoring are difficult to find and there appears to have been little effort dedicated to evaluating the impact and cost-effectiveness of monitoring initiatives. This is surprising given the high and sustained costs associated with monitoring. Coupling the lack of solid evidence for impacts of past monitoring initiatives on decision-making and their cost-effectiveness, it is perhaps not surprising that few initiatives have actually been successfully sustained. It is both difficult to value the information that is being collected; and often its final uses are not known at the outset. As such it remains a major challenge to assess the cost-effectiveness of monitoring frameworks, particularly those that focus on environmental metrics that have no direct market value.

We recommend linking monitoring to specific decisions and conducting Value of Information Analysis to economically justify future monitoring initiatives. This allows systematic identification of what are the high value measurements, as well providing information on the economic value of conducting further measurements and how much expenditure can be justified on making them (see Applied Information Economics).

Budgetary resources tend to be limited for monitoring systems—thus “reaching for the stars” with perfect monitoring systems is likely to fail, unless resources are abundant. An implicit consequence of budgetary constraints is that there is little investment in the continuity and development of sustainable monitoring systems that endure beyond the grant cycle. Institutions responsible for collecting, maintaining, and disseminating statistical information cannot sustain monitoring systems and baseline data when grants do not provide for the complete life cycle of costs required. Consequences of this project-based funding structure are unmaintained websites, out-of-date or static data repositories, high staff turnover, and the routine loss of institutional knowledge and capacity.

Communicating results, feedback and language

Most monitoring frameworks use websites for communication of results and several invite feedback. Other means of communication used are policy briefs, participation in international conferences and scientific publications. Both biophysically oriented monitoring initiatives and those focused on outcomes for poor people also use various means to influence global policy processes, such as the CBD, the Rio+20 process, among others. Although this review focuses on databases available in English, French and Spanish, there is a need to include monitoring initiatives and metrics of relevance to low- and middle-income countries in the UN languages, Chinese, Russian and Arabic to complete the picture and provide recommendations for feedback systems. Although, we expect that many of the weaknesses and principles identified will hold for those initiatives as well.

The lack of solid evidence for impact of monitoring initiatives on decision-making, and the fact that most development decision-making relies on scientific evidence to only a limited degree (SciDev survey), together suggest the need for a strong capacity building initiative in evidence-based decision-making. Given the importance of uncertainty in agro-ecosystems and sustainable development, strengthened capacity in decision-making under uncertainty would seem to be of the utmost importance.

Conclusions

The purpose of this review was to identify lessons and opportunities for the derivation and use of data from monitoring initiatives in the sustainable intensification of agriculture. Based on literature review, a conceptual framework was initially developed for viewing monitoring systems looking at the interface between agricultural productivity, ecosystems and poverty and nutrition outcomes. The framework was represented as a set of 34 criteria against which monitoring initiatives that had sound conceptual frameworks were assessed. A total of 103 monitoring initiatives were identified, from which a subset of 24 initiatives was selected for detailed review against the full set of criteria. The mapping of the initiatives against the criteria is presented in spreadsheet form. The criteria included aspects of (i) conceptual framework, objectives, scope, and links to decision-making, (ii) statistical sampling frames, design of indicators, trade-off analysis, and interpretation, (iii) stakeholder involvement, data sharing, communication of results, and evaluation.

Insights from monitoring experience in other fields and from an external review group were also integrated into the assessment and contributed to the synthesis of lessons learned and opportunities for the design of data systems to measure sustainable intensification of agriculture.

Lessons learned

The evidence points to common weaknesses in a number of areas of design, although emerging monitoring initiatives are beginning to address some of these deficiencies. The areas for improvement identified are especially critical for monitoring initiatives for sustainable intensification of agriculture, where linkages between scales and between human behavioural and biophysical factors are so important. The key needs identified, are for:

- A clear conceptual framework to demonstrate an understanding of the system under study. In particular theories of change on how the monitoring results would affect behaviours and explicit linkage to specific decisions are weak or lacking.
- Clear definition of the target inference space (geography, population) and how that is sampled. This is critical for making sound inferences from the monitoring results in terms of their wider applicability.
- Well-defined sample units or strata. It should be clear hoe units represent a sample of a larger area for which inference is desired.
- Consistent and well-documented measurement protocols, so that there is opportunity for aggregation and meta-analysis of results towards the development of generalizable knowledge and provision of a reliable picture of state and trends.
- Build scale hierarchy explicitly into the sampling design and statistical analysis methods, which is particularly critical for decision-making on sustainable agricultural intensification. Tools for doing this, for example through multilevel sampling, and use of mixed effects statistical models, have recently become more easily accessible.
- Determined efforts to integrate biophysical and socio-economic indicators both conceptually and in sampling frames. A particular challenge is how to link sampling units used in biophysical monitoring initiatives (e.g. fixed area sampling or watershed delineations) with units commonly used in socio-economic monitoring (e.g. households, villages).
- Designs that allow attribution of impacts of interventions. Use statistically sound study designs where possible. Disaggregate indicators across different levels of important conditioning variables (e.g. by gender, income group). Monitor variables along the impact pathway to accumulate evidence of intervention impacts.

- Link choice of variables and indicators to objectives, value of additional information, sample units, and measurement methods. Provide guidelines for interpreting indicators for management or policy decisions.
- Represent uncertainty, both conceptually and in communicating results. Make trade-offs trade-offs among objectives explicit and separate material from preference trade-offs.
- Make data and information generated by research and government institutions accessible and reduce costs associated with access.
- Put in place active mechanisms for dissemination of results to target audiences, beyond web-based dissemination.
- Collect relevant data to be able evaluate the impact and cost-effectiveness of monitoring initiatives, to help make a better case for sustaining initiatives.

An over-riding lesson is that despite investment of massive amounts of resources in monitoring initiatives, there is a surprising lack of evidence for their actual impact on decision-making or management, except in cases where initiatives have been explicitly designed to inform specific decisions (e.g. regulation). This may be a principal reason why few monitoring initiatives are actually sustained.

Opportunities

There is enormous interest in monitoring initiatives, as indicated by the large number of initiatives identified that are relevant to sustainable agricultural intensification (103 in English language alone), and significant recent new investments in monitoring systems in data sparse countries. There are important opportunities for increasing the returns on these investments by better integrating monitoring systems with development decision processes and thereby increasing impacts on development outcomes. Useful insights into how this could be achieved were gained by drawing on experience in public health surveillance, systems thinking in industry and public services, and decision sciences. These opportunities are summarised below in terms of new thinking versus common existing practices:

Existing situation	New thinking
Measurements and indicators are designed first and then attempts are made to apply them to influence policies and practices	Decisions are identified first and then measurements are designed to reduce critical decision uncertainties
Current knowledge and beliefs are not formally represented	Current knowledge and beliefs, including our degree of uncertainty, are represented in quantitative models
Impact pathways or theories of change are restricted to mental models or simple qualitative descriptions	The impact pathway is forecast in a quantitative model, with uncertainties represented
Uncertainty is largely ignored	Representing uncertainty is seen as fundamental to dealing with issues such as food security, adoption, resilience, and economic evaluation of information
Variables that are seemingly intangible or that appear too difficult to measure (e.g. environmental services, equity) are ignored	All variables that are perceived to be important to a problem are modelled, and ways of measuring them are found if they are of economic importance to a decision

Ex post impact assessment is used as the main justification for current investment decisions	Quantitative *ex ante* impact assessment is used to justify current investment decisions, using estimates based on past experience
Monitoring is focused on state and trends	Monitoring is focused on identification and tracking of behavioural risk factors
Monitoring initiatives are generally not designed to provide evidence on attribution of interventions on development outcomes	Use study designs, and the monitoring of sensitive variables identified by impact forecasting models, to provide cumulative evidence on attribution.
Trade-offs are only qualitatively addressed	Material and preference trade-offs are distinguished and made transparent through quantification
Monitoring is focused on compliance with specifications, regulations and targets	Monitoring is focused on understanding how to improve performance
Decision-making is predominantly based on perceptions or current knowledge of decision makers, with little systematic use of evidence in decision-making processes	Evidence-based decision-making can be improved through capacity building in structured approaches to decision-making under uncertainty
Data remains with individual projects or initiatives	Move data onto platforms that are independent of individual initiatives to facilitate dissemination, reuse and sector-wide learning

Further recommendations for good practice in design of monitoring initiatives, and on what donors and those designing monitoring initiatives should do differently, including a proposal for new decision analytic conceptual approach, are given in Appendix 9 Detailed recommendations. The opportunities identified in this report are especially relevant to the implementation of the CGIAR results-orientated strategy (Strategic Results Framework), which emphasizes the better linking of research to international development outcomes, through the new CGIAR Research Programmes.

Bibliography[6]

Ballantyne P, Maru A, Porcari EM. 2010. Information and Communication Technologies—Opportunities to Mobilize Agricultural Science for Development. *Crop Science* 50: S-63–S-69.

Barder O. 2012.*Global development: views from the center*. Washington DC: Center for Global Development.

Beddington J, Asaduzzaman M, Clark M, Fernandez A, Guillou M, Jahn M, Erda L, Mamo T, Van Bo N, Nobre CA, Scholes R, Sharma R, Wakhungu J. 2012. *Achieving food security in the face of climate change: final report from the Commission on Sustainable Agriculture and Climate Change.* CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS).Copenhagen, Denmark. Available online at: www.ccafs.cgiar.org/commission.

Bond R, Kapondamgaga PH, Mwenebanda B, Yadav RPS, Rizvi A. 2007. Monitoring the livelihood platform: reflections on the operation of the livelihood asset-status tracking method from India and Malawi. *Impact Assessment and Project Appraisal 25: 301–315.*

Bond R, Mukherjee N. 2002. Livelihood asset status tracking: an impact monitoring tool? *Journal of International Development* 14: 805-815.

Booth D, Nsabagasani X. 2005. *Poverty Monitoring Systems: an analysis of institutional arrangements in Uganda*. Working paper no. 246. London: Overseas Development Institute.

Briggs A, Claxton K, Sculpher M. 2006. *Decision modelling for health economic evaluation*. Oxford: Oxford University Press.

Callahan JT. 1984. Long-term ecological research. *BioScience* 34: 363-367.

Cash et al. 2003. Knowledge systems for sustainable development. *Proceedings of the National Academy of Sciences* 100: 8086-8091.

Cates RW. 2011. *What kind of a science is sustainability science*? Proceedings of the National Academy of Sciences.Vol 108: 19449-19450.

Cattin MB, Andrianirina N. 2012. *Natural resources and livelihood concepts in monitoring rural development in Madagascar: issues of measurement and interpretation*.

CDC. 2012. *CDC's vision for public health surveillance in the 21st century.* Morbidity and mortality weekly report vol.61. U.S, Atlanta: Centers for Disease Control and Prevention (CDC).

Chape S, Harrison J, Spalding M, Lysenko I. 2005. Measuring the extent and effectiveness of protected areas as an indicator for meeting global biodiversity targets. *Philosophical Transactions of The Royal Society* 360: 443-445.

Chartres C. 2011. *Metrics, monitoring and incentives to support sustainable intensification of small-holder agriculture*. Beijing: IWMI International Water Management Institute.

Cochran WC. 1977. *Sampling techniques*. New York: John Wiley & Sons.

Consultative Group on International Agricultural Research (CGIAR). 2011. *A strategy and results framework for the CGIAR*. Montpellier, France: CGIAR.

[6] The bibliography contains additional relevant references that are not cited in the report as a resource for readers.

Deming WE. 2000a. *Out of the crisis*. Massachusetts, USA: MIT Press.

Deming WE. 2000b. *The new economics: for industry, government, education*. Massachusetts, USA: MIT Press.

Etches V et al. 2006. Measuring population health: a review of indicators. *Annual Review of Public Health* 27: 29–55.

European Commission. 2004. *Common indicators for monitoring rural development programming* 2000-2006.

Fritz S, Scholes RJ, Obersteiner M, Bouma J, Reyers B. 2008. A conceptual framework for assessing the benefits of a global earth observation system of systems. *IEEE Systems Journal* 2: 338-348.

Global Donor Platform for Rural Development (GDPRD), FAO and World Bank. 2008. *Tracking results in agriculture and rural development in less-than-ideal conditions: a sourcebook of indicators for monitoring and evaluation*. Bonn, Germany: GDPRD, FAO and World Bank.

Gregory R et al. 2012. *Structured decision-making: a practical guide to environmental management choices*. Chichester, UK: Wiley-Blackwell

Hàk T, Moldan B, Dahl, AL eds. 2007. *Sustainability indicators*. Washington: Island Press

Havemann T. 2009. *Measuring and monitoring terrestrial carbon: the state of the science and implications for policy makers*. Washington DC: The Terrestrial Carbon Group, Heinz Center.

Hawkesworth S, Dangour AD, Johnston D, Lock K, Poole N, Rushton J, Uauy R, Waage J. 2010. Feeding the world healthily: the challenge of measuring the effects of agriculture on health. *Philosophical Transactions of The Royal Society* 365: 3083-3097.

Hayati D, Ranjbar Z, Karami E. 2010. Measuring agricultural sustainability. In E. Lichtfouse (ed*.), Biodiversity, Biofuels, Agroforestry and Conservation Agriculture*, Sustainable Agriculture Reviews 5. Springer Science+Business Media B.V.

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

International Institute for Sustainable Development (IISD). 2010. *Compendium of sustainable development indicator initiatives*. Winnipeg: International Institute for Sustainable Development.

Jacobs R, Goddard M, Smith PC. 2007. *Composite performance measures in the public sector.* Policy Discussion Briefing. York, UK: Centre for Health Economics, University of York.

Jackson LE, Kurtz JC, Fisher WS. 2000. *Evaluation guidelines for ecological indicators.* Report No. EPA/620/ R-99/005. Washington DC: Environmental Protection Agency.

Jaspers S, Shoham J. 2002. *A critical review of approaches to assessing and monitoring livelihoods in situations of chronic conflict and political instability*. Working paper no. 191. London: Overseas Development Institute.

Kahneman D. 2011. *Thinking fast and slow*. London: Penguin Books Ltd.

Karl J. 2000. *What is an inference space?* http://wiki.landscapetoolbox.org/doku.php/general_design_topics:inference_space.

Keeling CD. 1998. Rewards and penalties of monitoring the earth. *Annual Review of Energy and the Environment 23: 25–82*.

Keeney RL. 1992. *Value-focused thinking: a path the creative decision-making*. Cambridge, Masachusetts, USA: Harvard University Press.

Knapp AK, Smith MD. 2001. Variation among biomes in temporal dynamics of aboveground primary production. *Science 291*: 481-484.

Lindenmayer DB, Likens GE. 2009. Adaptive monitoring: a new paradigm for long-term research and monitoring. *Trends in Ecology and Evolution* 24: 482-486.

Lindenmayer DB, Likens GE. 2011. Effective monitoring of agriculture. *Journal of Environmental Monitoring* 13: 1559–1563.

Lindley DV. 1985. *Making decisions*. Chichester, USA: John Wiley & Sons.

Lindley DV. 2007. *Understanding uncertainty*. Hoboken, New Jersey, USA: John Wiley & Sons.

Masset E. 2011. A review of hunger indices and methods to monitor country commitment to fighting hunger. *Food Policy* 36: 102-108.

McKenzie N, Henderson B, McDonald W. 2002. *Monitoring Soil Change: principles and practices for Australian conditions*. Technical report, May 2002. Australia: CSIRO Land and Water.

McKenzie NJ, Dixon J eds. 2006. *Monitoring soil condition across Australia: recommendations from the expert panels*, Prepared on behalf of the National Coordinating Committee on Soil and Terrain. Expert panel reports. Canberra ACT: National Land & Water Resources Audit.

McMahon SM, Diez JM. 2007. Scales of association: hierarchical linear models and the measurement of ecological systems. *Ecology Letters* 10: 1-16.

Moldan B, Billharz S, Matravers R. Eds. 1997. *Sustainability indicators: report on project on indicators of sustainable development*. Chichester: John Wiley.

Moldan B, Dahl AL. 2007. Challenges to sustainability indicators. In: Hàk T, Moldan B, Dahl A.L eds. *Sustainability Indicators* p. 1-26. Washington: Island Press.

Nisbet E. 2007. Cinderella science: On-the-ground monitoring is unglamorous work, seldom rewarded by funding agencies or the science community. But we neglect it at our peril, warns Euan Nisbet. *Nature* 450: 789-790.

Office for National Statistics. 2012. *Measuring national well-being: summary of proposed domains and measures.* London, UK: Crown. Available online at: www.ons.gov.uk.

Olsen AR, Sedransk J, Edwards D, Gotway C.A, Liggett W, Rathbun S, Reckhow KH, Young LJ. 1999. Statistical issues for monitoring ecological and natural resources in the United States. *Environmental Monitoring and Assessment* 54: 1–45.

Pain A. 2012. *Understanding and monitoring livelihoods under conditions of chronic conflict: lessons from Afghanistan*. Working paper no.187. UK, London. Overseas Development Institute.

Pannell DJ, Glenn NA. 2000. A framework for the economic evaluation and selection of sustainability indicators in agriculture. *Ecological Economics* 33: 135–149.

Parris TM, Kates RW. 2003. *Characterizing and measuring sustainable development.* Annu. Rev. Environ. Resour. 28:559–86

Pencheon D. 2008. The good indicators guide: how to use and choose indicators. *Annual Review of Environment and Resources* 2003: 559–86.

Pereira HM et al. 2013. Essential Biodiversity Variables. *Science* 339: 277-278.

Ramos YJR. 2012. *SciDev.Net Global Review 2012*. UK**:** SciDev.Net. Available online at: www.scidev.net.

Rapport DJ, Hildén M. 2013*. An evolving role for ecological indicators: from documenting ecological conditions to monitoring drivers and policy responses.* Ecological Indicators 28: 10–15.

Ravallion M, Sen B.1996. When method matters: monitoring poverty in Bangladesh. *Economic Development and Cultural Change* 44: 761-792.

Remans R, Flynn DFB, DeClerck F, Dirus W, Fanzos J, Gaynor K, Lambrechtz I, Mudiope J, Mutuo PK, Nkhoma P, Siriri D, Sullivan C, Palm CA. 2011. Assessing nutritional diversity of cropping systems in African villages. *PLoS ONE* 6: 1-11.

Rempel RS, Andison DW, Hannon SJ. 2004. Guiding principles for developing an indicator and monitoring framework. *The Forestry Chronicle* 80: 82-90.

Renkow M. 2010. *Assessing the environmental impacts of CGIAR research: toward an analytical framework*. Mitch Renkow working paper for SPIA. Washington DC: CGIAR.

Reshef IB, Justice C, Sullivan M, Vermote E, Tucker C, Anyamba A, Small J, Pak E, Masuoka E, Schemaltz J, Hansen M, Pittman K, Birkett C, Williams D, Reynolds C, Doorn B. 2010. Monitoring global croplands with coarse resolution earth observations: the Global Agriculture Monitoring (GLAM) project. *Remote Sensing* 2: 1589-1609.

Rickard et al. 2007. Ensuring policy relevance. In: Hàk T, Moldan B, Dahl A.L. eds. *Sustainability indicators* p. 65-79. Washington: Island Press.

Rinne J, Lyytimäki J, Kautto P. 2012. From sustainability to well-being: lessons learned from the use of sustainable development indicators at national and EU level. Ecological Indicators (in press).

Rural Industrial Research and Development Corporation. 2000. *Determining the effectiveness of vegetation management programs: measures and methodologies - literature review.* Kingston: Rural Industries Research and Development Corporation.

Sachs JD, Remans R, Smukler SM, Winowiecki L, Andelman SJ, Cassman KG, Castle D, Fries R, Denning G, Fanzo L.E, Leemans R, Lehmann J, Milder CJ, Naeem S, Nziguheba G, Palm CA, Pingali PL, Reganold JP, Richter DD, Scherr SJ, Sircely J, Sullivan C, Tomich TP, Sanchez PA. 2012. Effective monitoring of agriculture: a response. *Journal of Environmental Monitoring* 14: 738-742

Sachs J. 2010. Monitoring the world's agriculture: *Nature* 466: 558-560.

Salzer E. Willoughby's. 2001. *Sample design for rangeland assessment and monitoring*. http://wiki.landscapetoolbox.org/doku.php/general_design_topics:sample_design.

Savage SL. 2009. *The flaw of averages: why we underestimate risk in the face of uncertainty.* Hoboken, New Jersey: John Wiley & Sons Inc.

Schreuder HT, Ernst R, Maldonado HR. 2004. *Statistical techniques for sampling and monitoring natural resources*. General technical report RMRS-GTR-126. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Seddon J. *Systems thinking in the public sector: the failure of the reform regime… and a manifesto for a better way*. Axminster, UK: Triarchy Press.

Tomich TP, Brodt S, Ferris H, Galt R, Horwath WR, Kebreab E, Leveau JHJ, Liptzin D, Lubell M, Merel P, Michelmore R, Rosenstock T, Scow K, Six J, Williams N, Yang L. 2011. Agroecology: a review from a global-change perspective. *Annual Review of Environment and Resources* 36: 193–222.

UNEP 2012. *Land Health Surveillance: An evidence-based approach to land ecosystem management. illustrated with a case study in the West Africa Sahel*. Nairobi: United Nations Environment Programme.

Webster R, Lark M. 2013. *Field sampling for environmental science and management*. Oxon, UK: Routledge.

Wesemael BV, Paustian K, Andren O, Cerri CEP, Dodd M, Etchevers J, Goidts E, Grace P, Katterer T, McConkey BG, Ogle S, Pan G, Siebner C. 2011. How can soil monitoring networks be used to improve predictions of organic carbon pool dynamics and CO_2 fluxes in agricultural soils? *Plant and Soil* 338: 247-259.

World Health Organization (WHO). 2002. *World health report*. Geneva: The World Health Organization.

Zhen L, Routray JK. 2003. Operational indicators for measuring agricultural sustainability in developing countries. *Environmental Management* 32: 34-46.

Acknowledgements

We acknowledge valuable comments from an external review team, including Professor Margaret Gill (DFID), Kate Schneider (Bill & Melinda Gates Foundation), Jerry Glover (USAID), Ken Cassman (University of Nebraska), Monika Zurek (Climate Focus), Cheryl Palm (Earth Institute of Columbia University), Neil McKenzie (CSIRO), Kathleen Beegle (World Bank - LSMS); as well as Andrew Noble (IWMI). We thank DFID, especially Professor Margaret Gill and Rachel Lambert for providing the opportunity to conduct this review, including financial support, and for overall direction.

Appendix 1 DFID call

Call for proposals: May 2012

Review of the evidence on indicators, metrics and monitoring systems for agriculture, poverty and environment:

Background

Agriculture needs to expand and develop in order to meet the food requirements of a growing population in the face of natural resource constraints and climate change. The Royal Society report in 2009, "Reaping the Benefits: Science and the sustainable intensification of global agriculture", identified the need for sustainable intensification of agriculture in which yields are increased but without adverse environmental impacts or cultivation of more land. The Foresight report on Global Food and Farming Futures further highlighted the need for the development of standards and metrics for monitoring the performance of agricultural systems and trade offs between productivity and environmental services and welfare outcomes. The CGIAR Science Forum 2011 identified the need for such metrics that could potentially include monitoring and certification of food production systems to ensure that agricultural systems, and their products (crops, livestock, timber, fish), are produced using practices that do not harm the environment or contribute to climate change and loss of biodiversity.

The scientific challenge is to ensure that such methods are based on good science and are unbiased. According to the Science Forum, areas where there could be harmonization of research methods and data collection include: analyses related to climate change (carbon capture and storage), soil quality, GHG emissions and nutrient losses per unit of economic yield, impact on biodiversity, contribution to local food security and nutrition, system resilience and yield stability, energy and water productivity, etc.

The ultimate goal is to provide decision-makers with tools which they can use to explore trade-offs between food security, environmental and socio-economic goals. Thus other benchmarks will also be needed for social and economic well being, including measures of poverty, hunger and nutrition. Many of these are lacking at the agricultural production system level, but to make progress in informing decision-makers, we need to learn lessons from experience to date from other sectors on what is feasible. An important caveat raised by Rosenzweig and Tubiello (2006), however, is to keep 'metrics as simple as possible – but not too simple as to become useless'.

Purpose

DFID propose to initiate a process to collate existing information through:

1) Commissioning a review of existing evidence and a gap analysis exercise.

2) Convening an external expert group, to validate the mapping and review of evidence, and identify next steps in terms of both DFID's role in supporting relevant current and future research as well as how other stakeholders, including the CGIAR, could build on the review in developing a strong scientific basis and appropriate methods to underpin the measurement of ecosystem related variables, and to analyse the tradeoffs between productivity, ecosystem and welfare outcomes from agricultural R and D. The expert group will reflect the broad geographic scope of stakeholders and research, the multidisciplinary nature of the research, and include academic, operational and policy partners.

Building on the Global Foresight report on Food and Farming Futures, and recent initiatives including the Integrated Monitoring System for African Landscapes, the purpose of the review is to:

Identify lessons and opportunities for the derivation and use of data from monitoring systems in the sustainable intensification of agriculture (i.e. enhancing productivity with less impact on scarce natural resources and which do not contribute to climate change and biodiversity loss).

In particular, this review will:

- Map out key initiatives in data monitoring systems relating to agriculture, paying particular attention to those which also acknowledge the impact on ecosystem health, and/or poverty and well-being; including their objectives, indicators and metrics used.
- Summarise any pertinent findings from monitoring of data related to ecosystem health and poverty/well-being not directly connected to agricultural systems.
- Classify the indicators identified in relation to their primary intended users (e.g. scientists vs decision-makers).
- Undertake a gap-analysis of systems, indicators and metrics, identifying strengths and weaknesses in methodology and use.
- Undertake an evidence-based synthesis of lessons based on a number of criteria including application/use at different levels, standardisation of indicators, composite metrics, coverage and replicability, validity, cost of indicator systems, level of disaggregation (age, gender, wealth etc).
- Synthesise lessons learned, guided by an expert panel, drawing on the evidence collated in the mapping and gap analysis, with particular emphasis on the indicators which would be of most use in helping to inform decision-makers on the evaluation of trade-offs.

The analysis will inform the development of any future DFID research investments and engagement with stakeholders in this area. Specifically it has the potential to inform:

- the development of systems to measure the impact of CGIAR investments (of relevance to DFID as a significant funder) at the level of the 4 system outcomes (http://consortium.cgiar.org/our-strategic-research-framework/) and to link outcomes at CRP level with higher level impacts.
- Mechanisms to analyse the impacts and trade-offs associated with sustainable intensification at different scales (sub-national, national, regional).
- Future research on value for money metrics for measuring agriculture, ecosystem and poverty and nutritional outcomes.

Other Users

This assignment is relevant to other funders of research, particularly those investing explicitly in sustainable intensification (eg. BMGF, USAID amongst others).

It has a specific relevance to the CGIAR as the new research program rolls out through a portfolio of CRPs, each contributing to the overarching Strategic Results Framework and system level objectives. Each CRP is being developed separately, which highlights the need for common measures to ensure progress and the ability to adequately compare the impact of CGIAR research on natural resource management goals across sites and regions. The Beijing CGIAR Science Forum in 2011 brought together policy makers, private sector, NGOs and researchers to identify what is needed in terms of science and research to promote metrics, monitoring and certification to support sustainable intensification of small-holder agriculture in developing countries.

Scope

Geographical

The mapping will focus on work of relevance to low- and middle-income countries. A preliminary list of relevant research programmes and institutions activities is provided in Annex 1.

Timeframe of interest

The exercise will review the evidence on monitoring systems and metrics undertaken in the last 20 years.

Sources

The exercise will cover monitoring systems in developing as well as developed countries, in order to derive evidence based lessons. It will look at data collection and metrics developed by the private sector as well as public sector bodies for analysing key ecological processes and biodiversity under conditions of agricultural intensification. It will systematically review the evidence in relation to a number of criteria including the validity and costs of such systems.

Workshop

Applicants should also present their views on how they would help DFID facilitate a half-day virtual workshop with the external group of experts (see above).

Product

The output should be a short, clear and succinct analytical report (no more than 25 pages excluding annexes) which sets out:

- A conceptual framework for viewing monitoring systems looking at the interface between agricultural productivity, ecosystems and poverty and nutrition outcomes.
- Analysis of initiatives – setting out in a coherent and accessible form the objectives, indicators and metrics and other key aspects of different monitoring systems, which operate at different scales, and which have looked at issues of agriculture, ecosystems and environment, and poverty and nutrition.
- What the evidence tells us on key issues of scale, validity, appropriateness, cost, coverage and replicability, usefulness.
- What the evidence tells us on the extent to which data/indicators that are collected allow not just aggregate pictures to be described but a disaggregated view by key socio-economic characteristics (gender, age, income group, farm size etc).
- Gaps – an analysis of research problems and questions that are not being adequately addressed in current initiatives and the strengths and weaknesses of different systems.
- Mapping in Annex – setting out the monitoring systems being undertaken: primary research questions or objectives, outputs, costs and timeframe, donors and research partners, geographical scope, indicators and metrics.
- Based on the inputs of the external group, synthesis the lessons learned and opportunities for the design of data systems to measures sustainable intensification of agriculture.

Management

The DFID Senior Research Fellow will provide overall technical support and oversight to the process, working closely with DFID's Agriculture research team. Subject to external interest, the external experts will be engaged at project conception stage to provide guidance on relevant research initiatives, to validate the project findings and provide high-level input on the way forward.

Person specifications

Specifications for this work include expertise on:

- Indicators, metrics and monitoring systems in ecosystems, environment and biodiversity;
- Indicators, metrics and monitoring systems in agriculture, food systems and poverty and nutrition.
- Development and design of monitoring systems and databases in developing countries.
- Strong research expertise including in systematic review approaches and methods.

It is envisaged that 120 people days will be required for this assignment. This is not a rigid ceiling but a guideline. Applicants are welcome to submit variations, but full justification should be given if additional effort is thought necessary.

How to apply

Applicants should submit a technical proposal with a max of 6 pages, with short CVs in a separate Annex (details on length in Annex B), plus a financial proposal showing days input for each team member, daily fee rates along with clear costs for any relevant expenses. All financial information should be presented in Sterling (£). The bid questionnaire at Annex B may be used as a guide for structuring your proposal. Further information may also be found in the ITT Instructions document.

An electronic version of your application (in Word 2003 format please – **not** a PDF) must be received no later than **14:00, UK time, on Tuesday 5th June 2012**. Bids received after the due time and date will not be considered.

Bids should be emailed to EvidenceReview@dfid.gov.uk and file names must include the name of the bidding organisation (e.g. ABC-bid.doc). Receipt of all bids will be acknowledged.

DFID will provide some feedback to all successful and unsuccessful applicants dependant on the level of response that we receive to this call. At any stage of the process, before the award of contracts, DFID reserves the right to annul the tendering process and not make any award should there be no suitable tenders.

Indicative timetable

Activity	**Deadline**
Deadline for submission of bids	5th June 2012
Evaluate Proposals	19th June 2012
Supplier selected	26th June 2012
Issue Contract	2nd July 2012

Contract Starts	asap

Evaluation process and Scoring Methodology

Proposals will be evaluated using the following scoring methodology:

6	Excellent, addresses the requirements of the ToR and all ITT issues, and where relevant demonstrates fine tuning, to make a match with Client expectations, and is of a quality and level of detail and understanding that provides confidence in certainty of delivery and permits full contractual reliance (where applicable)
5	High degree of confidence that they can meet the requirements of the ToR (and where relevant strong evidence they have tailored their response to meet these). Demonstrates they have a thorough understanding of what is being asked for and that they can do what they say they will; translates well into contractual terms (where applicable)
4	An understanding of all issues relating to delivery of the ToR and tailoring the response to demonstrate that proposals are feasible so that there is a good level of confidence that they will deliver; can be transposed into contractual terms (where applicable)
3	Understands most of the issues relating to delivery of the ToR and addresses them appropriately with sufficient information, but only some relevant tailoring and so only some confidence that they will be able deliver in line with expectations
2	Some misunderstandings of the issues relating to delivery of the ToR and a generally low level of quality information and detail. Poor appetite to tailor when asked and so fails to meet expectations in many ways and provides insufficient confidence.
1	ToR issues are scantily understood and flimsy on quality information, with minimal tailoring if anywhere relevant. Provides no confidence that the issues will be addressed and managed at all in line with expectations
0	Complete failure to address the requirements of the ToR.

The above scoring methodology will be applied to each of the Criteria detailed on the table below. The Total Score for each Criteria will comprise of the score awarded (0 to 6) multiplied by the weighting allocated to each Criteria.

Evaluation Criteria

The evaluation criteria and weightings that will be applied to this tender are detailed below.

Criteria	**Weighting**
Previous relevant expertise	10

Technical proposal to undertake the work	30
Skills, expertise of the team and how they propose to work together	35
Evidence of familiarity with existing initiatives	10
Commercial	15
Overall Total	100

Contacts and guidance

DFID is unable to respond to individual enquiries about the tender. Any questions emailed to EvidenceReview@dfid.gov.uk will have responses posted at: http://www.dfid.gov.uk/What-we-do/Research-and-evidence/How-we-do-research/Funding-opportunities/Evidence-Review-competition/

Annex A

Provisional (but by no means exclusive) list of relevant monitoring initiatives across agriculture, food security and ecosystems

UNEP World Conservation Monitoring Centre	
Earth Institute, CIAT-TBSF, ICRAF	
CarboAfrica	www.carboafrica.net
African Soil Information Service (AfSSI)	www.africasoils.net
Bioversity International GEF project	www.bioversityinternational.org
Consortium for Improving Agricultural Llivelihoods in Central Africa (CIALCA)	www.cialca.org
Integrated Monitoring System for African Landscapes (Conservation International with Earth Institute, CSSR)	
Millennium Ecosystem Assessment	
Intergovernmental Platform on Biodiversity and Ecosystem Services'	http://www.ipbes.net/
Landscapes for People, Food and Nature	http://landscapes.ecoagriculture.org/documents/global_review_flyer

Global Landscape Initiative	
The Economics of Environmental Systems and biodiversity (TEEB)	
Natural Capital Project	Keveira et al 2011
GEOSHARE	
GEO-AG	Scholes et al 2008
Tropical Ecology and Assessment Monitoring (TEAM)	www.conservation.org
Committee of sustainability assessment (COSA)	
GECAFS	
Earth System Science Partnership	http://tropi-dry.eas.ualberta.ca www.agrobiodiversitas.org www.globalandproject.org
China National Soil Fertility Fertiliser Efficiency Long-term Monitoring Network	
LSMS-ISA	
Global Foresight report on Food and Farming Futures	Masset, E, 2010, Food Policy A few of hunger indices and methods
Harvest Choice	
OPHI/IFPRI/USIAD Women's Empowerment index	
FAO	
IFPRI Global Hunger Index	
FEWSNET	
LSMS-ISA	http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,contentMDK:22287818~menuPK:6194952~pagePK:64168445~piPK:64168309~theSitePK:3358997,00.html
DHS surveys	

Appendix 2 Scoring of Initiatives

Spreadsheets giving the scoring of the individual monitoring initiatives against the criteria are give in the accompanying Microsoft Excel files:

Scoring_Monitoring_Initiatives.xlsx

Scoring_Monitoring_Initiatives_Detailed.xlsx

Downloaded files on individual initiatives are given in the accompanying zipped folder *Monitoring Initiatives.zip.*

Appendix 3 Narrative of detailed reviews

Agriculture & Ecosystems

UK National Ecosystem Assessment

The objectives of the assessment were (1) To produce an independent and peer-reviewed assessment of the state and value of the UK's natural environment and ecosystem services; (2) Identify and understand what has driven change observed in the natural environment and the services it has provided over the last 60 years, and what may drive change in the future; (3). Foster better interdisciplinary cooperation between natural and social scientists to assist in strengthening policy making in order to ensure effective management of the environment and ecosystem services in the future; (4). Ensure full stakeholder participation and encourage different stakeholders and communities to interact with each other; (5). Use the key messages from the assessment to raise awareness among society of the importance of the natural environment to human well-being and economic prosperity.

The primary users of the national ecosystem assessment are the UK government. However the involvement of a wide range of public, private and third sector decision-makers and stakeholders through a User Group, as well as a range of wider stakeholder meetings, helped to shape the assessment process and ensure that the outputs are relevant for a variety of different audiences. The 300-plus authors involved, managed by a group of coordinating lead authors (largely natural scientists, but including economists and social scientists), were drawn from more than 50 academic institutions, together with representatives from over 15 government agencies, more than 10 NGOs and 11 private sector institutions.

The assessment is so wide-ranging that the number of indicators and types of metrics cannot be summarised and the greatest achievement of the assessment is not the monitoring of individual indicators, instead the application of the MEA framework using broad habitats as the unit of analysis and a consistent compilation of existing measurement efforts (Figure 6). No measurement protocols had been used but the assessment itself reviews knowledge gaps and acknowledged more research was necessary to develop monitoring protocols.

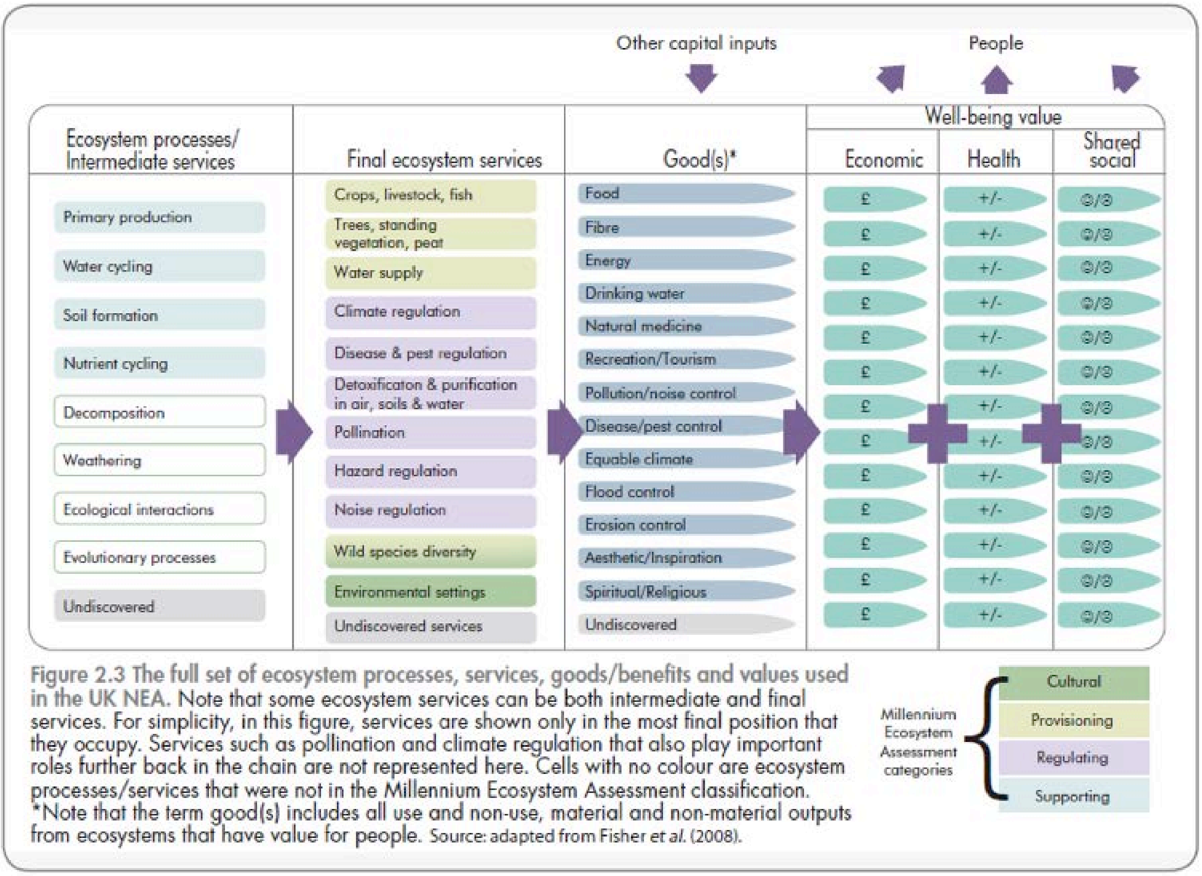

Figure 6. UK National Ecosystem Assessment conceptual framework used for organising indicators

There is an implicit trade-off analysis due to the large range of drivers, and services that are included in the assessment; this is facilitated by an economic valuation of ecosystem services. Case studies are employed to further investigate trade-offs between services.

Scenarios have been developed as a means to conduct very broad policy, macroeconomic and societal level ex ante analyses with a strong link to agricultural development

CarboAfrica

The monitoring initiative is designed to provide general recommendations on potential management of (natural) ecosystems for carbon sequestration. The initiative seeks to consolidate and expand terrestrial carbon and other greenhouse gas flux monitoring networks of Sub-Saharan Africa and ultimately to quantify, understand and predict the greenhouse gas budget of Sub-Saharan Africa and its associated spatial and temporal variability. This would lead to an assessment of the current land use change and an evaluation of the potential for carbon sequestration in Sub-Saharan Africa in the context of the Kyoto Protocol.

The greenhouse gas flux measurements are intended primarily for scientists and the main partners are European academic institutions. CarboAfrica is coordinated by the Department of Forest Science and Environment (DISAFRI) of University of Tuscia, the Max-Planck-Institute for Biogeochemistry (MPI-BGC) will coordinate site characterisation and development of model inversion tools, the Department of Physical Geography and Ecosystems Analysis (INES) at Lund University coordinates modelling interactions of climate-fire-vegetation dynamics at regional and continental scale. Communications are coordinated by Global Terrestrial Observing System (FAO-GTOS) of the Food and Agriculture Organization of the United Nations, while the *Centre de Coopération Internationale en Recherche Agronomique pour le Développement* (CIRAD) leads the evaluation of a sustainable sequestration potential in relation with the Clean Development Mechanism (CDM).

Broad ecoregions within Africa have been targeted while the temporal scope is a constant measurement regime of carbon, water and energy fluxes at specific locations. A consideration of scale hierarchy is not evident in the results that are presented, but spatial resolution is an issue in the modelling for up-scaling. Neither the objectives of the initiative nor the sampling strategy allow the evaluation of a specific agricultural research intervention although links are made between land-cover and carbon fluxes and interventions such as deforestation, afforestation or other land conversions.

The monitoring framework is very specialised and does not consider other ecosystem services nor any other conditioning factors related to agricultural development or livelihoods.

Land Degradation Surveillance Framework of the African Soil Information Service

The objectives of the surveillance framework are to assess and monitor land degradation in 60 sites across Africa to provide data for the African Soil Information Service (AfSIS) and to provide a standardised methodology for soil and land use monitoring and digital soil mapping. The surveillance (sentinel) sites represent an area that is statistically representative of about 18.1 million km^2 of sub-Saharan Africa and provides a baseline measurement of soil health; in order to become a monitoring initiative the sentinel sites would be resurveyed at 10 year intervals to assess soil condition changes and their causes, although this is outside the timeframe of the project itself, thus there are risks that monitoring would not happen.

The kinds of outcomes that are measured are limited to soil health (and crop yields) rather than livelihoods outcomes. Nevertheless the link between research (on soil fertility trials) and crop yield outcomes is explicit. In selected sentinel sites special attention will be given to key socio-economic determinants (e.g. gender and market determinants) for adoption of ISFM technologies, but these factors are not considered in the sampling framework of the sentinel sites. Diagnostic trial sites are characterised according to the farming system, and there is an effort to link the soil properties, and the results of the diagnostic trials, with household surveys which take into account farm types, resource endowment and gender, but no results are obviously available.

The initiative was designed and implemented by CIAT-TSBF, ICRAF, and the Earth Institute of Columbia University provides information to AGRA and its grantees, particularly its Soil Health Program. National research and development organization teams (NARS) work with the sentinel site survey teams and receive training in protocols. Broad based recommendations regarding options for ISFM (opportunities and constraints for larger regions based on observed general soil health status) are targeting beneficiaries such as CAADP, FARA, AGRA, HarvestChoice and the donor community. Outcomes will also inform regional organizations like ASARECA, CORAF and SADC. While national level recommendations will target ministries, national policy institutes, research organizations and other stakeholders, local level beneficiaries will include farmers and farmer groups, district officials and locally operating NGO or CBO. Recommendations for local level beneficiaries will refer to specific ISFM options, adapted to the local conditions and farmers circumstances.

The results of the monitoring are evaluated and, when combined with the results of the diagnostic trials, are transformed into recommendations for farmers and NARS. However there does not seem to be a plan for how the results could be used to modify the indicators - possibly because there is a high confidence in the utility of the chosen indicators and a need to maintain consistency in the use of indicators in order to facilitate monitoring (over time and space).

The concept of soil health is linked strongly to concepts of sustainability, especially regarding sustainable agricultural development and recommendations for farmers would involve trade-offs. The mechanisms for this trade-off analysis are less clear, with references to the NUANCES framework, but there is little evidence at this stage that the link has been made.

The project proposal claims that the soil-health surveillance service is cost-effective, which gives the impression that the trade-offs have been considered from the start. However the cost of maintaining the sentinel sites limits their number and potentially the area in Africa which is covered and the random selection of sites may cause practical problems but known security hot-spots have been avoided.

Integrated Monitoring System for African Landscapes (Vital Signs)

VitalSigns has been initiated by the Earth Institute at Columbia University, Conservation International and the Council for

Scientific and Industrial Research (CSIR) in South Africa to quantify sustainability and provide analytical tools to evaluate risks and trade-offs among agricultural development, ecosystem health and human well-being outcomes, but has yet to be implemented.

No information available about the frequency of measurements but the spatial extent is Africa with 5 'regions' mentioned as well as three specific countries where the initiative will be started - Tanzania, Ethiopia and Ghana. The initiative considers the different measurement scales of agricultural landscapes and range from the plot, household, through the landscape to a larger region.

The three institutions coordinating the initiative are seeking collaboration with a wide range of partners including governments, NGOs, the private sector and academia. The range of potential users - farmers, farmer associations, NGOs, national governments and donors - is large giving the impression that links between research outputs and outcomes could be explored, but there is currently not enough information to confirm this. An aim of the initiative is to use a dashboard-type visualisation (Figure 7) to enable decision-makers to interpret the results of the monitoring initiative, with a proposal to use 'holistic indicators' for different themes, such as a local food sufficiency index, and biodiversity health index.

Careful consideration is being given to sound design of hierarchical sampling frames (www.vitalsigns.org).

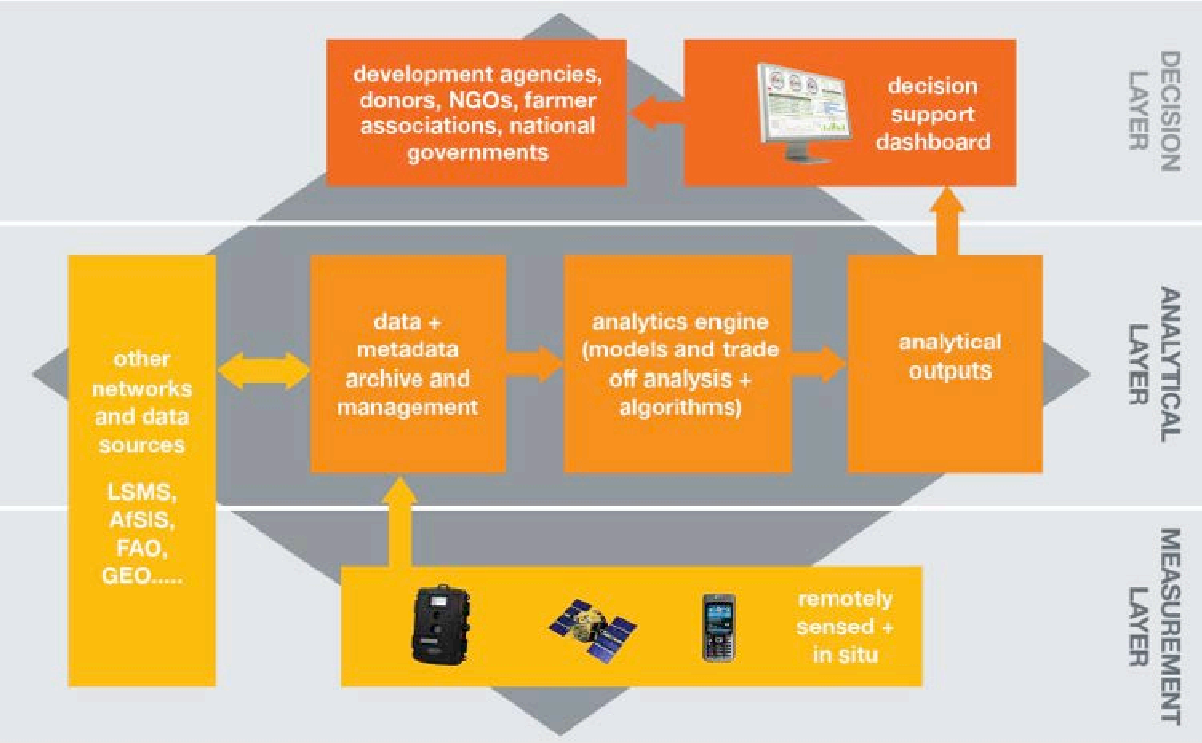

Figure 7. Conceptual model of Vital Signs showing three layers.

There is no evidence that the value of the information collected has been assessed although the cost of measuring and managing the data (1-2% of existing investment in agriculture) has been calculated.

Millennium Ecosystem Assessment

This global assessment is a compilation of existing measurement efforts and has been designed to address policy rather than research outputs and is too broad to support specific decisions. Specifically the assessment sought to evaluate the current spatial extent and condition of ecosystems as well as the quality, quantity, and spatial distributions of services provided by the systems. This implied knowledge of the human inhabitants of each ecosystem and the ecosystem services they used and the condition of these ecosystems are valued with regard to their contribution to the well-being of the global population. Despite not being a monitoring initiative the assessment also sought to document trends in ecosystem condition and their services in the recent (decades) and more distant past (centuries) and understand how ecosystem condition, and in turn ecosystem services, respond to the drivers of change for each system.

The MA was conducted as a multi-scale assessment, with interlinked assessments undertaken at local, watershed, national, regional, and global scales although the basic reporting unit is a global ecosystem (marine, coastal, inland water, forest, dryland, island, mountain, polar, cultivated and urban). The temporal scope of the assessment varies according to indicator.

The assessment brings together information from different monitoring initiatives. It uses the conceptual framework of ecosystem condition to organise the findings rather than harmonising the raw data, and there is no evidence for the use of standardised protocols and recognition that case studies use different protocols or none at all. The sampling frame reflects both the ecosystems in which ecological processes occur as well as the regional framework that allows for locally relevant policy and decision-making. However, the sampling may appear ad hoc (and potentially biased) since any institution or country was able to undertake an assessment as part of the MA if it agreed to use the MA conceptual framework (Figure 8), to centrally involve the intended users as stakeholders and partners, and to meet a set of procedural requirements related to peer review, metadata, transparency, and intellectual property rights. Multiple authors contributed to the MA, which was overseen by MA board, an assessment panel (of mainly academics) and supported by a secretariat (government agencies, UN agencies and academic departments).

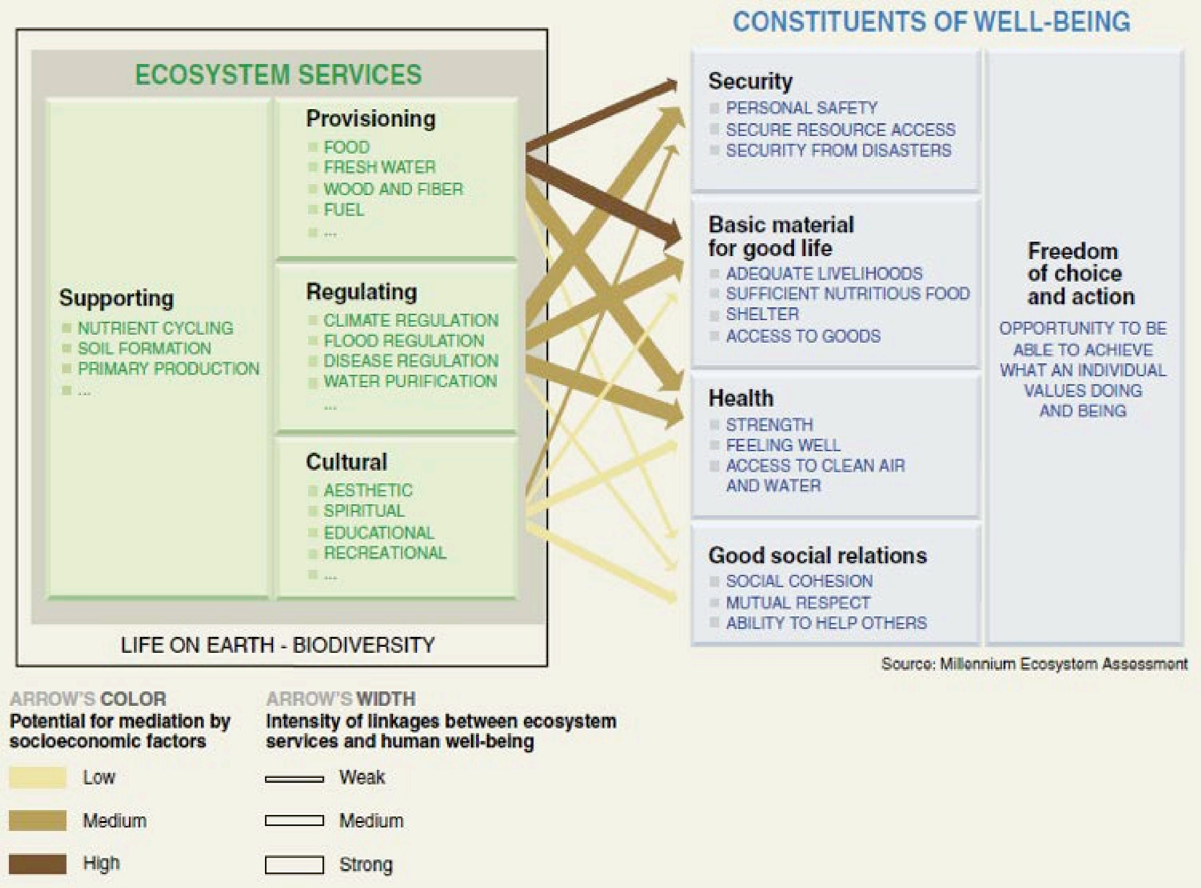

Figure 8. Millennium Ecosystem Assessment linkages between ecosystem services and human well-being.

There is no evidence of stratification, and while socioeconomic differences are considered in the links between ecosystem condition and vulnerability it is not possible to disaggregate results according to farm types or household resource endowment classes.

Core datasets chosen for consistent reporting across chapters and to allow for changes to be monitored, but data on the spatial distribution, quantity, and quality of regulating, supporting, and cultural services such as nutrient cycling, climate regulation, or aesthetic value have generally not been collected, and it was necessary to use indicators, modelled results, or extrapolations from case studies as proxy data.

The assessment discusses the need for trade-off analysis, and gives some examples of how this can be done, and how the use of common metrics can aid trade-off analysis. There is also mention of dose response (for fertiliser application) of both benefits (crop yields) and costs (nitrate leaching). The assessment also applies trade-off analysis in cultivated systems, although this analysis is based on case-studies rather than a comprehensive global analysis. Valuation of all ecosystem services (except supporting due to measurement problems) is explicitly considered and scenarios have been developed as a means to conduct very broad policy, macroeconomic and societal level ex ante analyses with a strong link to agricultural development.

Landscapes for People, Food and Nature

This initiative is conducting a global review to assess the 'track record' of EcoAgriculture experiences and the performance of an approach - integrated landscape management - in different contexts and see what works best and why. Specifically the current extent and future potential of integrated, multi-objective management of rural landscapes (EcoAgriculture and related approaches) to meet global needs for food and fuel, achieve food security and good nutrition, conserve biodiversity, and ensure the provision of critical ecosystem services. The results of the global review have not yet been released to the public but a key goal is to communicate the results "to support practitioners, policymakers, and advocates to develop and scale-up effective practices" via numerous channels including peer-reviewed journals, set of briefs for awareness-raising, outreach, and advocacy, a Landscapes manual/source book, multi-media, and curriculum materials for training practitioners and landscape leaders.

The co-organisers of the initiative are EP, Bioversity International, CI, FAO, IFAD, UNEP, WRI, ICRAF, UNU-IAS, Dutch ministry of economic affairs, agriculture and innovation. There is a larger group of advisory organisations including more local or regional organisations, and a separate group of funding organisations. It is, however, unclear, which partners will undertake the global review. The sample appears to be EcoAgriculture case studies which are defined at the landscape scale. The approach used to interpret 'landscape inventories' and potentially to design the review is the Landscape Measures Resource Center (LMRC) – a collection of ideas and tools which have been developed by EP and Cornell University (not a partner in the initiative). Most of the information about the review methods is therefore derived from LMRC documentation rather than information about the review itself. The LMRC documentation suggests that links between research activities, outputs and outcomes will be made, and apart from the question on current extent and potential for EcoAgriculture there are further issues relating to technical and policy interventions that can be addressed.

The case study approach, provided the sample is large enough, is appropriate to answer the 8 key questions that form the focus of the review. The LMRC also offers guidance for designing and carrying out the data collection. The Global Review is self-selected, with institutions requested to suggest cases for analysis; in addition there may not be any assessment of non-EcoAgriculture landscapes (i.e. a counter-factual situation). However within the case studies if the LMRC is applied then bias should not be introduced. The LMRC is clear that monitoring needs to take place at various spatial levels from the plot to the landscape but what is unclear is whether this has taken place as part of the global review. The LMRC explicitly recognises the need to take equity into account when analysing access to resources, and household surveys are mentioned as an instrument to obtain this information and to allow disaggregation but there is no information on whether the conditioning variables are used in stratification.

It is unclear whether the Global Review itself will look at change or whether repeated measurements will take place, but the need for tracking change and providing feedback to modify the theory of change is well documented in the LMRC. The breadth of indicators allows for ex post cost benefit analysis to take place with the possibility of a meta-analysis, but perhaps not best suited to provide sufficient data for an ex ante analysis. The LMRC has a guide on cost-effectiveness of measuring the indicators, which we assume will be documented for the review(s) and which would assess the trade-offs between the value of the information and the cost of measurement, however there is currently no evidence that trade-offs have been analysed.

The LMRC guidelines acknowledge that trade-offs will be made but suggest a link between the size of the landscape and whether trade-offs (involving losses of environmental functions) are acceptable. In the definition of indicators the LMRC guidelines suggest including indicators from each 'leg of the stool' (i.e. livelihoods, conservation and production) in order to facilitate the analysis of trade-offs but the LMRC framework does not explicitly deal with valuation of ecosystems services so it appears that trade-offs are negotiated.

GEOSHARE

This initiative seeks to provide data that can address issues such as climate change mitigation, environmental impacts of biofuels, offsite pollution from agriculture, and the preservation of biodiversity, among others. Some of the problems have components that are observable - such as land use and climate, and these are the main focus of the monitoring initiative. The principal actors are northern universities (Purdue, McGill, Stanford, Bonn) and CG centres (CIAT, IFPRI and IRRI). The universities are coordination and thematic leaders (economic analysis, land use, climate and water use) while the CG centres are regional nodes looking at agriculture, poverty and the environment (LAC, SSA, Asia). Donors of a proof of concept project are DFID, DEFRA, USADA, HubZero and CCAFS. IT infrastructure called HubZero will be used to communicate the results and allow analysis, although apart from the donors it is unclear who the exact users will be.

The initiative seeks to explore relationships between agricultural productivity and potential drivers and/or stressors, but does not look at specific research outputs. An objective of the initiative is to provide data that allows for policy-makers and other actors to evaluate decisions, and there is a strong economic analysis component, but the specific decisions are not mentioned.

The target population of the initiative would be the global population but in the short term the initiative will focus on two countries in South Asia and six in SSA in a proof of concept phase. There is some evidence that the scale hierarchy of agricultural systems has been considered, and regional and global nodes are envisaged although these are more for administrative purposes and for taking advantage of regional expertise. Poverty is one of the key datasets that will be compiled and could be used as a conditioning factor but is seen more as an outcome of agricultural activity.

The most important technical goal of the project is to provide a time-series of data in order to show change, but there appears to be no commitment to continue with monitoring these datasets into the future, just a mention of the need for future updates. The benefit of combining datasets to produce a global time-series of key datasets has been made in the justification of the project but there is no evidence that an assessment has been made of the cost versus the benefit, although there is mention that this issue will be addressed in the proof of concept stage.

With regard to sustainability of agricultural production, an objective of the initiative is to understand the trade-offs between agriculture and the environment but there are no details of the indicators proposed, no mention of methods nor a clear conceptual framework.

GEO GLAM

The major impetus for this initiative was from G20 agriculture ministries and the need to monitor, prepare or mitigate food price volatility. Specifically the objectives of GEO GLAM are to reinforce the international community's capacity to produce and disseminate relevant, timely and accurate forecasts of agricultural production at national, regional and global scales. The initiative has not been designed to support a specific decision; instead the purpose of the monitoring is to prepare for changes in global food prices by monitoring key observable components of food price volatility - food production and cropping patterns. The focus was initially on the major grain crops (wheat, maize, rice and soybean, total grain) for the G20 + 7 countries covered by AMIS, but monitoring of countries 'at risk' is proposed with emphasis moving away from 'key' crops to crops specific to those countries, as well as pastures and rangelands

The main operational partners of GEO GLAM are the University of Maryland (USA), Joint Research Centre Ispra (EC), USDA FAS (USA), IRSA CAS, Beijing (China), ISRO (India), GEO Secretariat (Brazil), Agriculture and Agri-Food Canada, UCL (Belgium), IIASA and USGS (USA) who form the GEO Agricultural Monitoring Community of Practice. Other actors provide oversight at the international level, as well as thematic expertise and national organisations in charge of national monitoring.

There is a mix of data sources that provide resolution from a 5km grid to a 1m grid but these correspond more to the satellite sensors available rather than the system itself. The temporal resolution of the satellite imagery goes from hourly (for the largest grid size) to monthly (for the metre level resolution imagery). The national monitoring initiatives that are supported by GEOGLAM would need to link remotely sensed data to national statistics and expert knowledge - but there is currently no description of the tools required to do this nor any evidence of the use of conditioning factors. In terms of reporting there seems to be a desire to produce monthly bulletins of crop growth but it is unclear for which spatial units these will be reported.

Apart from providing information to policy-makers there is a global network of satellite based data dissemination systems providing environmental data to a worldwide user community using GEONETCast. GEO GLAM will attempt to make these data sets freely and openly available not only to the GEOGLAM community, but also to the general public, however the documentation notes that the existing policies on data sharing of different partners would need to be taken into account. There is also evidence that feedback on the different remotely sensed products (combination of platforms, sensors and algorithms) will be incorporated, as well as indicators on food supply as a result of interaction with AMIS. Training in sampling procedures for national monitoring systems will be instigated and protocols will be developed to allow the use of data from different (satellite) platforms.

The total costs of GEOGLAM are estimated at US$7.5M per year, but there is no estimate of the monetary benefits that would accrue as a result of implementation or a social return on investment.

Committee on sustainability assessment (COSA)

COSA is a non-profit global consortium of institutions that seeks to provide an indicator set that decomposes 'sustainability' into directly observable indicators at the farm level to appraise the three pillars of sustainability (economic, environmental and social) over multiple years. The purpose of the COSA indicators is to support general investment decisions but some of the indicators could be used in a project monitoring and evaluation framework which would address specific decisions. The indicators assess sustainability at the (coffee and cocoa) farm level, while the coverage is focused on producers certified with sustainability labels such as Fairtrade, Organic, Utz Certified, Rainforest Alliance, COSA has the capacity to analyse the impacts of local sustainability initiatives and other "on farm interventions" on a case-by-case basis. COSA is facilitated by IISD and UNCTAD through their SCI, indicators were designed and evaluated by developing country producers, traders, leading companies, NGOs, standards bodies, intergovernmental agencies, and research institutions. COSA methods were developed by a broad group of stakeholders that range from producer groups and NGOs to international agencies and private firms. They are implemented by different organisations within specific countries according to the purpose (e.g. certification according to different organisations). The indicators have been chosen so that changes can be seen over short time periods, and while COSA appears to be the method which is used by other certification schemes they may impose their own scheduling of repeat measurements. Links with research are not explicit, instead the indicators seek to produce a general diagnosis, and when the assessment is repeated the monitoring initiative employs statistical analysis methods to make inferences about any interventions introduced on the farm in that time (and changes in the economic context) but in the example given the sample size did not allow statistically sound results.

The basic units of analysis are generally farms (with some practices allowing certification) and in some cases control farms and the coffee case study shows that sustainability initiatives, growing regions, farm size, agroecological zone, coffee types and production systems are considered at the sampling stage allowing disaggregation according to farm types, resource endowment, age, and gender. There is no general evidence that conditioning factors are used in stratification, although counterfactual farms were included in the coffee study (Figure 9). There is a potential for bias depending on how the counterfactuals are chosen, as a general monitoring tool (without counterfactuals) the initiative is not suitable since the farms are self-selected.

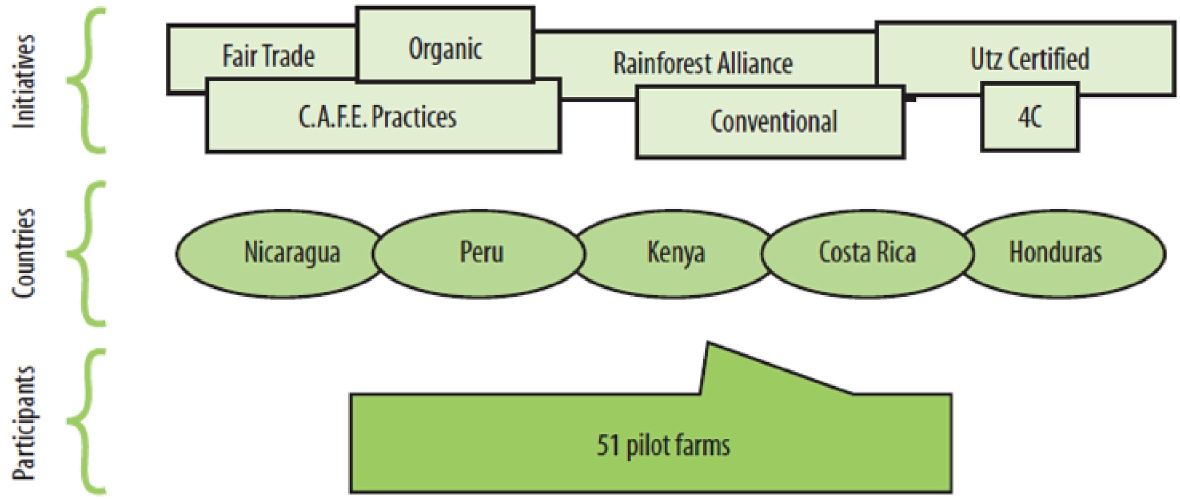

Figure 9. General sampling framework of COSA coffee study.

The cost of implementing COSA is not specified and no trade-off has been made between the cost of measurement and the potential errors and COSA focuses on tracking the existence of management systems, waste reduction measures and evidence of safe chemical use procedures rather than direct soil and water sampling. There is no evidence of the use of benchmark or critical values, especially as direct monitoring of natural resources and carbon sequestered were replaced with indirect measures. Protocols for standardised data collection have been developed in the form of a harmonised questionnaire, although the guidelines for assessing some indicators (e.g. using trees as a proxy for biodiversity) are unclear in the documentation available.

The organisations using COSA work directly or closely with the end-users such as producers, traders, certification bodies and policy-makers. Apart from the coffee report it is unclear how results will be shared. There is the intention to share data publicly (respecting the anonymity of individuals) but this has yet to be done. There is no evidence of the results of the monitoring being evaluated, although some feedback about the instruments used and the training required has been incorporated.

The COSA monitoring results can potentially be used for cost benefit analysis (either ex ante or ex post) of agricultural interventions although this would require valuation of some services or goals because sustainability issues are not considered explicitly, instead COSA makes reference to existing standards that make up the different indicators. Trade-offs between productivity, natural resource management and well-being are implicit and higher premiums are meant to compensate for additional costs (in essence putting a value on sustainability).

National Agri-Environmental Health Analysis and Reporting Program (NAHARP)

The objectives of the Canadian National Agri-Environmental Health Analysis and Reporting Program are to assess the environmental performance of agriculture over time, measuring the impact of adopting environmentally beneficial management practices (including tillage), and the effectiveness of agricultural policies and programs. The program has also been designed to provide information to enable the development of strategies and actions to safeguard areas and resources that remain at environmental risk, e.g. impact of erosion on crop yields, impact of trace elements on soil health, the impact of land use changes on salinisation, and the effect of best management practices on N contamination in water. The main actors in this initiative are, at the federal level Agriculture and Agri-Food Canada (AAFC), and provincial governments. AAFC envisage that the indicators would support farm level decisions such as how much fertiliser to apply, as well as more general policy decisions at the province or federal levels rather than evaluate research outputs. The program documentation does not mention the costs of measurements, although the regulatory framework behind some indicators might mean that measurement is a necessary cost to ensure compliance and is therefore non-negotiable and the cost has been factored-in at an earlier stage when the regulatory framework was devised.

The target (human) population is all the population living in the extent of agricultural production in Canada. The data used to calculate the individual indicators are collected at various temporal and spatial scales, but are integrated and displayed for a common unit - the Soil Landscapes of Canada (SLC). Indicators are generally reported for individual units that are consistent with the SLC dataset, while drainage areas are used for reporting some water quality indicators. The sampling frame and reporting units certainly allow for the questions to be answered at the national level, although more local decisions might require a different reporting unit. Statistical analysis and models are used to harmonise data from different sources and to combine measurements from different units into a common SLC unit for display and communication. Critical values of different indicators are known at local levels (especially for environmental quality) but are not used at the national level, instead indicator values are categorised in 5 classes ranging from good to bad, high to low etc.

Data appear to be updated every 5 years between 1981 and 2006 as a result of agricultural censuses and other surveys. Since many data are based on the census there was no sampling *per se*, however the differences in the sampling units between indicators has been communicated. Other indicators were based on data from specialised surveys whose sampling frame is communicated elsewhere. The sampling frame of the Farm Environmental Management (FEM) surveys is deliberately skewed to larger enterprises. Data from the 5-yearly agricultural census allows for some conditioning variables to be considered in the interpretation of the indicators. Composite indices are used extensively to help track performance, examples include the Soil Quality Agri-Environmental Performance Index (itself a combination of four indices), the Water Quality Agri-Environmental Performance Index (again a combination of four indices), and Air Quality Agri-Environmental Performance Index (itself a combination of three indices). A Beneficial Management Practices (BMP) Adoption Index is also being developed to fill in a knowledge gap. Proxy indicators are used for air quality, and specifically greenhouse gas emissions, which are estimated, based on land use or animal numbers rather than directly measured.

A summary report has been produced with interpretation of individual indicators and indices, in addition provincial summaries are included, and a web-mapping service shows key indicators. Summarised indicators at the SLC level are publicly available, although data tables and raw data are unavailable. Ethical issues are not mentioned in the indicator series, but are an issue for the FEM surveys and census where data are suppressed if it is possible to identify individual farms and farmers. The results are evaluated and where gaps are found there are efforts to fill those gaps. Trade-off analysis and cost-benefit analysis of policies is proposed, although a full cost-benefit analysis of interventions requires better valuation of non-marketed goods and services which is under development, at present AAFC stresses that all indicators should not be seen in isolation so that a tacit trade-off analysis is undertaken by users.

Joint Experiment for Crop Assessment and Monitoring

The overarching purpose of the Joint Experiment for Crop Assessment and Monitoring (JECAM) is to compare data and methods for crop area, condition monitoring and yield estimation, with the aim of establishing ‘best practices’ for different agricultural systems. The goal of the JECAM experiments is to facilitate the inter-comparison of monitoring and modelling methods, product accuracy assessments, data fusion, and product integration for agricultural monitoring. JECAM does not support specific decisions and instead is designed for general monitoring of agricultural production. Like GEO GLAM JECAM is coordinated by the GEO Agriculture Monitoring Community of Practice, and individual countries volunteer to participate. So far sites in 12 countries (Canada, Mexico, USA, Argentina, Paraguay, Belgium, France, Russia, Ukraine, Mali, South Africa, and China) have been identified where research is carried out; these partners make requests for satellite imagery (optical, radar and microwave) which is provided by imagery providers, facilitated by JECAM.

The spatial extent of monitoring is not clear and while the JECAM partners would like all major cropping systems to be included there is no definition of these, the sample points have been communicated but it is unclear how these fit into a more general global sampling frame of agricultural areas. Given the strong role for space agencies it is likely that the spatial and temporal framework will be defined by the different satellite sensors (see GEO-GLAM above) which are used to directly observe crop area, while crop yield is estimated using calibration at sites around the world for some key crops. Different sites are collecting different data (and at different resolutions) depending on local priorities, examples include changes in land use and cropping area, soil moisture, residue and tillage.

The development of international standards for monitoring and reporting protocols is one of the objectives of JECAM, but there is little information about the use of proxy indicators, or the methods employed to interpret the results. Data sharing is a fundamental requirement to the success of JECAM. In particular sharing In-situ data (ground measurement and observations) and free and open access to earth observation imagery. But these are within the partners of JECAM, and it is unclear if data will be shared outside of JECAM with the website used to communicate basic site information and specific in situ validation objectives to the general public.

The initiative does not consider sustainability issues and there is no analysis of the trade-offs between agricultural production, conservation and well-being.

Statistics and Ecological Monitoring Directorate - National Ecosystems Monitoring and Desertification Dynamics Programme

This initiative has been proposed by the Nature Conservation General Directorate (DGCN) of the Government of Burkina Faso, and seeks to ensure a continuous knowledge of the status and trends of the environment and ecosystems that make up Burkina Faso's natural capital. The initiative has been set-up to support general monitoring of ecosystems but with an emphasis on desertification, of which some components are directly observable. The main actor is the Statistics and Ecological Monitoring Directorate within the DGCN which takes a coordination role of scattered local monitoring initiatives, with some links at a higher level to UN-SPIDER.

The spatial extent is the country of Burkina Faso, but there is no information available about the temporal scope of inference or the frequency of repeated measurements. The spatial sampling considers the four ecoregions in Burkina Faso, but this is the only scale that has been considered and there is no evidence that other conditioning factors have been appreciated, apart from the need to consider areas that are both protected and in productive use.

There is some evidence that the initiative will use remotely sensed imagery (for wildfire detection for instance) and the documentation acknowledges the need for ground-truthing but there is no information on methods just a general framework for data management. There is also no information on what indicators would be used or how the results of monitoring would be interpreted, and no information on whether any analysis of the trade-offs would take place.

Long Term Ecological Research (LTER) Network

The LTER was set up and designed by the US National Science Foundation (NSF) with the implementation carried out by the various US universities which run each site. The Network brings together more than 2000 scientists and graduate students. LTER Network receives its greatest funding from NSF, but other Federal agencies such as the USDA Forest Service and Agricultural Research Services, the National Aeronautics and Space Administration, the US Geological Survey, the US Environmental Protection Agency, and the US Department of the Interior's National Park Service and Fish and Wildlife Service also support various projects at site and network levels. The costs of implementing the research network is well known, but it is less clear if any assessment of the value of the information has been undertaken, the 30 year review states "we feel that the existing network is of incredible value and represents one of the highest scientific values for the science dollar" but there is no mention of the methods used to determine the value-for-money of the LTER

The original objectives of the LTER were to evaluate the nature and pace of ecological change, to interpret its effects, and to forecast the range of future biological responses to change, with the focus on individual research sites. Over the past four decades the focus has shifted and the LTER now seeks to understand how humans perceive the critical services provided by ecosystems at multiple human scales, how these perceptions change behaviour and institutions, and how these changes in turn feedback to affect ecosystem structure and function and the ability of ecosystems to continue to deliver services over the long term (Figure 10). As such the research network as a whole does not support specific decisions but instead focuses more on providing fundamental knowledge on long term processes which take decades and up to centuries to manifest.

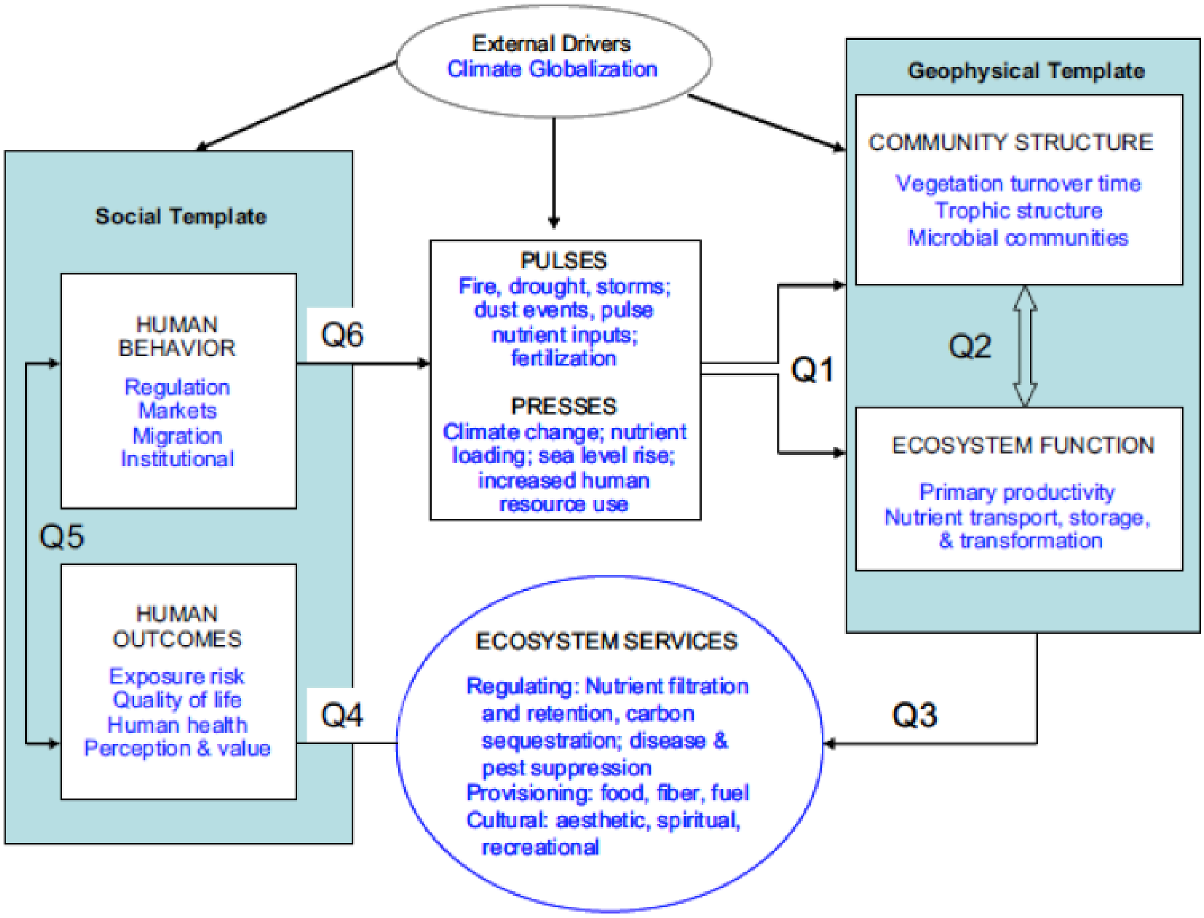

Figure 10. Integrated Science for Society and the Environment framework used as a basis for LTER research questions.

The extent of the monitoring is limited to the USA (and Antarctica), however "Geographical distribution of sites or ecosystem type never became active criteria in the selection process. Yet, the resulting geographic/ecologic diversity of the research sites is so broad as not to contain duplication" (Callahan, 1984). The sample units are very well defined and characterised sites and the sampling frame is appropriate for the ecological research, but does not represent the 95 sites that had been suggested by The Institute for Ecology (TIE) and while there does not appear to be any bias, agricultural lands are not well represented. It is unclear if scale hierarchy has been considered during the sampling, but it is considered in certain projects at specific sites (e.g. McMahon and Diez, 2007). Since the focus is on ecological research rather than socioeconomic research socioeconomic conditioning variables are not considered. The temporal scope of reference was clear when the monitoring network was setup, i.e. the ability to track environmental processes over decades and centuries.

The core research areas are: 1.Primary Production; 2.Population Studies; 3.Movement of Organic Matter; 4.Movement of Inorganic Matter, and; 5.Disturbance Patterns. All of which have some components which are directly observable. There is no evidence of the use of proxy indicators for fast variables since research is fundamental and fast variables such as greenhouse gas fluxes are measured directly (in the case of the KBS station which specialises in agriculture). Measurement protocols are being developed for the synthesis themes and are used for many experiments at the KBS station, but there is no evidence of any analysis of the trade-offs between measurement cost and potential errors. Statistical methods are used in the analysis of measurement notably in the cross-site analysis of NPP, but there is no evidence that secondary data are used for validation or for extrapolation. Critical values are not relevant for NPP measurements but an economic threshold exists for pest infection used as a variable in the natural predator research. There is also recognition of tipping points in the synthesis theme on landscape vulnerability and resilience to climate and land use change.

The LTER Network makes data collected by all LTER sites broadly accessible to other investigators in LTER sites but unclear if data are shared out with LTER. Local communication at all levels (from scientists to the general public) with face-to-face communication and hands-on. At higher levels there are brochures and profiles as well as technical reports and peer-reviewed papers

There are examples of research outputs being monitored and outcomes (on ecosystems) being subsequently observed, however the research has tended to be site specific. The agricultural interventions are currently not implemented at all sites but monitoring and evaluation are carried out as part of the many experiments that are conducted. Sustainability is considered in the Strategic Research Initiative of Integrative Science for Society and Environment, but it is not clear how the monitoring framework has been or will be modified to take sustainability into account. Trade-offs have been considered in some research but at very local levels (e.g. plant physiology), some discussion about information trade-offs (between certainty and timeliness, and also between scale and resolution), but little mention of trade-offs between development objectives.

National Ecological Observatory Network (NEON)

NEON has been funded by NSF and the main actor is NEON Inc. - a private company set up to run the observatory. NEON was designed by NSF and Oak Ridge National Laboratory (ORNL) and NEON Inc. will have its own staff, while individual sites will be run by third-parties. There has been a recognition of the cost of some measurements, notably the airborne observation platform, with a compromise between the cost and the area covered and trade-offs have been made between the costs of measurement and the potential errors in the sense that uncertainty in measurements was built into the spatial and temporal sampling which was included in the observatory design and costing.

NEON seeks to answer five key questions: (1) What are the impacts of climate change on continental-scale ecology?; (2) What are the impacts of land use change on continental-scale ecology?; (3) What are the impacts of invasive species on continental-scale ecology?; (4) What are the interactive effects of climate, land use, and invasives on continental-scale ecology?, and; (5) How do transport and mobility of energy, matter, and organisms affect continental-scale ecology? Although decision-makers are referred to, it seems that specific decisions are not being supported by NEON; there are relocatable sites that deal with agricultural systems, but not much information about how agriculture will be tackled with no mention of specific interventions.

The geographic scope is the continent of North America, while the temporal scope is 'decades' with plans for 30 years of observation. NEON will observe physical and chemical climate, land use, and biological invaders and biological responses (matter and energy fluxes, biomass and plant productivity, diversity and genomics of key organismal groups, infectious diseases and community, phenological and population indicators). The spatial and temporal measuring frames and the instruments have been designed explicitly to incorporate uncertainty and to be able to define trends over a 30 year period "NEON observations require sufficient measurement accuracy, precision, sampling, and replication to allow the detection of decadal trends against the background of diurnal, seasonal, and interannual variation. Many of the measurements planned for NEON have never been deployed to detect and diagnose long-term trends." A protocol was developed for the sampling (based on uncertainty in measurements) and protocols will be available for all other measurements. It is unclear if benchmark values will be used but there is some evidence of temperature thresholds for pest mortality.

The sample units are biophysical domains, and the scale hierarchy has been considered in both the sampling design and choice of sites which are representative of larger domains, as well as the integration of site based observations with remotely sensed data over larger areas to enable extrapolation. There is no evidence of bias, and sampling frame based on the variability of key variables such as growing season, precipitation, soil water holding capacity, and solar radiation. However, since the focus is on ecological rather than socioeconomic research conditioning factors such as gender or resource endowment are not considered.

Statistical methods have been utilised both in the design and in the extrapolation. Data from outside the NEON observatory is mainly on land use, using remotely sensed data as well as existing processed geospatial datasets but no information is available on the tools used to harmonise/integrate the data. With regard to interpretation of the data there is no mention of any trade-off analysis for agricultural systems evaluation there are relocatable sites that deal with agricultural systems, but not much information about how agriculture will be tackled with no mention of specific interventions.

Education and outreach are a significant component of NEON with an education director and the data provided by NEON will be freely available to a wide variety of users including scientists, students, educators, and the general public in near real-time and NEON, Inc. will endeavour "to archive and distribute data generated by individual investigators at NEON sites, provided that data and meta-data are produced in accordance with NEON formats. Data collected based on funding from public agencies (including data using NEON facilities) will follow agency/NEON policies for public release."

Biodiversity Information System for Europe

Streamlining European Biodiversity Indicators (SEBI) is a component of the Biodiversity Information System for Europe (BISE) which is a single entry point for data and information on biodiversity in the EU and serves as the EU Biodiversity Clearing House Mechanism to CBD. SEBI began in 2005 with the establishment of a Coordination Team and the involvement of six thematic expert groups. This involved around 140 experts from across the pan-European region and from international intergovernmental organisations and NGOs. Each group provided a range of technical expertise and geographical coverage. SEBI institutional partners are the European Environment Agency (and its European Topic Centre on Biological Diversity), the European Centre for Nature Conservation, UNEP- WCMC, the European Commission, the Joint Secretariat of the Pan-European Biological and Landscape Diversity Strategy (PEBLDS), and the Czech Republic (as lead country for the Kiev Resolution action plan on biodiversity indicators).

The objectives of SEBI are not clear although it does seek to address the following general policy questions referring to biodiversity: (1) What is changing? (2) Why is it changing? (3) Why is it important? (4) What are we doing about it? There is no evidence that SEBI or the general information system address a specific decision but the indicators should "measure progress towards the new European and global 2020 biodiversity targets" - which implies some decisions taken at some point. There is some mention of the cost of implementing a well-designed indicator system but that the cost would be less than the societal cost of inaction or the development of poor policies, but with no figures to back up this assertion. In addition where data gaps were identified there was an implicit trade-off between filling the gaps and the resources available.

Some of the questions that SEBI addresses have components which are directly observable - such as abundance and distribution of species, critical loads of contaminants, or the status of habitats and the system (i.e. the collection of indicators) has been designed to detect changes within the timeframe. The initiative was designed to compile indicators of both causes and effects so that relationships should be achievable and quantifiable in order to link pressures, state and response indicators. Critical levels are noted for many indicators, such as fish stocks, and nitrogen deposition. However, it is recognised that more research is needed to define the critical levels of other populations or pollutants. Composite indicators are used, for instance the red list species trend index and the Marine Trophic Index, but SEBI recommends that baseline values are used to enable a more general composite index of ecosystem quality. Despite the use of the pressure-state-impact –response framework there is little regard for conditioning factors and biodiversity indicators appear to be interpreted in isolation of other types of indicators. The initiative could potentially be used for cost benefit of policies but only if integrated with some valuation of the biodiversity benefits, and while there is a mention of sustainability there are no details about any trade-offs between different objectives.

The spatial scope is broadly Europe with some measurements allowing monitoring at the national level. The temporal frame is towards targets for 2010 and now 2020, based on the CBD and the 2003 "Kyiv Resolution on Biodiversity" and on more recent conventions such as the Aichi targets and the EU's long-term (2050) vision. Sample units appear to be nation states but it is difficult to assess and indicators come from a variety of sources and as such there has not been any sampling design at the European level although in a review of the indicators it is noted that indicators can be nested according to spatial scale.

All of the data are from external sources but it seems that the lack of harmonisation has restricted the indicators that could be used "The status indicators on species only cover birds and butterflies, since these are the only taxa/species groups for which harmonized European monitoring data are available." There was not enough information to assess the data sharing principles of the initiative but progress reports and indicator fact sheets have been produced and are available via the EEA website. The results of the whole streamlining initiative was evaluated but recommendations for changes to the indicator set seemed to be a response to the external policy framework rather than the scientific relevance of the indicators themselves.

World Database on Protected Areas

The objectives of the World Database on Protected Areas (WDPA) are not clearly defined, but the initiative ought to "provide a basis for assessing the extent of formal protection of global biodiversity, and a measure of conservation commitment at the global scale". There is no evidence that the initiative supports specific decisions, although the database helps track the global response to MDG7 target 9, and CBD targets. The major stakeholders of the WDPA are UNEP-WCMC, IUCN and World Commission on Protected Areas. The original coordinator of the initiative was the IUCN, and afterwards by UNEP-WCMC; other contributors are IUCN-World Commission on Protected Areas, American Museum of Natural History, BirdLife International, Conservation International, Fauna & Flora International, The Nature Conservancy, Wildlife Conservation Society, World Resources Institute and World Wildlife Fund, as well as the national bodies who contribute information about the protected areas.

The WDPA has been publicly accessible since 2003 and now has a Website dedicated to the publication of the database with download options and user contributions.

The spatial scope is worldwide, while the temporal scope is constant, starting from 1962. The monitoring initiative currently monitors the existence, location and legal characteristics of protected areas with all the data provide by third parties, requiring metadata standards to harmonise datasets with rules on minimum, core and enhanced attributes for each protected area. It is possible to detect changes in the number, or area of protected areas over time, as well as some indication of the level of protection, but there are no attributes about the condition of the protected areas Simple statistics are used to show changes in protected areas over time and contribute to other initiatives such as the Biodiversity Indicators Partnership (BIP). The protected status is a proxy for some variables as well as a response indicator in its own right and while there is the potential for links to be made between system monitoring and the evaluation of interventions (see Chape et al, 2005) this does not seem to have taken place.

World Agriculture Watch

The design of the World Agriculture Watch (WAW) initiative is the consequence of a debate between ministries in the Government of France and CIRAD. The other actors responsible for global coordination and promotion (FAO and IFAD) appear to have become involved to provide a more global mandate, while the implementing partners at the local level are the local partners and CIRAD. Local partners include *Réseau des Observatoires Ruraux* in Madagascar which has undertaken farm level surveys and analysis, *l'Institut d'Economie Rurale* in Mali who have a strong GIS and remote sensing unit, AGRHYMET and the *Direction des statistiques du Ministère de l'élevage* in Niger, IICA, the Center of Economic Policy for Sustainable Development at the National University of Costa Rica, and the Development Observatory (ODD) at the University of Costa Rica, and the Asian Institute of Technology in Thailand.

The objectives of WAW are (1) to document more accurately the diversity of agricultural production systems, their structural transformations, their resilience to current challenges and contributions to sustainable development; (2) produce comparative spatiotemporal analyses; (3) warn of possible crises and specific vulnerabilities, proposing possible policy options, and; (4) strengthen the capacities of local, national and regional observatories and stakeholders to collect and analyse relevant information, and use it to fuel the policy debate. The global monitoring initiative will link different national monitoring initiatives, although as yet unclear how this will be achieved and there is no evidence of any assessment of the cost of the measurements versus the beneficial value of the information obtained.

The spatial scope is a range of "low to high income countries and differing farming systems, where significant structural transformations are in progress or foreseen". The temporal scope is not addressed. Current pilot schemes are underway in specific districts or sectors in Niger, Mali, Costa Rica, Madagascar and Thailand but global sites have not yet been defined. The scale hierarchy of the natural and human system has been considered and monitoring takes place at three levels: (1) agricultural production units, (2) 'territories', and (3) markets (as chains). The sampling of these units has not been communicated, although there is some mention of sampling (intensively cropped areas vs. frontier areas) in the Mali case study. There is not enough information available to assess fully whether socioeconomic conditioning factors have been considered, but the descriptions of the farm level surveys carried out in the case studies suggest that conditioning variable are very important because they help define the farm typologies which is a common conditioning variable in all cases. Some of the components can be directly observed, such as production systems diversity, but other components require analysis and modelling. Indicators have been, or are being, defined in specific case studies (agri-environmental indicators in Madagascar) but there does not yet seem to be any harmonisation between case studies and it is difficult to assess whether changes will be detectable.

The global initiative has not been designed to support specific decisions, however at the case study level there are links between system monitoring and the evaluation of interventions, examples include analysis of changes in farm systems from local markets to agro-export systems, on the impacts of production systems and choices on agri-environmental indicators. In some of the case studies (notably in Thailand) there is an explicit treatment of trade-offs between environmental impacts, water resource depletion, and cropping (mainly rice) system performances.

Currently there appears to be in-country communication of results with documents available online. Eventually there will be a "web portal hosted at FAO, connected to local WAW websites, providing access to: i) published documents; ii) relevant global and local data and reports; iii) on-going trends and issues; iv) forums." Databases are one of the deliverables of the initiative but it is unclear how freely accessible the data will be.

Poverty & Livelihoods

Women's Empowerment in Agriculture Index

The objective of the Women's Empowerment in Agriculture Index (WEAI) is to "track the change in women's empowerment levels that occurs as a direct or indirect result of interventions under Feed the Future, the US government's global hunger and food security initiative." Oxford Poverty and Human Development Initiative (OPHI) developed the WEAI and adapted the Alkire Foster method, which underpins the Index. Feed the Future, through support from USAID, defined the five domains of the Index (production, resources, income, leadership and time), provided technical input on the development of the pilot survey, and provided overall policy guidance for the Index. IFPRI provides overall coordination for the project, designed and implemented the household pilot survey, and developed the individual case studies, working with in-country collaborators.

As a whole the Index does not support specific decisions but the different domains of the index can be addressed. The scope of the assessment and objectives are addressed to policy output or on-farm decision-making, however since Feed the Future has some research components it might also be possible to link development outcomes with research outputs. There is no information about the cost of measurement but in this case the measurement and analysis are necessary to see if the interventions are supporting Feed the Future's target group. Statistical methods are used to analyse the results and the Sample size has been defined beforehand presumably so that results are statistically significant. The sample size is presumably the minimum size which is a trade-off between cost and statistical significance.

The spatial scope is currently limited to case studies in three countries where Feed the Future is active (Guatemala, Bangladesh, and Uganda), with the household as the basic unit of analysis. Most of the component indicators of the index are not directly observable but require a questionnaire to elicit responses from individuals. The sample population is both women and male members of rural households with socioeconomic conditioning variables were considered at the sampling stage with stratification according to wealth class and gender. Only one scale is considered and the choice of districts in the case studies was not clear in the case of Bangladesh or Uganda, and in Guatemala the districts were chosen because of the large proportion of indigenous population. The sampling frame is not biased but there is no mention of the choice of the pilot areas apart from the fact that these areas should represent the domain of the local Feed the Future mission. The temporal scope for performance monitoring requires that data for the Index should be collected on a biennial basis.

The Index has a well-developed protocol for data collection and the construction of the index, whereby each indicator has a threshold value, above which the achievements are considered adequate. The index is analysed for correlation with wealth, education, age and hunger, but there are no links to external variables such as cropping system. GPS coordinates of the sample households are not taken (but villages are noted) and names are removed from the digital database to protect the anonymity of respondents. At the moment each country is independent and the results are analysed independently, there is no information on the accessibility of data and it is unclear how the results will be communicated outside the three coordinating institutions.

After analysis it might be possible to conduct an ex-ante cost benefit analysis by extrapolating based on the performance of Feed the Future interventions on the WEIA index. The index does not consider wider sustainability issues but there might be the opportunity to link to other initiatives supported by Feed the Future such as AfricaRISING.

Global Information and Early Warning System

The Global Information and Early Warning System was designed by FAO "to provide policymakers and policy-analysts with the most up-to-date and accurate information available on all aspects of food supply and demand". More specifically "to keep the world food supply/demand situation under continuous review, issue reports on the world food situation and provide early warnings of impending food crises in individual countries." The main actors are FAO in a coordination, analysis and communication role, but individual countries often have their own early warning systems (often in statistical agencies) and contribute to regional systems. NGOs and other UN organisations are users of the system. There is no evidence that the initiative supports specific decisions but the initiative does support national Early Warning and Food Information Systems. There is an implicit link between initiative monitoring and the evaluation of interventions because country policies are also monitored, allowing links between food security outcomes and policies to be made.

The spatial and temporal scope of the initiative is not clearly communicated, but it appears that the basic unit of analysis is the nation state, and the initiative has access to satellite imagery to monitor NDVI (with a spatial resolution of 1km and a temporal timeframe of 10 days) and rainfall estimates at different locations for every 10 days. It is possible to assess the trade-offs between the frequency and cost of measurement by comparing the data that are required (from satellite remote sensing) to forecast food availability with those that are used by the initiative. However, it is not clear whether this has been done.

There are some components of food supply which are observable (e.g. crop growth and yield) and NDVI is used to estimate crop growth during the season and when combined with rainfall estimates is used to forecast yields for the season. NDVI is however a proxy for crop growth, which is a proxy for crop yield, which itself is only one component of food security. NDVI values can be assessed against previous critical values to assess the current status of plant vigour. There is also some evidence that the initiative receives official information about the status of crop growth from individual countries, although how this is incorporated into the initiative is not clear.

Socio-economic factors are not considered and disaggregation according to farm types or resources endowment is only possible if combined with third-party socio-economic data.

Country briefs are produced at least four times a year and are freely available via the website, but it is unclear if the data are shared outside the partners, with only some meteorological data freely available from the website.

FEWSNET - Famine Early Warning System Network

The objectives of the Famine Early Warning System Network (FEWSNET) are to provide timely and accurate information to pinpoint and assess emerging or evolving food security problems, assist in the capacity development of information providers, understand the underlying causes of food insecurity, and identify long term development needs, strengthen national and regional capacity for early warning and response planning, increase the usefulness of information for decision makers, improve the appropriateness of responses to food security related issues through a better understanding of food insecurity, improve the timeliness of responses to food insecurity, including fostering early policy action, and improve local monitoring and analysis. FEWSNET has a presence in 20 countries (17 in SSA, 2 in LAC and 1 in SA), but will implement remote monitoring in a further 14 in Central America, Africa, and Asia. FEWS NET is implemented by Chemonics International, Inc., with support from the United States Geological Survey (USGS), the National Aeronautics and Space Administration (NASA), the National Oceanographic and Atmospheric Administration (NOAA), and the United States Department of Agriculture (USDA); and with livelihoods research in collaboration with FEG Consulting.

It seems that FEWSNET is an adaptive framework, based on experience in the field and reflection. An important example is the adoption of the IPC phase classification and the dropping of FEWSNET's bespoke food security indicator. In addition the realignment towards remote monitoring of countries where FEWSNET does not have a presence was motivated by the need to increase the spatial scope of FEWSNET without opening offices. These adaptations imply that some analysis of the cost of measurement versus the value of the information has been undertaken. However, it is less obvious whether an assessment has been made of the trade-offs between cost of measurement and the potential errors or uncertainties in these new countries where benefit is expected to accrue from remote monitoring.

There are some components of food security which are observable (e.g. crop growth and yield) and NDVI is used to estimate crop growth during the season and when combined with rainfall estimates is used to forecast yields for the season at a fine resolution. One of the biggest contrasts between FEWSNET and GIEWS is that the basic unit of analysis is the livelihood zone which reflects larger areas of general agricultural production and consumption systems. The livelihood zones allow general livelihood conditioning variables to enter the food security analysis. These are considered as contributory to the direct measures of food security that are incorporated in the IPC classification and different livelihood profiles are developed for each zone (based on wealth) which describe different livelihood strategies and activities.

The initiative links data from remote platforms, as well as local data on household or community crop production indicators (such as sowing dates). There appear to be guidelines to aid the diagnosis and prognosis but it is unclear how the data are interpreted to produce the prognosis, although some indication that HEA approaches are used. Changes in WRSI and NDVI are crucial components of the initiative, but are less to the forefront of the initiative (compared say to GIEWS), instead the food insecurity phase is a more important indicator. Food insecurity has been classified using a more or less standard measure (the IPC version2 food insecurity phase) which has descriptive benchmarks for each phase. The IPC classification is a combination of nine factors that when combined define food security. These are replicable and allow comparison between countries (and even with other initiatives like GIEWS or MARS). The most critical need for protocols is in the livelihoods component of the early warning system. The IPC classification is a protocol for how to analyse and summarise indicators but there is no evidence for measurement protocols. There is no evidence of the use of statistical methods in the analysis of results, indeed it is suggested that "Achieving statistically comparable measures of food insecurity is not currently possible, nor is it necessarily required for early warning purposes." As with GIEWS the purpose of FEWSNET is not to explore research outputs, but there is a research component directed towards understanding the underlying causes of food security. The data on crop growth, rainfall and water requirements are interpreted in combination with livelihood profiles and baselines to estimate the impact of current conditions, but the analysis of these factors is not evident, and there is no consideration of longer term causes of food insecurity linked to natural resource degradation for instance.

The temporal frame is continuous although the normal reporting period is monthly. Early warning products are available online but unclear if raw data are available. The main communications vehicles are reports which are publicly available and which are directed towards USAID and other international aid agencies, as well as local decision makers. At the local level it is possible that FEWSNET could be used to support specific decisions to mitigate food security and there is the potential for wide-scale agricultural interventions to be analysed in the context of their effect on food security (e.g. irrigation schemes over very large areas or the introduction of a new drought-tolerant variety), but no evidence of this.

Living Standards Measurement Study-Integrated Surveys on Agriculture

The Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA) has not been designed to support any specific decision, instead the objective of the initiative is " to foster innovation and efficiency in statistical research on the links between agriculture and poverty reduction." LSMS-ISA is an extension to conventional LSMS household surveys but focuses on the African countries of Mali, Niger, Nigeria, Ethiopia, Uganda, Kenya, Tanzania, and Malawi. The initiative has been designed and coordinated by the Research Group of the World Bank, the major donors are BMGF (with support from USAID, DFID, and national governments), and the national surveys are implemented by the respective bureaus of statistics. There is no publication that shows the value of the information (apart from stating that it is very important to improve the parlous state of agricultural statistics in developing countries, particularly in Africa), but the costs of data collection are well known and are high enough to warrant funding from donors to carry out household surveys. Evaluation of interventions aimed at alleviating the food price crisis of 2007-2008 were a spur to the development of LSMS-ISA and Feed the Future (USAID) in Mali have discussed the possibility of working with LSMS-ISA to oversample areas where FTF is active so as to provide a baseline to aid future FTF evaluation.

The basic unit of analysis is the household and the spatial scope of inference for LSMS surveys are normally restricted to large regions within a country, and between urban and rural households, although there are techniques such as small area estimation to link the LSMS data with census data to increase the resolution of the scope of inference. The temporal resolution of the surveys is unclear, but the conventional LSMS surveys tend to be more frequent than the population census with 2-4 rounds expected in the pilot countries between 2005-2012. Both the agricultural and poverty variables are observable via informant recall or GPS measurement but the initiative is possibly better at showing changes over time rather than spatially. LSMS-ISA takes the scale of the human system into account, in the sense that LSMS surveys also have a community questionnaire component, but no higher levels are considered.

Target and sample populations are well defined as part of the conventional LSMS methodology, but the number of panel households in LSMS-ISA is smaller than the conventional cross-sectional sample. Socioeconomic variables are measured as part of the LSMS but there is no evidence that economic status is incorporated in the sampling via stratification apart from rural and urban. In general the sampling is well communicated for each country, and has been designed to remove bias. The surveys are practical and in the case of Malawi a remote island was excluded due to the small population and high costs of enumeration, and in general the frequency of data collection is trade-off between the demand for the data (and results) and the availability of funding.

In the past proxy measures were sought for crop yield and farm/plot area which was easier than measuring areas or timing survey visits to coincide with crop harvests. There is now a move towards diaries and use of GPS measurement. Questionnaires and sampling methodology are standardised across LSMS although each country has differences according to the context. There is also evidence of tracking protocols used in panel surveys

LSMS-ISA has the potential to allow the relationships between research outputs and outcomes to be explored via a means-ends framework because the LSMS-ISA is a panel survey, although information would be needed on the interventions themselves and it might be difficult to attribute changes to those interventions. The most common form of interpretation of the results is to relate the agricultural variables with the welfare variables and to look for correlations between the two (e.g. land size or assets and consumption or income). In addition the household income variables can be compared to national poverty lines to assess their poverty status. Trade-offs between productivity and well-being objectives are possible, but as yet the proposed modules on NRM or wider environmental conditions have not been implemented.

The Public Access Policy of LSMS-ISA is in general Open Access and free of charge, with data available within 12 months of data collection and can be found on National Statistics Office websites and the World Bank LSMS website and should also include questionnaires, manuals, sample, problems and the format of the data. All datasets are anonymised prior to release, and GPS points are not released, instead "all surveys are geo-referenced at the household level and to overcome the obvious problem of distributing geo-references, we have developed a protocol to create a set of 'geovariables' at the household level which can also be safely distributed without compromising confidentiality." The initiative allows for data from other sources to be integrated via GPS and GIS to create these 'geovariables', but the LSMS-ISA is essentially a stand-alone initiative.

Household Economy Approach

The Household Economy Approach (HEA) approach was developed in the early 1990s by Save the Children UK and the Global Information and Early Warning System (GIEWS) of the Food and Agriculture Organization (FAO), Since then other partners have come on board notably the Food Economy Group (FEG Consulting) and livelihoods have taken centre place in the approach. Oxfam has also been a partner in the development of the HEA framework. The original aim of HEA was to improve FAO's ability to predict short-term changes in a population's access to food. The current objectives are not so clear but the initiative seeks "to provide a clear and accurate representation of the inside workings of household economies at different levels of a wealth continuum and in different parts of the world." The objectives are not problem driven instead they seek greater understanding of livelihood strategies and activities, which are directly observable but also require secondary information and information from households themselves.

HEA is not intended to support specific decisions instead it is a general approach which can be adapted to different countries, it can be used for the evaluation of interventions and can be used for both ex-ante analysis (this is a fundamental part of HEA) if the agricultural interventions are considered as part of the coping strategies or a part of the livelihood strategies or activities, as well as ex-post in the evaluation of agricultural interventions. HEA has been stated as being cost-effective but it is difficult to find information on the metrics or studies used to come to this conclusion.

HEA has been applied in Sub-Saharan Africa, Central America, the Balkans and Asia. The spatial scope of the monitoring is limited in each case to a particular country with the inference domain a livelihood zone within a country, while the temporal scope has normally been restricted to a fixed point in time (a baseline) against which projections can be made for future situations. However, a study in Sri Lanka applied a HEA before and after a specific intervention to enable monitoring and evaluation. The livelihood descriptions, including the livelihood activities, potential hazards and coping strategies help determine which indicators should be monitored but this monitoring is not part of the HEA approach. All of the analysis is undertaken at one scale - the livelihood zone but within each livelihood zone different wealth classes are identified. Livelihood zones are homogenous zones in terms of livelihood options, with differences within the zones based on households' ability to take advantage of different livelihood options based on wealth, assets etc. Population within zones are not consistent. The sampling is not biased but because livelihood zones differ in area and population the interpretation must take into account these attributes in order to best plan interventions. The sampling relies principally on expert knowledge and secondary data to define livelihood zones, which is more practical than surveys designed especially for defining livelihood zones.

Critical values are used to judge variables with 'survival' and 'livelihood protection' thresholds (based on food and income) used in the analysis of outcomes as part of the ex-ante exploration of potential risks, as well as responses during and immediately after a hazardous event (Figure 11). During the baseline description critical thresholds (of assets for example) are used to define boundaries between wealth classes. There does not appear to be any use of strict measurement protocols, instead comprehensive implementation guides are utilised. Secondary data are used in the definition of livelihood zones as a preparation for stakeholder workshops, and in the baseline description and in the outcome analysis. The guidance for the analysis of secondary data is not explicit and it is left up to the implementer to best use the data available.

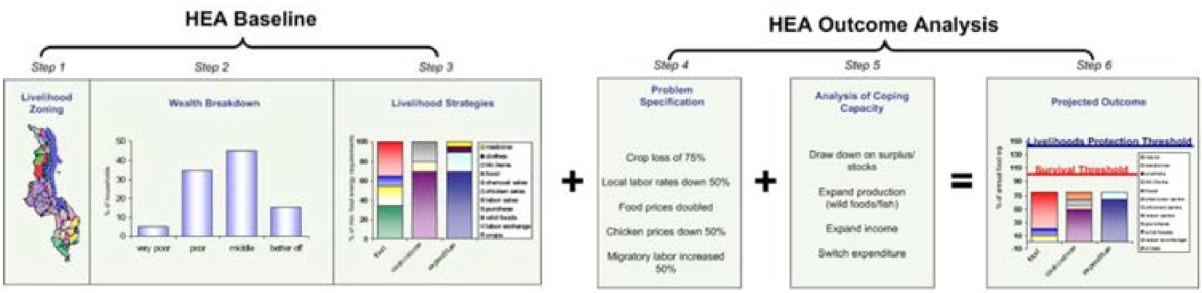

Figure 11. Framework for Household Economy Approach.

Reports are shared publicly via the FEG website, but it is likely that this is only a subset of all the HEA studies that exist and which are used by international NGOs and national agencies. In some cases it is possible to obtain the livelihood zones as GIS files, but there does not appear to be a general policy on data sharing. The baseline is best interpreted as an input on which other variables (which are assumed to be faster moving) have an impact resulting in household level development and welfare outcomes (there there are links to FEWSNET which uses the HEA baselines and more actively monitors the other drivers). Sustainability issues are considered, mainly in connection to sustainable livelihoods although the sustainability of the natural resource endowment is not addressed explicitly and no trade-off analysis between development objectives is obvious.

It is evident that the HEA has been modified over time, but it is not clear if there is a constant re-evaluation of the methodology.

Basic needs mapping

The objectives of the methodology are to develop poverty indicators (rather than a monitoring initiative per se), which provide a direct method of identifying the poor, taking into account aspects that are not necessarily reflected in the income of households, and taking advantage of the immense potential of the geographic disaggregation that is permitted by census information. The main actors are CEPAL (ECLAC) who have designed the index, and coordinate the activities of the national bureaux of statistics in LAC. The initiative has not been designed to support specific decisions, nor does the initiative allow the relationships between research outputs and outcomes to be explored via a means-ends framework, given the frequency of the surveys and the difficulties in attributing change to a particular intervention. However it is possible to evaluate changes in levels of NBI over time and make some exploratory analyses if the distribution of interventions is also well known. It is therefore potentially possible to compare the costs of data collection and the benefits that might accrue to a large-scale intervention (like subsidies, safety nets or infrastructure) through better targeting

The basic unit of measurement is the household but the spatial scope of inference depends on the whether the data are from a census, in which case the scope of inference is the enumeration district, or from a survey for which the inference domain would likely be a sector (urban/rural) or a sub-national region. The temporal scope of inference is the date of the population census or household survey from which the basic needs index is composed, in Latin America this is normally a 10 yearly period. Two levels in the human system are recognised - the principal level is the household for which basic needs satisfaction index is calculated, and at which scale some analysis is possible, but more common is the analysis at a higher aggregated level which could be the spatial inference domain, or a higher level still like the national level where comparisons between countries are possible. Sampling frames for individual countries are available and the location (rural vs. urban) is often used in the stratification. Since the research questions are not defined beforehand the sampling frame reflects more general political or decision-making territories. Since the index is based on census information there is the possibility to differentiate the outcomes based on gender and age, although farm typologies might not be possible to define.

The basic needs index is composed of a number of different components reflecting different basic needs. Each basic need (e.g. access to education) is composed of various dimensions which are then matched to variables in the census - so the conceptual basic need can only be observed after these links are made. Some of the variables in the census are approximations of the dimensions of the components of the index or there are dimensions which are not captured at all and each component of the index, and each dimension needs a critical level to be defined so that satisfaction can be assessed. These critical values are difficult to apply and may change according to the context of any particular household - for instance what is a satisfactory dwelling in tropical lowlands may not be satisfactory in cold highlands. There are no protocols defined by ECLAC, but good practice guidelines exist.

The choice of indicators is evaluated in each country but there is no evidence that it is evaluated after every census. The methodology is replicable, and there is an attempt to harmonise the indicators across the LAC region, but inevitably the values of the variables that make up the basic needs will change according to context. The results, however, should be comparable across contexts and countries. Trade-offs between the cost of measurement (frequency, access, equipment etc.) and the errors (of omission and commission) have been acknowledged (for instance by ECLAC) and this limits the explicative power of the index, suggestions to improve their power include combining with indirect (income or consumption) poverty measures to give an idea of the chronic or acute nature of poverty.

The monitoring initiative employs statistical analysis methods in the analysis of the results but statistical methods are generally not used to link to secondary data (*cf*. the use of geovariables in LSMS-ISA). Since the data are generally from one source which has collects only basic household data there is no explicit treatment of trade-offs among productivity, environmental and well-being objectives and sustainability is only considered in the household's longer term economic capacity, using indicators such as the educational levels of children and enrolment in formal education. The monitoring results are not really appropriate for cost benefit analysis of agricultural interventions, and while the index could be used for targeting interventions it is unlikely that agricultural interventions would have an impact on one or more of the components of NBI in the short-term.

The data sharing policy is dependent on each country, but generally there is a lag, and data are not always freely available. Data are anonymised (not always successfully) anated to the enumeration district level or higher administrative levels. The most common communication mechanisms of the results are reports of the analysis of the 10-yearly census although this is dependent on each country, or by third-parties.

Village Dynamics in South Asia

Village Dynamic in South Asia (VDSA) is the new name for an older initiative called Village Level Studies (VLS) which was active in India and some African countries in the 1970's and 1980's. ICRISAT is the coordinating institution while NCAP collaborates with partners in the Indian national program in Eastern India and collaborates with IRRI and ICRISAT in the rice belt and SAT zones, respectively. IRRI has oversight and responsibility for the VDSA work in Bangladesh and the donor of VDSA is the BMGF.

The purpose of VDSA is (1) to track changes in the livelihood options of the rural poor; (2) to understand farmers' response to changing markets, policies and technologies; (3) to understand farmers' perceptions on climate change and their coping mechanisms; (4) to provide feedback for designing policy interventions, setting research priorities and refining technologies, and; (5) to understand the dynamics of agricultural transformation. One of the development objectives of the study is "by providing insights and development pathways to identify and understand socio-economic, agro-biological, policy and institutional constraints to agricultural development"; this implies that different interventions will be assessed but the initiative has not been setup to evaluate any particular intervention or support specific decisions. There is so far no evidence that the monitoring results can be used for cost benefit analysis of agricultural interventions, but a review of the publications that resulted from the first generation of VLS suggest that the VDSA could be used for ex-post analysis.

The extent of the monitoring is the semi-arid tropics of India, as well as humid tropics in eastern India and Bangladesh. The sample population was farm households within the villages in districts chosen for their cropping systems (which corresponded to the mandates of the major implementing partners). The spatial inference domain is unclear but might be limited to the 42 selected villages themselves. There is an effort to compile district-level biophysical and socioeconomic data from other sources but it is unclear which tools are used to harmonise or manage the secondary data. The villages surveyed were part of VLS dating back to 1975 which lasted for a decade, the same villages were selected for monitoring from 2009 and monitoring frequency varies from monthly in SAT areas in India and Bangladesh to yearly or half-yearly in E India. Within each village farm households were stratified according to farm size (from landless to large farms) with critical values used for land holding size within each village. The sampling frame may not be representative for all semi-arid or sub-humid areas in India, but the households are representative of the villages (and farm typologies). There may be some over-representation of small farms compared to large farms due to the smaller variance of key variables and a smaller required sample size for this farm typology class, there might be bias towards areas which grow crops that are the mandate of the partner organisations but this is transparent and appropriate for the purposes of the study The process is practical, although the increase in study villages in the second generation of the initiative has caused some problems with manpower and coordination.

The hierarchy in the human/production system has been recognised with information collected (or compiled) at the district, village, farm, and plot levels. The problems being addressed can be decomposed into components that are directly observable and the initiative is able to detect change in the indicators or measures proposed within the temporal and spatial measuring frames. There is no evidence that proxy indicators have been sought for ‘fast’ variables that are difficult or costly to measure. The questionnaires and sampling methodology are standardised, but there is no information about how the monitoring initiative deal with confidentiality issues and other ethical considerations.

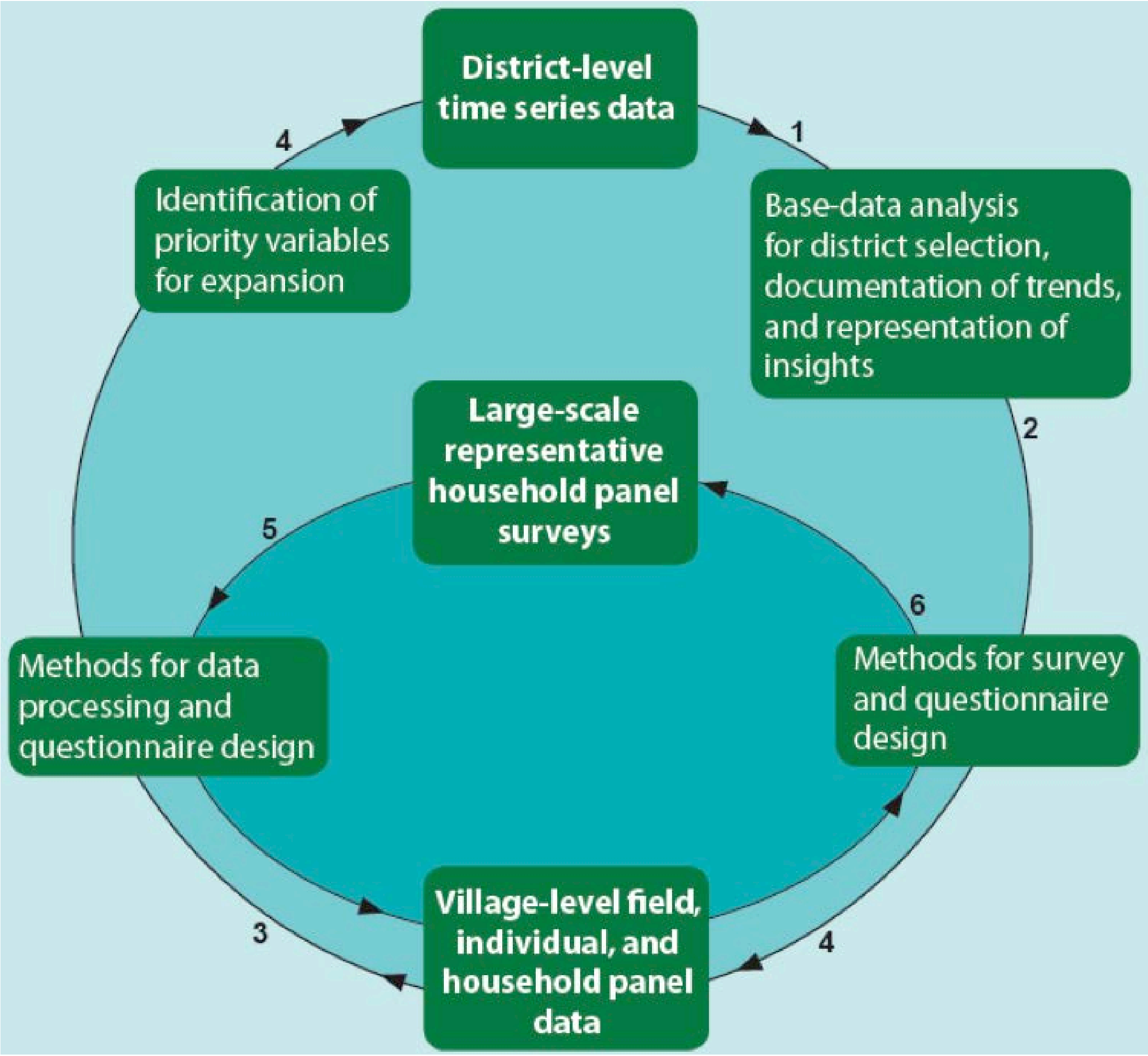

Figure 12. Linkages among spatial scales and instruments in Village Dynamics in South Asia.

No analysis of the monitoring has yet been completed but it is very likely that the initiative will need to employ some statistical analysis. In the documentation there is no evidence of explicit treatment of sustainability (e.g. NRM issues at the farm level) or of trade-offs among productivity, environmental and well-being objectives. It has also not been possible to assess whether the research will make use of the composite indices. The initiative potentially allows the relationships between research outputs and outcomes to be explored via a means-ends framework.

The survey is a longitudinal study and it is unlikely that the main variables would change during the short-term as a result of evaluation, but this cannot be confirmed. The trade-offs between the cost of measurement (frequency, access, equipment etc.) and the errors (of omission and commission) are partly addressed (also by LSMS-ISA) by investigation into the accuracy of recall by respondents of certain variables by comparing with diaries (in the case of diet and cropping activities) and by comparing farmers' assessments of plot sizes with other measurements such as GPS or manual measurements. It is possible to assess the trade-off between the value of the analysis of the information collected and the cost of measurement if the value of the monitoring initiatives can be measured in publications, and the other more tangible outcomes that resulted from the first generation of village level studies

Policies on data access are unclear, but funding was provided by BMGF so it can be assumed that data will be made freely available after one or two years. There are also activities in the VDSA workplan for data release. There is a VDSA database which may be accessible online via the VDSA website. Access to the interpretations of the results ought to be accessible via the publications page, but there are no recent publications.

Appendix 4 Insights from public health surveillance

The below materials are based on the below publications:

Shepherd K.D. et al., (2012a). General Concepts and Analytical Approaches In: UNEP 2012. *Land Health Surveillance: An Evidence-Based Approach to Land Ecosystem Management. Illustrated with a Case Study in the West Africa Sahel*. Nairobi: United Nations Environment Programme.

Shepherd, K.D. et al. (2012b). Land health surveillance and response: a framework for evidence-informed land management. Working Paper. Nairobi: World Agroforestry Centre (ICRAF), Draft.

Surveillance science

Public health surveillance, which has been active since the 1940s, has been defined in simple terms as “the routine, ongoing collection, analysis, and dissemination of data to those in public health who need to know”. It provides the scientific and factual foundation and database for informed decision-making and appropriate public health action. Surveillance information tells us where the problems are, whom they affect, where programmatic and prevention activities should be directed, and how well they are working.

Agricultural and natural resource management problems share many of the complex features of human health problems (Table 6), especially those of non-communicable diseases, and thus similar scientific approaches may be applicable (Table 7).

Table 6. Common complex features shared by public health and agro-ecosystem health

1. A rapidly increasing burden of health problems in developing countries, partly linked to demographic factors; and the problems become chronic if not addressed.
2. Problems often occur together as syndromes, with a common set of symptoms, rather than in isolation.
3. Health problems exist as a continuum and there is difficulty in defining the normal case and diagnosing poor versus good health.
4. Problems are associated with a range of physical, biological, social and economic determinants (or risk factors), both at individual and population levels.
5. Risks are often inter-related and act together to cause a health problem. They range from proximal risks acting directly to cause the problem, to distal risks that are further back in the causal chain.
6. Risk factors are often separated from outcomes in time, sometimes by many decades, making it difficult to establish causality.
7. Risks are generally greater for the disadvantaged in our societies, and poverty is a major risk factor.
8. Evaluating the cost-effectiveness of alternative preventive and rehabilitation interventions is complex and must consider many factors. Interventions may cause unintended side effects.
9. Different stakeholder groups (scientists, policy makers, public) perceive risks differently, requiring two-way communication processes.
10. Few resources are allocated to surveillance in developing countries, presenting

challenges such as how best to communicate surveillance and research findings in this setting, which still needs much local research.

Table 7. Science principles used in public health surveillance

1. A focus on the health of *populations* rather than individuals.
2. Sampling designs are used to make inferences from the sample to the population.
3. The frequency of health problems in populations, such as prevalence and incidence, is measured.
4. Standardized protocols are used for data collection.
5. Case definitions are used to specifically and consistently diagnose health problems.
6. Screening tests are used for assigning individuals to cases.
7. The association between health problems and risk factors is measured and assessed using statistical (risk quantification) models, including significance testing and confidence interval estimation.
8. Spatial and syndromic surveillance are commonly employed to detect spatial patterns in health problems and in symptoms that commonly co-occur.
9. Future needs are projected using mathematical forecasting models to detect patterns in data collected over time.
10. Cost-effectiveness analysis is used to screen alternative interventions.
11. Intervention impacts are rigorously evaluated using experimental designs with controls.
12. Meta-analysis is performed and used as the primary source of information for design of public policy and health programmes.
13. Operational surveillance systems are built into everyday health policy and practice, for both design and evaluation.

There are several key components of a surveillance and response framework (Figure 13). The basic surveillance work, shaped by feedback from by stakeholder consultations, aims to build up a good understanding of the current state and expected future trends in problems and their associated risk factors. This includes developing knowledge of the shape of the relationship between risk factors and problem intensity, which has a major bearing on the type of intervention strategy required. For example, a linear dose-effect relationship over the whole range of exposure implies that any level of the risk factor should be regarded as hazardous, pointing to the need for a population-wide strategy to reduce average levels of the risk factor, as opposed to targeting a high risk sub-population. Intervention targeting involves assessing the potential impact of interventions on problem burden and intervention cost-effectiveness. Considerable resources are assigned to evaluating interventions in the real world using study designs. Results are communicated to stakeholders in the form of actionable messages and responses also fed back into improved system design. Because health surveillance is by definition oriented towards action, the surveillance system itself is also evaluated, for example in terms of whether surveillance information has been communicated to those who need to know, and whether the information has had a beneficial impact on the health problem.

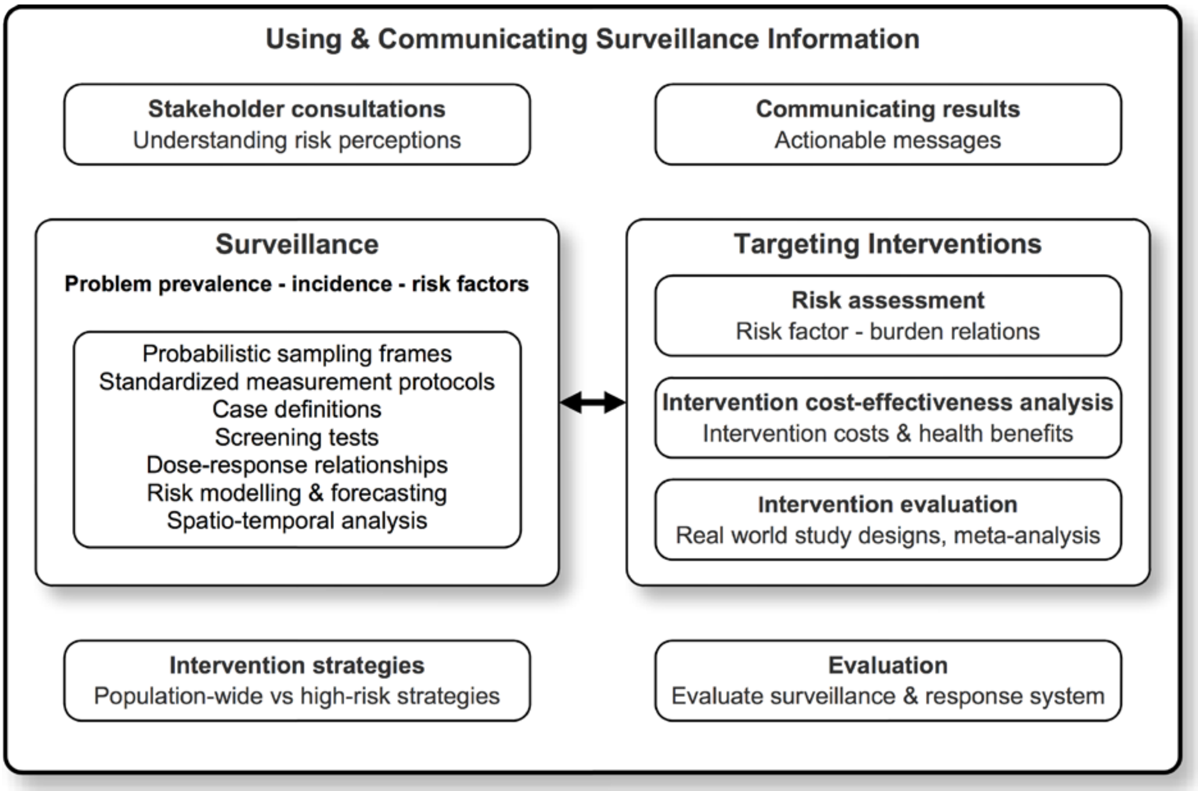

Figure 13. Key components of a health surveillance and response system. Connecting lines indicate two-way flows and iterations.

Intervention science

Surveillance is an essential tool for measuring the real world impact of interventions and public policy. Evaluation of intervention efficacy or invariably employs study designs to allow scientifically rigorous evaluation. Such studies are typically designed to statistically test the questions:

1. Is the post-intervention assessment significantly different from the pre-intervention assessment?

2. Is the change between the pre-intervention and the post-intervention assessment of the intervention group significantly different than that of the control or comparison group?

3. Is the outcome for the intervention group significantly different than for control or comparison group?

Designs vary in the strength of evidence they provide for causation, and to cater for different circumstances, such as ease of accessing sample units and ethical considerations. There are also many options with respect to randomized vs non-randomized studies, types of randomization (e.g. stratified, clustered, unequal allocation) and nesting. With intelligent design and analysis some evidence can be gained on intervention effectiveness in most situations, although difficulties increase for interventions operating on more distal factors. There is often opportunity for superimposing intervention evaluation studies on population-based surveillance studies, such as conducting case control studies to evaluate interventions after the fact.

Analytical tools are available to deal with a number of situations, to: (i) adjust for baseline differences between treatment groups, (ii) account for unmeasured confounding effects, (iii) match cases and controls on the basis of critical variables, (iv) investigate the effect of covariates, (v) conduct analysis on time series of observations on outcomes, and (vi) conduct sensitivity analysis with respect to input data.

Measuring intervention impacts on problem incidence in a real world setting often necessitates long-term studies (ten or more years), well beyond the life of a typical donor-funded project. Such studies may require, for example, multiple measurements of the outcome prior to and after the intervention. However early results can be achieved by identifying proximal markers that are highly predictive of the final outcome variables of interest, providing more outcomes in shorter follow-up times.

Randomized control trials (RCTs) are beginning to be used more frequently in poverty reduction studies (e.g. Poverty Action Lab; www.povertyactionlab.org/about-j-pal), but are still under-utilized in the agricultural sector. In contrast, evidence-based public health recommendations are strongly based on critical reviews of evaluations based on study designs, for example the Cochrane Collaboration (www.cochrane.org) provides a good model that could be applied to agro-ecosystem health, and the principles are already being applied to evidence-based conservation (http://www.cebc.bangor.ac.uk/).

Surveillance communication

Effective surveillance systems do not simply make the assumption that data will be used, but purposively plan use of surveillance findings (Table 8). Thus the design of surveillance systems must revolve around the recognition that the value of surveillance data comes from their use and the framing questions: "what information is demanded" and "who will use the data and how?".

Groups with an interest in surveillance data and information can be categorized into stakeholders, a subset of direct users, and those who have principal responsibility for defining and establishing the surveillance. Stakeholders and users may span a number of organizations including policy, research, academic, practitioners, community groups, private sector, and media. Successful surveillance systems find ways to manage the collaboration among these groups and ensure enduring partnerships. Although there has to be some centralized coordinating body or design team, involvement of target audiences in the design of surveillance systems is critical for effective utilization of results. There are challenges due to fact that decisions that affect agro-ecosystem health management are spread among different ministries (e.g. forestry, agriculture, livestock, water, environment, lands), and that agricultural, forestry and environmental extension infrastructures are mostly very weak in developing countries.

Making surveillance data more broadly accessible increases the value of surveillance systems. Surveillance information is a non-rival public good with multiple, primary, secondary and tertiary uses and users. Everyone can use the information freely without diminishing its value to any individual user.

Table 8. Considerations for communication of surveillance information

1. Design of communication processes must take careful consideration of the different perceptions of risk among different stakeholder groups (e.g. scientist, policy makers,

land users).

2. Recognizing and communicating uncertainty in surveillance findings is essential to creating openness and trust between the public and scientific community.
3. Trust is reinforced when results are released by an independent professional agency and presented by experts, rather than by politicians or political bodies.
4. Considerable resources may be required to achieve the essential social consensus required for tackling risks. Surveillance has a key role in identifying the few key risks on which governments should prioritize resource allocation.
5. Many of the most commonly used approaches to keeping practitioners informed (e.g. guidelines, reports, education programmes) have minimal impact or give mixed results, and more active dissemination strategies (e.g. delivery of messages to health professionals through medical representatives and key opinion leaders) are required to get guidelines used. However, research on such aspects in developing countries is still limited.
6. Policies are often impeded in areas where there is still scientific uncertainty over the importance of a problem and surveillance findings supplied to fulfil this need have a high chance of being used.
7. Representing uncertainty in surveillance data in presenting findings should be strongly encouraged and is increasingly being done, but there is also need for capacity building of decision-makers in how to interpret and use this information.

Surveillance insights

From reviewing surveillance science principles used in public health and their potential application to improving scientific rigour in agro-ecosystem health, a number of conclusions can be reached (Table 9).

Table 9. Science insights from public health surveillance in the context of agro-ecosystem health

1. Application of surveillance principles could accelerate progress in tackling major global problems related to agro-ecosystem health and management, especially in developing countries.
2. Standardized methods provide data that are more useful for planning successful interventions than locally adapted methods.
3. Risk factors are important over large enough areas and broad enough ranges of contexts to make broad scale surveillance and intervention planning useful.
4. A large part of the burden of agro-ecosystem degradation in many developing countries may result from a small number (5 to 10) of risk factors.
5. Interventions based on systematic surveillance will be more cost-effective than alternatives.
6. For diffuse degradation problems (such as loss of biodiversity, soil fertility depletion, and soil erosion) population level and preventive interventions will generally be more cost-effective than those focusing on high risk sub-groups.
7. Wider use of study designs in agro-ecosystem management projects will reduce the

time required to identify cost-effective interventions.

8. Better evidence on cost-effectiveness and impact of interventions will increase government and donor support for promising interventions.
9. Guidelines and standards for the design and consistent reporting of results of intervention studies will facilitate meta-analysis for policy guidance.
10. National agro-ecosystem health surveillance systems could be established at relatively low cost by re-aligning existing resources.
11. An international scientific and technical support unit could efficiently build capacity in surveillance systems, improve data and information quality, and facilitate international meta-analysis.
12. An international advocacy body, such as the United Nations Organization, could effectively accelerate adoption of national surveillance systems and promote global standards.

Appendix 5 Applied Information Economics

For application of approach, see separate file: The Need for an Intervention Decision Model.pdf

For example see file: Applied Information Economics Example

The basic Applied Information Economics approach has the following steps (based on Hubbard, 2010):

1. **Clarify the decision problem**. The first question to ask is what is your dilemma? If you are asking "How do we measure X?" then you are putting the cart before the horse". All the variables relevant to a dilemma are defined and ambiguities, such as "how to measure sustainability", are resolved.

2. **Determine what you know**. Quantify your uncertainty about unknown quantities in the identified decision. This is done by representing existing knowledge, including knowledge of experts, in terms of ranges or probabilities. Anything that you think is important to a decision can be included, however seemingly intangible and however uncertain you may be about its range. The amount of uncertainty helps determine how much risk (or insecurity, such as food insecurity) is involved. This is typically done in a spreadsheet modelling tool.

3. **Compute the value of additional information.** Information has value because it reduces risk in decisions. A Monte Carlo simulation is run that plays through the probability distributions to provide a risk-return or return-on-investment distribution. Decision maker preferences (e.g. risk averseness) are normally included. The spreadsheet computes the value of information for each variable. This guides us as to what to measure (or research), how to measure it, and how much we should spend on measuring it.

4. **Measure the high information value variables**. Prioritize measurements (or research) according to the ratio of information value to cost[7]. The measurement pecking order is usually: try variable decomposition, next historic data, and finally surveys or trials. You iterate the model at each step as new information is acquired (Bayesian updating) and rerun the model to see what residual information values exist and determine whether further measurement or research is justified.

5. **Make a decision and act on it.** When the economically justifiable uncertainty has been removed, decision makers face a risk-versus-return decision. The preferences and attitudes of the decision maker can be included in the analysis. The variables with highest information values are also important variables to monitor during implementation, as they are the most likely ones to go off track.

Connecting interventions to CGIAR development outcomes

The CGIAR Program on Water, Land and Ecosystems is applying the applied information economics framework to establish what are the high value information and measurement needs of the range of interventions being considered by the program. The Intervention Decision Model (Appendix 5 Applied Information Economics) will be the basis for prioritizing

[7] Measurements that can reduce the uncertainty of high value variables the most at lowest cost of measurement are accorded highest priority.

interventions, determining how to measure the impact of interventions on development outcomes, and calculating the value of the research itself.

Some of the potential benefits of the approach are:

1. Provide a means for including the full range of on- and off-site costs and benefits for interventions, and their trade-offs, including uncertainty; and as a result make better intervention decisions.

2. Pinpoint high value information and research needs for interventions, and especially reveal counter-intuitive ones.

3. Make a sound business case to governments, investors and other stakeholders on proposed interventions – thus enhancing potential impact. Also a tool for engaging stakeholders.

4. Make a sound business case to donors on the value of our research and forecasted impacts on development outcomes.

5. Iteratively improve intervention designs by focusing on variables that are responsible for producing negative outcomes or higher risks.

6. Develop plans for monitoring intervention programmes based on those variables most likely to go off track during implementation. Set up initiatives to track how good are your intervention forecasts. Iterative learning/improvement.

7. Improve multi-stakeholder decision processes by revealing preferences and trade-offs.

Appendix 6 Methodological issues of composite indices

This summary is based on an online discussion between Sabina Alkire and Martin Ravallion and on the new ‘Multidimensional Poverty Index’ (http://www.oxfamblogs.org/fp2p/?p=3092) as well as OECD Handbook on Constructing Composite Indicators (www.oecd.org/dataoecd/37/42/42495745.pdf)

Composite indicators (CIs) are increasingly recognized as a useful tool in policy analysis and public communication. The number of CIs in existence around the world is growing year after year (for a recent review see Bandura, 2006, which cites more than 160 composite indicators). Such composite indicators provide simple comparisons of countries that can be used to illustrate complex and sometimes elusive issues in wide-ranging fields, e.g., environment, economy, society or technological development.

Essentially composite indicators describe a phenomenon that consists of independent sub-components that can be measured each by individual indicators. The composite indicators is calculated be applying appropriate weights to each of the sub-component indicators and aggregate them to a single composite index. One of the most commonly known composite indices is The Human Development Index (HDI), which consist of life expectancy, education, and income indices. It was created by economist Mahbub ul Haq, followed by economist Amartya Sen in 1990 and published by the United Nations Development Programme to rank countries into four tiers of human development.

While composite indices are easy to interpret and provide good summaries of otherwise, complex multidimensional realities there are a number of methodological issue that need attention. The most important ones are summarized in here:

- Need for a strong conceptual framework

The quality of a composite indicator as well as the soundness of the messages it conveys depend not only on the methodology used in its construction but primarily on the quality of the framework and the data used. A composite based on a weak theoretical background or on soft data containing large measurement errors can lead to disputable policy messages, in spite of the use of state-of-the-art methodology in its construction.

- Setting appropriate thresholds for each indicator

In order to aggregate the individual sub-indicators, a threshold defining the status must be defined. The thresholds constitute the standards to be compared with the actual situation of the unit of analysis (household/ecosystem). For each of these standards the unit is classified as complying or not complying. Establishing the standards for the different indicators can be based on an absolute or a relative view. The relative approach entails sampling the population first and establishing the threshold relative to the sampling results. A threshold is absolute when it is fixed independently of the indicator’s actual distribution among the population. Examples were absolute thresholds are usually used are needs assessments such as nutrition, education and health.

- Weighting & Aggregation

The Oxford Poverty and Human Development Initiative (OPHI) has recently launched a Multidimensional Poverty Index (MPI), and calculated it for over 100 countries. The MPI is a composite of indicators selected for consistency with the UNDP’s famous Human Development Index (HDI). The HDI uses aggregate country-level data, while the MPI uses household-level data, which is then aggregated to country level. The index has ten components; two represent health (malnutrition, and child mortality), two are educational achievements (years of schooling and school enrolment), and six aim to capture “living standards” (including both access to services and proxies for household wealth). The three broad categories–health, education, and living standards–are weighted equally (one-third each) to form the composite index. The index has been critesed to compare “apples and oranges” as it adds up multiple, non-comparable, dimensions of poverty; e.g. the death of a child is not equivalent to having a dirt floor.

Weighing of individual sub-indicators is another problem. Income poverty aggregates within a country using actual or imputed prices for fixing the income poverty standard across countries and time. The World Bank’s “$1 a day” poverty measure relies on economic theory, which says that (under certain conditions) market prices provide the correct weights for aggregation. While setting prices is not unproblematic in practice most complex phenomena do not have a scientific theory to rely on weights are set as value judgments. A decision has to be taken, and no consensus exists on how the multiple dimensions should be weighted to form the composite index.

- Availability of data

In order to construct a composite index based on several sub-indicators, these sub-indicators need to be available at the same unit of observation; e.g. household, ecosystem, farm level. While scientific sound composite indices can be developed from strong conceptual frameworks, in reality the best indicators are often not available for the unit of analysis. Thus the choice of the precise indicators used are in fact not chosen because they are the best available data on each dimension of phenomenon to be evaluated, but because they happened to be the indicator available for that unit of observation.

As indices increase in value with increasing time horizon of the measurement another fallacy is that new indices try to build on existing ones in order to keep the element of comparativeness. Thus there is a strong trade-off between developing a better index based on advanced theoretical understanding and better quality data and its ability to observe historical trends.

- Computation methods

Composite indicators are much like mathematical or computational models. As such, their construction owes more to the craftsmanship of the modeller than to universally accepted scientific rules for encoding.

The choice of methodology can have significant effects. For the 2012 report on The State of Food Insecurity in the World 2012 (FAO, 2012) used a new methodology to estimate the prevalence of undernourishment. “The new estimates suggest that the increase in hunger during 2007-2010 was less severe than previously thought,” the report says. The effect of high prices during the 2008-2009 economic crisis “was less pronounced than was assumed at the time, while many governments succeeded in cushioning the shocks and protecting the most vulnerable from the effects of the price spike.”

Appendix 7 Information systems for biodiversity in agro-ecosystems

Plant Genetic Resources

Plant genetic diversity for food and agriculture constitutes an important element for the livelihood strategies of smallscale farmers and continues to be fundamental in trying to achieve global food security. Crop genetic diversity provides adaptation to marginal or heterogeneous environments, rainfall variability, variable soil types, and can provide insurance against environmental risk. This diversity allows farmers to have the option of meeting changing market demands, adapted to changes in adult labor availability, respond to social and cultural obligations, and to improve dietary diversity and ensure nutritional well being. The utility of crop varietal diversity within the production system also has the potential to provide ecosystem services, to regulate pest and diseases, to support below ground biodiversity and soil health, and to sustain pollinator diversity. This can in turn reduce the financial and health risks of high levels of agricultural inputs, such as fertilizer and pesticides to small-scale farmers and the environment. Several global information systems are in place to monitor the amount and distribution of this diversity maintained both by farmers in situ and by genebanks ex situ.

World Information Sharing Mechanism on the Implementation of the Global Plan of Action for the Conservation and Sustainable Use of PGRFA (WISM-GPA)
http://www.pgrfa.org/gpa/selectcountry.jspx

At the FAO International Conference on Plant Genetic Resources, in 1996, following the adoption of the rolling Global Plan of Action (GPA), 150 countries agreed that "overall progress in the implementation of the GPA and of the related follow-up processes would be monitored and guided by the national governments and other Members of FAO, through the Commission on Genetic Resources for Food and Agriculture" (http://www.fao.org/nr/cgrfa). To this end, the Commission was asked to set the formats for receiving progress reports from all the parties concerned and establish criteria and indicators to assess progress in the implementation of the GPA.

This has been implemented through the development of WISM-GPA. WISM-GPA provides access to National Information Sharing Mechanisms (NISM)' portals and databases on conservation and sustainable use of PGRFA, established by 73 countries worldwide with the participation and contribution of more than 1,150 public institutions, non-governmental and private organizations, including farmers' associations, from the PGRFA world community that, day by day, conserve, monitor, multiply, improve, exchange and make available these resources essential to our and our planet's life. The NISMs are either hosted or mirrored by the World Information Sharing Mechanism (WISM), which during 2011-2012 has been visited by more than 340,000 users for about a million of downloaded pages. The interface of both the Web and stand-alone GPA monitoring systems now function in 26 different languages.

Main components of the National Information Sharing Mechanisms (NISM)

- A list of indicators for monitoring the implementation at country level of all priority activity areas of the Global Plan of Action;
- A reporting format, which is a structured questionnaire based on these indicators;
- A computer application, which has been developed to facilitate and simplify recording, processing, analysis and sharing of the information addressed by the indicators and the questionnaire; and
- Guidelines for initiating and coordinating this process, including guidelines for the involvement of stakeholders and for establishing a national information-sharing mechanism.

Under the WISM-GPA database there are at present more than 125,000 answers to the questions of the reporting format for monitoring the implementation of the 20 priority activity areas of the Global Plan of Action. These answers have been provided by more than 1,100 stakeholders from the 71 countries, where the National Information Sharing Mechanism has been established.

In 2011, FAO member countries agreed to a Second Global Plan of Action on Plant Genetic Resources for Food and Agriculture. The Second GPA contains 18 priority activity areas and reflects the new challenges and opportunities that have emerged with respect to the conservation and use of PGRFA since 1996, including climate change, as well as new developments in the policy environment, in particular the entry into force of the International Treaty on Plant Genetic Resources for Food and Agriculture. For monitoring the implementation of the Second GPA, the indicators and the reporting format are being revised to adequately reflect the changed priority activities of the Second GPA. Such indicators, once approved, will also contribute to the Aichi Biodiversity Targets of CBD and assess the contribution of PGRFA to food security and sustainable agricultural development.

WIEWS World Information and Early Warning System on Plant Genetic Resources for Food and Agriculture (http://apps3.fao.org/wiews/wiews.jsp)

The World Information and Early Warning System (WIEWS) on Plant Genetic Resources for Food and Agriculture (PGRFA), established by FAO, is a world-wide dynamic mechanism to foster information exchange among Member Countries, by gathering and disseminating information on plant genetic resources for food and agriculture, and as an instrument for the periodic assessment of the State of the World's Plant Genetic Resources for Food and Agriculture.

It is an important component of the FAO Global System on plant genetic resources for food and agriculture. WIEWS comprises the following databases:

- the Country Profiles database, including the structure of 190 national PGR programmes and activities;
- the Ex situ Collection database, containing summary records of PGR holdings (more than 5 million accessions belonging to more than 18 000 species) reported by more than 1 500 national, regional or international genebanks;
- the PGRFA and Seed Laws and Regulations database (70 countries);
- the World List of Seed Sources database (approx. 8 000 entries from 150 countries); and
- the List of Crop Varieties database (about 65 000 varieties from 1 249 cultivated crops).

Animal Genetic Resources

Livestock biodiversity is essential to food and livelihood security, particularly in the developing world. Livestock provide meat, milk, eggs, fibres, skins, manure for fertilizer and fuel, draught power for cultivation and transport, and a range of other products and services. Many of the world's rural poor – an estimated 70 percent – keep livestock and rely on them as important components of their livelihoods. Domesticated animals also contribute to the ecosystems in which they exist, providing services such as seed dispersal and nutrient cycling. Genetic diversity underpins the many roles that livestock fulfil and allows people to keep livestock under a wide variety of environmental conditions. As a result, domestic animals survive in some of the most inhospitable areas on Earth – from Arctic tundras and high mountains to hot dry deserts – where crop production is difficult or impossible.

Livestock exposed to extreme climatic conditions develop adaptive characteristics that help them survive and produce where other animals would succumb. They adapt to local feed resources and develop resistance to diseases and parasites.

The Domestic Animal Diversity Information System

The Domestic Animal Diversity Information System (DAD-IS: http://www.fao.org/DAD-IS) is a multilingual (currently available in English, French and Spanish), dynamic database-driven web-based information system supporting the implementation of the Global Plan of Action for Animal Genetic Resources (adopted in 2007). The backbone of DAD-IS is the Global Databank on Animal Genetic Resources, into which countries enter data on their livestock breeds, including data on the size and structure of breed populations. Countries take full responsibility for the completeness and quality of the data entered. More than 14 000 national breed populations (representing more than 8 000 breeds) from 182 countries are recorded. DAD-IS is the centre of an expandable global network of information systems that exchange breed-related data. A regional system (the European Farm Animal Diversity Information System –EFABIS) has been established in Europe and national systems have been established in a number of European countries. Data from DAD-IS have for many years been used as a basis for reporting on the global status and trends of animal genetic resources – three editions of the *World Watch List for Domestic Animal Diversity* (1993,1995, 2000), *The State of the World's Animal Genetic Resources* (2007) and biennial status and trends reports produced for the Commission on Genetic Resources for Food and Agriculture (2008, 2010,2012). A set of indicators for monitoring the status of the world's animal genetic resources, to be based on DAD-IS data, has been developed. In addition to the breed databank itself, DAD-IS features a library of full text publications and links to other web-based resources, a number of tools for analysing breed population data, news on events and developments in animal genetic resources management, and the contact details of National Coordinators for the Management of Animal Genetic Resources.

Aquatic Genetic Resources:

Maintaining aquatic biodiversity, including fish genetic diversity, in capture fisheries is fundamental to guaranteeing the productivity of fish stocks, their resilience and their adaptability to environmental change. The first State of the World's Aquatic Genetic Resources for 2013, the Commission has launched a review of existing information systems, and will work to develop a more streamlined reporting system for national and international organizations. With the number of farmed fish strains, hybrids and other genetic resources increasing in aquaculture, information systems are needed to identify and determine their relative contributions to farmed fish production. Similarly, better information on the genetics of wild fish populations should contribute to better understanding the needs for conservation and sustainable use.

FishBase (http://www.worldfishcenter.org/fishbase) Biodiversity Information System

FishBase is a Biodiversity Information System on all fishes of the world (finfishes), covering over 32,000 species. FishBase is a website for fish-related natural resource management both for conservation and exploitation issues. It is used extensively by fisheries managers and scientists in the developing world to estimate important biological parameters like mortality, annual reproductive rate, growth, and status of fish stocks. As a Global Species Database, it records a wide range of information on all fish species currently known in the world about their biology, ecology, taxonomy, life history, trophic features, population dynamics and uses, as well as historical data reaching back 250 years, citing at least 47,000 references. More than 51,000 images are gathered for more than half of the species, with the help of many contributors. FishBase provides also a range of country, regional, and ecosystem-specific information. State of the art analytical and graphical tools allow users to transform raw data into information that can be used to assess fisheries and identify management techniques to restore depleted fish stocks.

AlgaeBase is a database of information on algae that includes terrestrial, marine and freshwater organisms. At present, the data for the marine algae, particularly seaweeds, are the most complete. For convenience, we have included the sea-grasses, even though they are flowering plants. At present 131,891 species and infraspecific names are in the database, 16,205 images, 48,616 bibliographic items, 215,080 distributional records.

FishStat Plus (http://www.fao.org/fishery/statistics/software/fishstat/en)

This global database is part of the **FIGIS** and includes :

- Total Fishery Production, 1950-2010
- Capture Production, 1950-2010
- Aquaculture Production (Quantities and values),1950-2010
- Fisheries Commodities Production and Trade. 1976-2009 and
- Several regional databases from Regional Fishery Bodies.

FIRMS Fishery Resources Monitoring System (http://firms.fao.org/firms/en)

FIRMS is a resource monitoring system which provides access to a wide range of high-quality information on the global monitoring and management of fishery marine resources. Information provided by the partners is organized in a database and published in the form of fact sheets. FIRMS draws together a unified partnership of international organizations, regional fishery bodies collaborating within formal agreement to report and share information on fisheries resources. For effective fisheries information management, FIRMS also participates in the development and promotion of agreed standards. FIRMS is part of the Fisheries Global Information System (FIGIS) http://www.fao.org/fishery/figis/en) and linked to the National Aquaculture Sector Overview (NASO) (http://www.fao.org/fishery/naso-maps/naso-home/en/)

Fisheries Global Information System (FIGIS) is a single entry point to an integrated system comprising strategic data, information, analyses and reviews of issues and trends on a broad range of fisheries subjects. The FIGIS system provides Internet-based tools to promote more effective and focused cooperation between information sources, within FAO as well as between FAO, regional fisheries management organizations and national centers of excellence. Included in FIGIS are species fact sheets for both aquaculture and capture fisheries (http://www.fao.org/fishery/factsheets/en).

National Aquaculture Sector Overview (NASO), part of FAO Fisheries and Aquaculture Department regular programme. The NASOs contain detailed information on the history of aquaculture; human resources involved in the sector; farming systems distribution and characteristics; main cultured species contributing to national production; production statistics; description of the main domestic markets and trade; promotion and management of the sector; and development trends and issues at the national level.

Appendix 8 UN Global Pulse

Global Pulse is an innovation initiative in the Executive Office of the UN Secretary-General, harnessing new digital data to help the public sector understand sooner how communities are coping with the impacts of global crises. To this end, Global Pulse is developing a 21st-century approach to impact monitoring, building social software for evidence-based decision-making, and establishing a network of "Pulse Labs" to mainstream the use of real-time data analysis into development policy. Pulse Labs are national facilities established to support the public sector by developing analytical and technological capacity for harnessing real-time data to strengthen social protection.

The Assistant Secretary-General for Policy and Planning, Robert C. Orr explains the UN Global Pulse initiative in the following terms: "Global Pulse's mission is to help us seize this historic opportunity to improve how we combat hunger, poverty and disease – getting there, however, requires first building awareness of the opportunity in both the big data community and the development community, forming strategic partnerships, developing innovative approaches, and demonstrating their potential to change outcomes."

Produced by UN Global Pulse, the report – *Big Data for Development: Opportunities and Challenges* – explores how data from social media and other sources can yield information about population well-being, including responses to job losses, changing food prices and other economic indicators. The report features a particularly salient example of how Social Media Analytics have been applied to food security issues.

Food Price Inflation
A recent joint research project between Global Pulse and social media analysis firm Crimson Hexagon analysed over millions of Tweets related to food in Indonesia and analysed trends in these topics in conjunction with themes such as "afford," showing how the volume and topics of the conversations changed over time reflecting populations' concerns. Interestingly, changes in the number of Tweets mentioning the price of rice and actual food price inflation (official statistics) in Indonesia proved to be closely correlated (Figure 14). A linear regression of using the twitter topic categories to estimate food price inflation values from the national government demonstrated 89 percent predictive power.

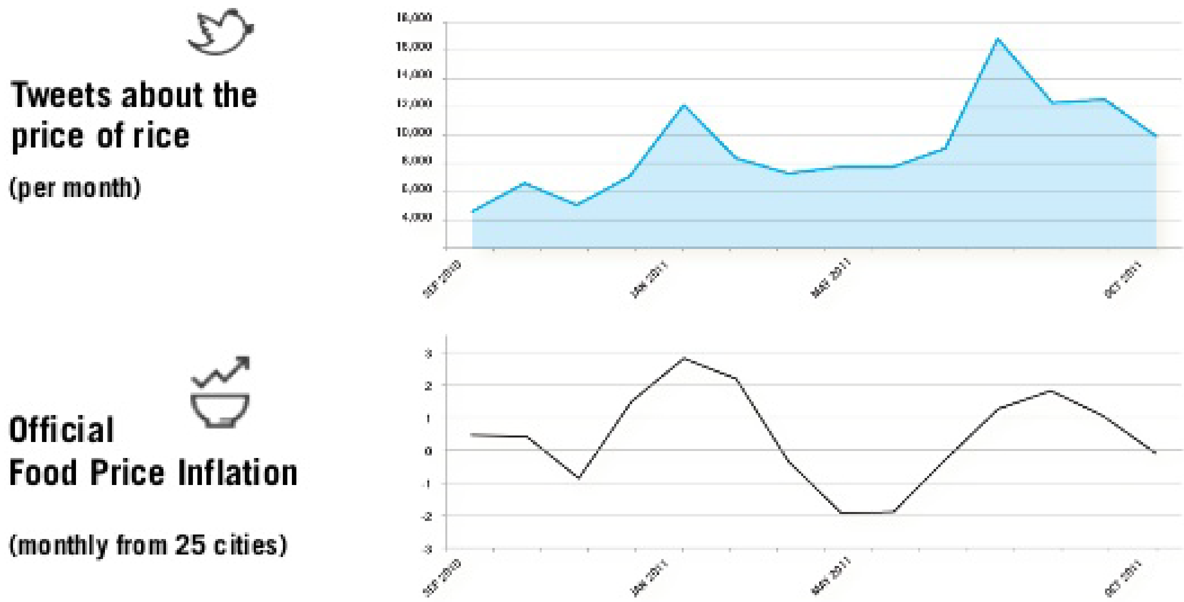

Figure 14. Indonesian Twitter Social Media correlation with food-price-basket inflation (rice, chilies, fish, sugar, corn, cooking oil).

Pulse Labs
UN Global Pulse has established 'Pulse Labs' in Jakarta, Kampala and New York. Pulse Labs are in-country innovation centers that will tap into local knowledge and innovation, establish key partnerships, pilot real-time monitoring approaches at the country level, and support the adoption of proven approaches globally. Approaches for exploration include: Social Media & Twitter Analytics, Aggregated Mobile Phone Data Analysis, Rapid Mobile Surveys, Geo-Spatial Mapping, and Development of a Gazetteer Framework (http://www.csiro.au/gazetteer).

Pulse Lab Jakarta is forming partnerships with Indonesian-based companies and organizations to collaborate on research in the following ways: Become a Data Philanthropy partner and share your data for analysis, Become a Research partner and work on a project with the Pulse Lab, Become a Technology partner and build prototypes or test new tools, Host a training session on Big Data analysis tool or method, Sponsor a Fellow to join the Pulse Lab and work on data visualization, data analysis, GIS mapping, or software development, Provide funding, or in-kind, support for the Pulse Lab's work. Initial funding support for Pulse Lab Jakarta has been provided by contributions from AusAID, the Government of Indonesia's Ministry of National Development Planning, and UNICEF.

In Kampala, there is strong support for a Pulse Lab by the National Planning Authority (NPA) of the Ugandan Ministry of Finance, Planning and Economic Development; the Ugandan Bureau of Statistics; the UN Resident Coordinator; and Makarere University among others. The Lab will build on UNICEF's work in Uganda on technology innovation and mobile phones for rapid data collection and analysis.

Data Philanthropy
http://www.unglobalpulse.org/blog/data-philanthropy-public-private-sector-data-sharing-global-resilience
Finally, UN Global Pulse is asking companies to consider a new kind of Corporate Social Responsibility referred to as "data philanthropy" by making anonymized data sets available for analysis, by underwriting technology and research projects, or by funding our on-going efforts in Pulse Labs. The same technologies, tools and analysis that power companies' efforts to refine the products they sell, could also help make sure their customers are continuing to improve their social and economic wellbeing. Governments and the United

Nations can use these initiatives to become more agile in understanding the needs of and supporting the most vulnerable populations around the globe, which in terms boosts the global economy, benefiting people everywhere.

D4D Challenge: Data for Development - http://www.d4d.orange.com/home

The Orange “**Data for Development**” - D4D - is an open data challenge, encouraging research teams around the world to use four datasets of anonymous call patterns of Orange's Ivory Coast subsidiary, to help address society development questions in novel ways. D4D is led by global telecommunications company France Telecom-Orange, in partnership with GSMA Development Fund, MIT MediaLab, University of Louvain (UCL), Université de Bouaké, the World Economic Forum and UN Global Pulse. The data sets are based on 2.5 billion anonymized Call Detail Records extracted from Orange’s customer base, covering the months of December 2011 to April 2012, the largest ever released mobile phone dataset.

Research teams wishing to take on the challenge and participate to the development of Ivory Coast society have access to the data to analyse it and cross-compare it with other types of data to find useful insights. The goal of the D4D challenge is to contribute to the socio-economic development and well-being of populations. Knowledge of typical behaviours of mobile telephone users can be very useful, for example to identify early signs of epidemics, to be reactive in times of crisis, to measure the threat and resultant impact of droughts, to optimize the usage of certain infrastructures, etc.

The data collection took place in Cote d’Ivoire over a five-month period, from December 2011 to April 2012. The original dataset contains 2.5 billion records, calls and text messages exchanged between 5 million users. The customer identifier was anonymized by Orange Cote d’Ivoire. All information about user’s whereabouts were recorded when a user sent or received a call or a text message (SMS). The cell tower that routed the call is recorded. We provide slightly blurred geographic locations for about 1,200 towers. Random user identifiers have been generated separately for each dataset.

Aggregate communication between cell towers.
For every one-hour period we provide the number of calls and the total communication time between every pair of cell towers. We also provide the information about which cell tower initiated the call. Calls starting in a particular one-hour period are associated with this time period, irrespective of their termination time.

Mobility traces: fine resolution dataset.
In every two-week period, we choose a random sample of active users. We provide timestamps and the cell towers those users made calls and texts from during the two-week period. This process is repeated for successive two-week periods with different random samples.

Communication sub-graphs for Social Network Analysis.
We have chosen several thousand of random users (egos) for whom we constructed communication graphs obtained by considering all communications between ego and his/her contacts at up to two degrees of separation from the ego. The communications inside each sub-graph are aggregated by two-week time window over five months. Random identifiers are attributed separately in each graph and remain unchanged over entire observation period.

The organizing committee will invite the winning researchers to present their projects at the NetMob International Conference on "Analysis of Mobile Phone Datasets and Networks", scheduled for Spring 2013, at which the results will be presented to the scientific community.

Appendix 9 Detailed recommendations

Good practice guidelines for monitoring initiatives

Based on the main weaknesses identified in the review, a set of good practice guidelines are advocated in Table 10. The criteria given in Table 1 can also be helpful.

Table 10. Good practice guidelines for design of monitoring initiatives

- Develop a clear conceptual framework to show current understanding of the system under study.
- Develop a means-end objective framework to show how objectives relate to each other and how the set of indicators measure their degree of achievement.
- Have a clear theory of change on how monitoring results will be used
- Involve stakeholders/users early on in the design process to ensure that they have a stake and own the data that is being collected both to improve performance of the monitoring initiative and to ensure that the data is used.
- Define the inference space in terms of the geography or population being sampled. This is the target area or population to which results will be applied.
- Sample the inference space or the variation in major factors influencing the variables of interest. Use stratification of major influencing variables to reduce within stratum variance.
- Clearly define sample units so that they can be unambiguously identified and consistently measured.
- Select indicators that:
  - Measure the achievement of the fundamental objectives or outcomes.
  - Measure performance of drivers as well as states of a system and consider scale hierarchies.
  - Capture biophysical and socioeconomic factors under a unified framework. Focus on behavioural factors that produce outcomes.
  - Are interpretable and have defined thresholds for action.
  - Have meaning to end users and are likely to relate in change in management or policy decisions.
- Use standardized measurement protocols and consistently apply them over space and time.
- Ensure that remote sensing measurements are adequately calibrated and validated with ground data.
- When monitoring interventions, use study designs with controls or counterfactuals and replication wherever feasible.
- Have a clear plan for statistical analysis of the results and consult a statistician at the design stage.
- Ensure that uncertainties in indicators are adequately represented when communicating results.
- Produce a data sharing agreement in participation with stakeholders during the design phase. Develop mutually acceptable agreements with traditional knowledge holders.
- Consider how the monitoring initiative will be sustained over the long term.
- Consider how trade-offs will be represented and analysed. Separate objective and

preference trade-offs.
- Build in procedures for evaluating the monitoring initiative itself, including costs and impacts.
- Incorporate decision analysis methods to focus the effort and increase efficiencies.

What should research organizations and those designing monitoring initiatives do differently?

Based on the findings of the review recommendations for research organizations and those designing monitoring initiatives are given in Table 11.

Table 11. Recommendations for research organizations

- Fully implement the above good practice guidelines in monitoring initiative design and execution.
- Develop capacity in probabilistic decision analysis and Value of Information Analysis.
- Use these tools to forecast impacts of alternative interventions on development outcomes and identify high value research areas and information products, based on Value of Information Analysis. This should include analysis of interventions at different scales (specific technologies, farm household decisions, large scale programmatic interventions, information systems).
- Deploy risk-return analysis of specific technologies to inform decision makers on whether technologies are ready for implementation or whether further research is warranted.
- Use probabilistic forecasting models to provide a business case (risk-return analyses) for research or dissemination projects, including for example schemes for providing rewards for ecosystem services.
- Use the models to guide and help prioritise long-term monitoring initiatives as well as justify investments in new measurements, including improving efficiency of statistical designs.
- Include a decision analysis component in new research and development projects to engage key stakeholders, help identify research priorities, and provide a framework for forecasting and monitoring impacts.
- Produce and implement guidelines on use of study designs and standard reporting methods for intervention evaluation, based on approaches used in public health surveillance.

What should DFID and other donors consider doing differently?

Based on the findings of the review recommendations for DFID and potentially other donors are given in Table 12.

Table 12. Recommendations for DFID

- Support the use of the above good practice guidelines in supported projects.
- Consider using probabilistic decision analysis to design their project portfolios based on forecast impacts on desired development outcomes. This would provide a solid business case for investments.

- Support joint work with research organizations to forecast intervention impacts and justify further research and measurement initiatives.
- Invest in monitoring initiatives that are designed to validate model forecasts of intervention impacts.
- Support capacity building of stakeholders in probabilistic decision analysis to improve structured and evidenced-based decision-making. Capacity building for high-level decision makers might focus on simple awareness raising on decision-making under uncertainty and on how to work with scientists and decision analysts in this area. More intensive capacity building might focus on funding decision analysts/scientists to work at high levels in government departments and other institutions, especially on analysis of integrated interventions across sectors.
- Support the development and integration of probabilistic decision analysis and Value of Information Analysis in university curricula, especially agricultural, environmental natural resource, social, and economic sciences. Encourage new systems science disciplines that integrate these different sciences for development.
- Harmonize reporting requirements of grantees around key indicators by sector so that analyses can be made over time and across countries regardless of donor or implementing organization. Requiring fund recipients to use the same outcome-indicators across projects and in coordination with global national statistical institutions (such as the Millennium Development Indicators) will facilitate consistent monitoring and evaluation, and allow for systematic reviews of environmental intervention approaches.
- Require proposals to include project close-out stages requiring implementers to document and hand off data to appropriate institutions that will maintain data over the longer term. Moving data onto platforms independent of individual projects is essential for dissemination, reuse and sector-wide learning

New conceptual approach to measurement and monitoring

There appears to be a major conceptual limitation in the approach to measurement and monitoring with respect to expected impacts on real world decisions. If there is no vision of how monitoring data will influence decisions or outcomes (i.e. a theory of change) then monitoring initiatives have no value beyond scientific curiosity. If there is no evidence for how monitoring initiatives having changed decisions or influenced outcomes, then initiatives have no economic value, and without some at least projected impact there is no basis for evaluating their cost effectiveness.

It has been difficult to find evidence to support the widespread perception of the usefulness of current monitoring initiatives and indicators of agro-ecosystems and livelihoods. Indeed some of the most experienced scientists in this field consider that monitoring efforts have had little or no impact on decision-making. There is almost no data or analysis on impacts or cost-effectiveness of existing or past initiatives.

The current practice is to design measurement and monitoring initiatives, or to develop indicators from existing initiatives, and then see how to interpret and communicate results to influence management and policy decisions. Measurements are designed on the assumption that they will have value in improving decisions, whereas many studies have shown that our intuition in judging what measurements have potential value towards changing decisions is often wrong. Perhaps it is not surprising that policy and management decisions are making only limited use of scientific evidence, if we are providing low value information.

There is a general lack of perception about the real purpose of measurements within a decision context – that is to reduce uncertainty in variables that could result in a high cost of being wrong and where there is a high chance of being wrong. Conceptual frameworks of past and current initiatives do not explicitly embrace this concept.

Our actions are dictated by our beliefs but monitoring initiatives up until now do not start by representing and analysing our current beliefs and our degree of uncertainty about them. If there is no uncertainty then further information has no value. Thus critical analysis of uncertainty should be the starting point for the design of any measurement initiative. But knowing the degree of uncertainty about a variable is not sufficient. The critical question is whether further reduction in uncertainty of a variable has the potential to affect outcomes. In other words, would a decision be taken differently as a result of being more certain about a variable, and if so would that have led to important changes in outcomes?

As a simple example, if a donor is seeking to prioritise funding allocations for sustainable intensification of agriculture in Africa based on the distribution of poverty levels, gathering further information on poverty distribution will only have value if it has potential to change the decision compared with using existing information on poverty distributions with its current level of uncertainty.

The above implies that no matter how uncertain we are about the impact of actions or interventions on outcomes, we should be forecasting their impacts, using our current levels of uncertainty. In other words providing a quantified theory of change with uncertainties represented. Only in this way can we know the value of gathering more information or conducting more research. The approach is iterative in that we update our degree of belief in a variable by reducing uncertainty through measurement and then evaluate the value of further reducing uncertainty through more measurement. The analytical framework is Bayesian decision theory and Value of Information Analysis. This framework is increasingly being used to prioritise public health interventions and research; and Applied Information Economics is a practical application of the framework that has been used in a number of different fields.

Hence a principal recommendation of this review is to advocate Value of Information Analysis for the design of measurement initiatives that seek to answer questions like:

- How to evaluate alternative research and development strategies in terms of their potential impact on productivity, environmental services and welfare goals, including trade-offs among these goals?

- How to cost-effectively measure and monitor actual effectiveness of interventions and general progress towards achieving sustainable development objectives?

Thus the first question can be reformulated as a decision problem – given a set of research and development alternatives, what is their potential impact on productivity, environmental services and welfare goals, and how much risk or uncertainty is there on these returns? Forecasting can in fact be considered as a form of measurement of likely future outcomes, when measurement is seen as a quantified reduction in uncertainty.

If we cannot model the impact pathway and represent our current uncertainty of all the things that we believe to be important for the outcomes, then we have no real basis for knowing what to usefully measure. Thus to tackle the second question above, we need to be (i) measuring variables with high information value for improving the decision and (ii) monitoring whether our forecast is turning out to be correct. The variables that have higher information values and large uncertainty are most likely to go off track during implementation and are thus are highest priority for monitoring.

Tracking high value variables in the forecasting model along the impact pathway can provide cumulative evidence for the impact of the intervention on outcomes and help attribute intervention impacts on the higher-level outcomes. For example if the uncertain intermediate variables (e.g. adoption rates) track close to projected values then there is more confidence that the outcomes are due to the intervention than if adoption rates were lower than forecast.

An example of a framework for forecasting interventions on CGIAR system level outcomes is given in Appendix 5 Applied Information Economics. How a forecast is valued relative to alternatives involves resolving stakeholder priorities and is dealt with below.

Implications for metrics and indicators

The evidence provided provides a strong rationale for putting decisions before measurements, and that it is wise to define metrics or indicators based on a forecast of the impact of alternatives on likely future outcomes and Value of Information Analysis. Thus there would be little value from taking indicator sets from previous monitoring studies and simply applying them to the above questions. There is a strong likelihood that we would choose the wrong measurements without a forecast and Value of Information Analysis.

Furthermore what constitutes a value for money metric for measuring agriculture, ecosystem and poverty and nutritional outcomes can only be defined through probabilistic forecasting and Value of Information Analysis. A metric that does not reduce uncertainty, or where the cost of being wrong is small, will have little value. On the other hand a measurement that provides a very small reduction in uncertainty on a variable where the cost of being wrong is large could be worth much.

The value of a measurement will also change with decision context and may become valueless once uncertainty is reduced to a threshold beyond which further uncertainty reduction does not affect the decision, or does not become economic in terms of yielding greater pay back than the cost of the measurement. Because different people also perceive uncertainty, or degree of belief differently, the value of a measurement or metric will also depend on the user. However there are many re-occurring decisions and types of stakeholder groups that would point to metrics that would have common value. But the analysis to identify them has yet to be done.

Implications for analysis of trade-offs

Trade-off analysis has been given scant attention in monitoring initiatives probably because of a lack of connection to decision analysis. A decision analytic framework clarifies trade-off analysis by distinguishing material trade-offs (e.g. crop productivity at the expense of biodiversity) from preference trade-offs (e.g. how to weight one objective versus another, time and risk preferences).

Material trade-offs and their uncertainties are explicitly represented in probabilistic forecasting models. Preference trade-offs are implicit in any decision and can have a large influence on decision outcomes but are usually not articulated. Therefore representing preference trade-offs of different stakeholders in decision analytic models, including the uncertainty in them, is a key benefit of in decision modelling.

An important benefit of this process is that quantifying the benefits of ecosystem services in terms of provisioning services and impacts on human welfare effectively provides a business case for maintaining ecosystem services. For example the value of biodiversity may be valued in terms of its effect on pollination services expressed as impact on long-term crop yields. Concepts such as ecosystems services and resilience, or soil quality then become hardened or tangible and measurable, even if the degree of uncertainty is high.

Probabilistic decision analysis as a unifying framework

A probabilistic decision analysis framework including Value of Information Analysis has potential to help overcome the deficiencies identified in the gap analysis of current approaches to monitoring, and provide a sound basis for evaluation of investments in agricultural research and monitoring initiatives. How decision analysis could overcome these gaps is summarized below:

- The purpose of the initiative is well defined by the specific decision problem. The process of clarifying the decision problem often provides a large contribution towards improving the actual problem situation.
- The conceptual framework is clearly defined as a forecast of the impact of decision alternatives on outcomes.
- Defining the specific decision problem automatically defines the inference space (e.g. geographic scope or target population).
- Uncertainty and risk are inherently represented. This is key for analysing agro-ecosystem management and livelihood outcomes; as well as for computing value of information.
- Measurements, metrics or indicators that have high value are systematically identified and how much it is worth spending on them is calculated.
- Impact monitoring is defined by the need to validate the model with a focus on the most uncertain variables.
- The objectives of measurements are well defined in terms of reducing uncertainty in key variables. This in turn makes it easier to design appropriate sampling frames, units, etc.
- Uncertainty analysis can be used to identify efficient research designs (e.g. sample sizes; when to stop a trial).
- Biophysical and socioeconomic variables of high value are identified under one framework. Seemingly intangible costs, benefits and risks can be represented even if they have large uncertainty. For example impacts on say women's empowerment may be represented based on current expectations of observable effects – and the analysis would indicate whether further research to narrow the uncertainty would influence the decision on alternative intervention designs.
- Credibility, legitimacy and salience for decision-making are implicit (built in). The initiative is designed for the purpose of improving important decision dilemmas facing stakeholders.
- The cost-effectiveness of the measurements and the decision analysis effort itself can be readily computed.
- The analysis provides a business case for how interventions are expected to impact on outcomes. The business case can help donors or government ministries make wise investments.
- A quantitative forecasting model of the impact chain through to higher level outcomes helps to define which variables can be tracked to accumulate evidence on impacts of interventions and on attribution of outcomes to the intervention.
- Uncertainties over political influences on decisions or other external factors (such as changes in policies) can be represented in the models.
- Communication of results and impact on decisions is potentially enhanced because decisions of stakeholders are being analysed and high value information relevant to improving decisions is provided.

- The decision analytic framework provides an end-to-end learning system:
    - Iterative intervention design to maximize returns and minimize risks through ex ante evaluation of alternative interventions in terms of risks and returns.
    - Design of efficient monitoring schemes. The value for information analysis identifies both which additional measurements would improve the decision and which variables are of high priority to monitor during implementation of interventions. Monitoring divergence of actual versus forecast performance provides opportunity for early adjustments during implementation. Lack of divergence provides evidence towards attribution of impact.
    - Validation and reliable learning. Tracking actual versus forecast variables, especially for highly uncertain variables, sets up the next decision on further action. There is incremental updating of knowledge or beliefs.

Some limitations of the decision analytic approach include:

- The probabilistic decision analysis framework is conceptually quite different from current approaches to monitoring and research prioritisation, and may not be readily absorbed by scientists and government agencies. In particular uncertainty is largely ignored in most decision-making and current modelling efforts and trade-off analyses.
- Decision makers, scientists and donors are not generally well versed in decision analysis methods, and there therefore is a large capacity building agenda.
- Considerable skill is often required in clearly defining a decision problem.
- Those engaged in decision modelling exercises require calibration training to estimate probability distributions with reasonable accuracy. However this can normally be done in half a day.
- There is an added transaction cost in engaging decision analysts, who have the appropriate skills in model formulation and elicitation of probabilities.
- Software tools used in decision analysis may not be familiar to researchers, although simple Excel-spreadsheet tools are now available.
- There are potentially a large number of decisions that would have to be analysed, though there are opportunities for developing generic approaches so that in new situations it is more of a matter of populating the models with site specific data as opposed to changing the model structure.
- Uncertainties and preference trade-offs are different for different stakeholders but can have a large influence the model outcomes, however revealing these differences often improves overall decision-making.
- The way relationships are represented in models (as conditional probability distributions) can have large effects on model outcomes. However revealing these uncertainties can help guide high value measurements.
- As with any research process, the success of the process will rely heavily on the degree of participation of decision makers and stakeholders.