



CGIAR Research Program on Forests, Trees and Agroforestry (FTA)

Special workshop of the Independent Steering Committee (ISC)

Impact Assessment for Natural Resources Management and Policy Research in FTA

Tuesday 5 Nov 2019, Rome

Final Minutes

Participants

ISC Members

Anne Marie Izac (ISC Chair), Susan Braatz, Linda Collette, Richard Stanislaus Muyungi, Florencia Montagnini, Stephan Weise

Management Team

Vincent Gitz (CIFOR, FTA Director), Ramni H. Jamnadass (ICRAF), Fergus Sinclair (ICRAF), Michael Allen Brady (CIFOR), Peter A. Minang (ICRAF), Christopher Martius (CIFOR), Rene Boot (Tropenbos International), Li Yanxia (INBAR), Plinio Sist (CIRAD), Eduardo Somarriba (CATIE), Christopher Kettle (Bioversity)

Management Support Unit

Alexandre Meybeck (CIFOR), Monika Kiczakajlo (CIFOR), Fabio Ricci (CIFOR)

Gender

Marlene Elias (Bioversity)

MELIA

Federica Coccia (CIFOR), Brian Belcher (CIFOR), Karl Hughes (ICRAF), Elisabetta Gotor (Bioversity), Jean-Charles Rouge (CIFOR)

Background and Rationale

The CGIAR has a long history of quantitative ex post impact assessment of its research, especially for crop germplasm improvement and other technological innovations. Such assessments have typically estimated economic surplus by modelling yield gains from new varieties or technologies (hardly any such studies exist for tropical trees). More recently, there has been a strong push, from donors and led by SPIA, for impact evaluation that focuses more on empirical measurement of an impact (or causal effect) through comparison

with a counterfactual (what would have happened without the intervention) (Stevenson, Macours & Gollin, 2018).

However, for many reasons that the workshop will discuss, different approaches are called for, when research aims to influence natural resources management (NRM), policy and institutions, as part of complex development impact pathways, and when research adopts a systems' perspective.

The aim of this workshop is to strengthen a shared understanding within FTA of (i) NRM, institutional and policy research impact pathways and (ii) how we can reliably evaluate effectiveness and impact by addressing the challenges just mentioned. This will provide FTA scientists with research tools and approaches that yield credible and internationally recognized assessments of the effectiveness of FTA's research, as well as greater accountability towards FTA's stakeholders, including its donors.

A) Setting the stage, state of approaches

Chair opening

The ISC requested this workshop because impact assessment (IA) is an essential part of the research agenda. Not to be done after or aside, it should be planned from the very beginning. IA provides the necessary feedback loops that give scientists a tool to see their effectiveness, and it is linked to the Theory of Change. Impact assessment is therefore crucial for the strategy of FTA.

Given the CGIAR current context, being able to answer the question: what is the role of trees in agriculture? – will be fundamental for the future of CRPs (or whatever they will be called in the future). And IA is the way to answer this question. FTA is the research group by excellence who should be able to answer this question – we are not promoting any ideology, but concrete results.

Expected results from today:

- Where is FTA on this IA work?
- Is FTA going to be able to provide concrete data and results – not only qualitative?

Understanding and Evaluating Research for Development – Brian Belcher

First, it's important to try to clarify some concepts. The CGIAR is focusing on quantitative/experimental approaches, but in fact we need also a lot of qualitative ones – for FTA and other CRPs – and they have a lot of value too. There is a big discussion ongoing

on how we do research. Research needs to have more impact. Research needs to have social impacts.

In essence, research needs to be more engaged with clients and stakeholders. Knowledge generation has to include co-generation. To ask the right question we need to engage with the stakeholders. Key idea is that we are co-responsible for our results. The CGIAR reform has put this more under the spotlights. We have to do research IN development.

We have multiple impact pathways. IA is needed for learning. Research is done in different ways – we need to know what works or not. Default thinking of research is generally wrong: citations are not the point; the point is how is research being used to make change. Donors talk about mission level impact – is it making a difference on FSN, on environment, etc. CGIAR uses the idea of “spheres”.

But Innovation... often there is talk about impact of an innovation... but innovation usually goes through multiple pathways. Within the CGIAR, IA is emphasizing counterfactual approaches, and qualitative methods are seen inferior. But the kind of research we do usually doesn't lend itself to counterfactual approaches. We are often in situations where we don't have a replication possible (see policy-related topics).

It is not clear what donors want when they ask for impact-evidence: what is sufficient evidence needed? What type of information/data? Etc. There is a poor understanding of what donors want in terms of IA.

The theory of change evaluation is relatively qualitative. It is detailed. Is the ToC logical? Are there other results as evidence? Other elements?

We need careful evaluation design, rigorous data collection, triangulation, transparent analysis, peer review. But this is not sufficient. We need a portfolio of approaches:

- Nested ToCs adjusted per project
- Experimental and quasi-Experimental impact assessments of large innovations
- Project-scale theory based outcome evaluations
- Aggregating higher level outcomes modelled theoretically
- Ex-ante impact assessments
- And then linking indicators and metrics to results

What are the implications for FTA?

We need to improve application and documentation of social-change theory. Need systematic integrated nested ToCs and evaluation framework. Seek focus and coherence in overall program design.

We need to communicate and negotiate impact assessment expectations with partners, donors and SPIA.

Discussion

Fergus – innovation is a contentious term. In the HLPE report we stress that it is actually a process not a product – so innovation happens when people actually obtain something from the process.

Elisabetta – need a clarification on what means “good” data, as it is linked to impact. How is this handled in FTA – do we have good data? Needed to produce good causality...

Brian – what is the baseline? I.e. In policy: good to have baseline data, but what are you trying to achieve. Clearly for quantitative and replicable experiments it’s clear, but for social environments, the impact pathways are so many and the influencing factors are so diverse that the nature changes completely.

Richard – looking at engagement with stakeholders and clients is very important. Also engagement with donors in order to ensure that impact reflects their mission. In some instances, donors have strategic approaches in specific countries. How do your research relate with the development agenda and with the donor missions which in some instances have specific timeframes?

Brian – it is useful to discuss expectations from donors, often they expect direct consequences etc. but often it is not possible due to different time-frames and multiple causal pathways. It is useful to show the ToC and to link it to indicators, then we show evidence at outcome level to demonstrate that the project has a contribution. So, the causal line is then clear through the ToC: we cannot claim a full attribution, but we can be part of the change. Then this might have linkages with policy for more longer term projects. Is there evidence that the research has contributed to the policy discourse – through interviews, and less quantitative methods. Even the commodity centers are understanding that we can produce an improved variety but we still need to go through the social pathways for its adoption.

ISC Chair – SPIA was supposed to advise on donors’ expectations... what is happening?

Brian – I asked if there was a systematic information on expectations from donors but didn’t get a precise information. I think it works more on who is more vocal. SPIA does not seem to have a clear idea of what donors want.

Karl – SPIA is looking at innovations that have achieved significant impact – which could justify investment in the whole system. For SPIA, studies have to be statistically sound. This kind of pressure on the type of IA may skew the type of research that CGIAR does and lead it to focus only on research for which impact can be measured in particular ways... It is a vicious circle.

ISC Chair – Different donors have different expectations, so it is important to know what they are, more than what SPIA is advising. Importance of nested approaches. SPIA does not have the monopoly in determining what is a reasonable IA. CRPs should be able to stand up and speak for themselves, as long as they are credible.

ISC Chair – I would like to know from the FP leaders what they think, especially on the 6-7 principles at the end of the paper. Are these useful, are they fitting in their work, etc. Should we propose some modifications?

FP 1 – Ramni – it is a useful framework, but our case study we are illustrating today can show also a different approach mixing quantitative and qualitative – maybe Brian can guide us on where to stress which approach.

FP 2 – Fergus – very useful and need to be more explicit on how we do our IA. We have many different ways in which we attempt to have impact. Because we see different reasons why agroforestry is not being adopted at a wider scale. So, then we will need an aggregation of all the different approaches. To have impact at scale you often need to do different things. There is a challenge – how to simplify communication for a wider audience (donors, CGIAR, etc.) and communicate effectively across these results and approaches?

FP 3 – Michael – very helpful, especially the spheres approach and the discussion on the qualitative vs quantitative – in particular if related to SLO levels. We need to marry these two different methods.

FP 4 – Peter – overall it does reflect our way of working and the way outcomes are framed – one is more policy oriented, but there are some bits that are more quantitative in terms of measurement. It is open and flexible enough to accommodate both. Biggest challenge – where you have both paths for impact – how do you put them together?

FP 5 – Christopher – it aligns a lot with what we are doing in FP5. The ToC is a sort of map, very generic and it fits what we are doing, then going on the ground there are a lot of different things and different ways they relate to each other. It is a broad map, then we need the details.

B) State of the matter in FTA

SUCCESS: A case study for MELIA – Federica Coccia

SUCCESS – implemented by ICRAF in Peru – was selected due to its learning potential. It is a small project, but it used a co-generation of knowledge to influence policy in the agroforestry concessions. Methodology is going to be skipped (the ToC was reconstructed retrospectively, so evaluators had a workshop on this). Evaluating the full impact of the project will require more time, but what can already be said is that the project was successful to adapt its impact pathways to influence policy, even if it was relatively small in

budget. Findings are detailed – the most interesting part is that evaluation comes up with recommendations for more research – there is a learning process (see last slide).

Discussion

Ramni – how long did it take to evaluate this project?

Brian – around 9 months in total, but there were a lot of bureaucratic delays.

Ramni – how practical is it if it takes 9 months for such a small project? Is it worth the time and resources spent on evaluation?

Brian – we need to work in advance with the planning. This particular case needed the ToC to be reconstructed in retrospect. The purpose of this evaluation was to learn from this example as it represents a best case of co-generation of knowledge.

Peter – I like the idea of portfolio of ToC to capitalize on our approaches. But, in my opinion, we did not capture in this project all that we did in that area – if we had the multiple portfolio etc. We could have had better results.

Federica – the results (last slide) actually do refer to the multiple pathways.

Fergus – hugely useful report. There is almost an inverse relationship between funding and impact (see CROSS) in CIFOR-ICRAF Peru. Another incredible project where the initial project had very little funding, but due to its great results, it was followed up with a 1.5M project through BMZ. There is a difference between IA for learning and showing impact, and another for evaluation projects. This latter is less interesting to me. This report does not show how massively positive this project was. We are not selling ourselves well!

Elisabetta – I would like to stress that IA, evaluations and ToC are all different things. They require different expertise and serve different purposes – I feel we are mixing up things.

Brian – it is a mistake to think of an evaluation as “Did the project do what it said it would”? the right question is: “Did the project’s theory of change deliver what it theorized?” – these are different questions. There are multiple functions – scientists cannot do the policy and communication. Change goes through this path also. We need to be explicit on the way we expect change to happen and learn from what did not work. When things do not work it is also good – provided you learn from them.

Review of a portfolio of studies – Karl Hughes

FTA’s targets are immensely ambitious. I took stock of all the evaluation pieces – so I developed a framework, a review and some food for thought. Basically, I have broken down the spheres’ framework into a different and more personal framework. 31 evaluation studies, screened for the influence of FTA research or the impact of FTA on informed-

innovations, which delivered 14 reports. I then determined a pseudo-evidence gap map (mapped by FPs).

Main observations are that:

- There are a lot of gaps, areas where evaluation studies have not been conducted yet, especially on FP1
- We can only show a very small fraction of contributions to targets
- Most evaluations are in engagement and influence spheres, most exclude the concluding spheres
- Good, but focus on project-level so at the early phases of influence / initial stages of uptake and use. Need to broaden the scope.
- Some are more rigorous and some aren't – most are internal
- Studies focus at project level,
- Also need to broaden outcome/impact studies (see slide)

Food for thought:

- Need to move to more issue focused Program Evaluation – and look at the big issues FTA is dealing with: are they the right issues? Issues like: climate change, land degradation etc.
- We need to better communicate and sell our research. Harvest Plus did a lot of ex-ante work to demonstrate impact – we can learn a lot from them.
- To what extent are FTA innovations being developed and what is FTA contribution? To what extent are they been implemented so that they can be upscaled? What are the likely benefits? A lot of benefits will need time to emerge – there is a need to project in time. What are the “impact gems”? We should identify the “impact gems” with potential to justify future investments.

Discussion

ISC Chair – can you give some examples?

Karl: for FP5, REDD+ ; for FP2, Options by Context; but I admit I had not anticipated this question...

Richard – on the previous presentation – the 2nd finding: changes on policy as a result of the work – to me this is more evaluation (Eva) rather than IA.

Federica – in fact it was an Eva. Which said the outcomes in understanding where achieved, but none in policy – though it potentially had positive side effects: it contributed to the policy process, but not in policy outcomes, due to the short time-frame duration. It would be good to go back after some years and see the policy outcomes and how they link to the projects.

Stephan – I think we are struggling between IA and Eva. We really need to be clear. What are we trying to do? There are different ways of looking at the targets. FTA went through a process defining its priorities. I would not question our priorities and partners expectations – maybe these could be inserted in the learning process as a feedback?

ISC Chair – in this workshop we are focusing on **Impact** (assessment **and/or** evaluation).

Karl – for me impact (assessment) evaluation is under evaluation. We are not necessarily questioning the FTA priorities, but rather looking at what we are doing in the bigger issues and trying to package it better from an impact and evidence lens.

Li Yanxia – In the table there is a lot of white space to achieve the big targets – so if we take the whole programme, the big picture – are there any pieces missing in the puzzle for the overall evaluation?

Karl – maybe that would be useful for the future of FTA, but at the moment we are not really focusing on highlighting evaluation gaps.

Ramni: now we have the priorities (some are sub-set of others) – I believe it makes sense to arrange our impact/evaluation around the priorities. Those are our big issues.

Plinio – Honestly, when we wrote them, I was never really confident with the targets we chose. How are we going to deal with the (self-posed) targets that in the end we will not meet? What is the strategy? Also: we cannot have full attribution to all of them – for instance poverty alleviation is usually linked to a lot of policies – it cannot be only the result of the work of FTA. There is a danger of having an IA based on quantitative data – it is not all depending on research, many other factors influence the way society changes.

Karl – we need to elaborate on the fact that fully quantitative targets are more aspirational than realistic, because we can only show to be part of the change – we can never claim full attribution. We can provide some projections of the extent to which we can contribute to these numbers.

ISC Chair – we cannot be wedded to these aspiration targets – we were forced into these numbers. Donors know very well how this system works, that CGIAR cannot solve all the issues alone through its research, but that we are part of a wider picture.

Fergus – I disagree – see the GCA report: the German donor is linking the report to the improvement of the livelihoods of some 60 million people. We need to be careful not to under-sell our work. Our impact is big. Those numbers/targets are not coming from nowhere, we did think about them quite a lot. There is a national strategy in Ethiopia – let's be clear on what we will accept as evidence and what not. A lot will be a contribution from FTA; we are not the only player – but we have to show our impact.

ISC Chair – we need to be credible, but never underplay our role. FTA needs to do a credible job of assessing impact (qualitatively and quantitatively). Different donors are interested in different things – so we need to capitalize on the synergies and work on a multiple level

approach. FTA's research needs to continue in one form or another. The CGIAR cannot be just a crop-improvement group of centers, one center per crop.

Lars – it is very useful to have this stock taking exercise, to know the white spaces, it gives us a chance to fill them up. We as FP1 have 3 studies coming up that will be useful in this exercise. Also: donors have remarked that the SMO was careless in putting out a report on impacts of CRPs after 1 year of implementation.

Peter: We need to be clear that aspirations were linked to budget expectations that did not materialize. These aspirations need to be reviewed accordingly.

Stephan: On the one hand we are saying we need to fill the white space, on the other we are talking about identifying big issues (maybe through priorities). These are 2 different levels. We need to do both.

Karl – if we focus on the priority issues the white space would automatically get filled.

Chris – we need to reflect on the role of communication and to work in synergy with the impact and evaluation – few donors will read the actual IA and Eva reports, but they will read the twitter and FB posts...

Vincent – the objective of this session is to recognize that we were only partially equipped with methods that the program would use to measure progress and results at the very beginning. There were no formulas. The original intention was never to achieve the targets, which are nevertheless important for donors and as a driver for the program. At the moment we don't have a clear way to demonstrate the impact. So, we need to be more strategic in our approach towards a workplan of MELIA and of the FPs. Donors want stories and numbers. But also – how do we do it in practice? How do we make a story out of impact? There is an uneven level of understanding in the FPs on this.

C) Main challenges to measure and show progress along the theory of change at FP level

Detailed Examples of Key Impact Pathways – Peter Minang / FP4

*Concentrates on: Agroforestry Policy in India as an example of local governance and policies
Impact Pathway*

Impact of policy implementation:

- Direct investments from the Government: 140 million USD
- (for other elements – see ppt)

There are also indirect / ripple benefits – though they cannot be directly linked to the project, nonetheless they are still relevant.

How to measure impact? Possibilities: interviews, review of literature/report and legislation, historical timelines, analysis of media sources, case studies, social network analysis.

The attribution vs contribution question is important – we had a strong role in identifying the direction – but then it is hard to claim the full implementation vs just influencing effect. It is also a sensitive political issue as governments may resent claims that their policies have been influenced by other actors.

Time is a big question mark – if we started something in Phase II we might not finish it off in 2021. Having some ex-ante analysis and forecasting what it would entail, could it be legitimate evidence?

Counterfactual is another big question: India is the only country which has an agroforestry policy – where is the counterfactual? The absence can be a useful information, but there is no counterfactual. I like the idea of independent references – peer reviews. Are independent references to FTA contributions important to claim objectivity in the process? India has some university studies analyzing this.

In Nepal the process was slightly different, though similar in multiple activities and pathways. Mentions a publication on the process.

Discussion

Christopher – we had a similar story about the carbon accounting system in Indonesia from Australia – CIFOR did the work, but it was published by the Government – they were proud of it and were reluctant to give us attribution. We need to be very careful in this.

Brian – we are using at least 4 different definitions of impact. We need to talk about impact at mission level: people alleviated from poverty – areas of trees saved from deforestation, etc. Distinguish IA and Eva. Need to evaluate change in behavior – assessing outcomes, not outputs. But research contributing to outcomes is not really IA. How do we move then from pieces of evidence that research has changed outcomes – and then build up the rest, estimating impact? What else would we need to do, to move from what we know from project-level impact to mission level impact? For example, what would we need to do to gain an idea of impact in India of the agroforestry policy project?

Karl – This shows the real need for an integrated approach. We would need an assessment of the extent to which this policy was implemented – complemented by adoption studies from the field. This would give an idea of the contribution of that policy. A lot of policy Eva cannot be done on counterfactual premises. Counterfactual thinking can be useful but there is a flaw in the approach: just because something would have happened anyway, it does not mean that research (e.g. ICRAF) did not contribute or have an impact. This example actually could be a gem.

Peter – there is the need to construct the counterfactual argument in a completely different way. In terms of numbers, we could link investments to catalytic effects of the policy itself or even beyond policy. For example, the agroforestry mission came after the policy. Deregulation of species has been important – what effects has it had over timber supply? We have quantitative elements that we could actually use to build an interesting reporting. There is a lot of scope – could also be applied to Nepal, Cote d'Ivoire where we have the concession system. There are many examples to feed the portfolio. Attribution/contribution is a big question. Once ICRAF has made recommendations, the Indian Government took over. Could look at knowledge brought in and convening (in terms of contributions).

Ramni – how will you deal on attribution vs contribution? could we involve the Government and get their blessing on the level of contribution? ICRAF's role could be teased out in the interviews.

Peter – Attribution/contribution is a big question. Once ICRAF has made recommendations, the Indian government took it over. Could look at knowledge brought in and convening (in terms of contributions).

Fergus – this cuts across all FPs; FPs should interact more. In Ethiopia we have a scientist embedded in the national administration. And these people are developing a national agroforestry strategy. He was invited from the Government at an FTA meeting, so it is very explicit and documented. Similar in Uganda, Rwanda. Sometimes it is strategy rather than policy. In Rwanda they are looking at inter-ministerial orders to streamline the strategy. If aspects of FTA are actually embedded in the strategy. e.g. when they embed an option-by-context approach, then the FTA attribution issue is very clear. We need to be clever in how we brand innovations we are producing to be able to make FTA visible.

Marlène – on the implementation of the policy and its impact – we need to be careful on who we are partnering with. Also – we cannot do it alone; India alone has more than 1 billion people.

Elisabetta – It is key to have a network mapping for IA, to understand who are the partners, who the recipients, and the role of actors as it changes along the impact pathways. Scaling is an important element in assessing impact, CGIAR programs are developing innovations – they are adopted by our target populations, the scaling process should start so that we can then see the impact at larger scale after. This takes time. FTA can only go so far.

Peter – for us it is more a technology pathway, which starts locally. The policy pathway is broader. My sense would be that 2 years down the line, look carefully in India, even using national statistics, there could be some proxies that could be used to get to the numbers and data we are looking for. My target would be to look for these proxies.

Susan – the targets are hard to measure, we should instead collect data on the innovative implementation levels and not try to translate it towards the target levels, as it becomes

hard. This from Peter is a very good example on how we can link numbers to broader impact. It is a compelling story even if it may lack the numerical level.

Vincent – it is a great example: emblematic of the long impact pathway of these processes. Two approaches, one at flagship level, another one at a more complex level. How could we work out a process (including also ex ante assessment)?

Chair ISC – other CRPs that produce policy change have been struggling with the same issues (e.g. PIM – where at the beginning the director refused to deal with IA because of the attribution/contribution problem). Do we know what they are doing and how they are doing it? In particular at PIM, because they probably have hard economists and are in direct contact with donors. Similar problems might have already been addressed – it would be good if FTA reaches out to them.

Private Finance Sector Engagement (P17) – Michael Brady / FP3

The ToC for FP3 is still very relevant (e.g. logic model of P17). We need to simplify it and the priority leads need to be coached in this ToC. We are comfortable that these 5 outcomes are logically linked to our work. These are mission-level impacts. FP3 focuses on the private sector – but I want to focus on the difference between commodity value chains and finance for sustainability, on which I will put the emphasis.

We are struggling with some base-line targets and their assumptions. If we take the example of Outcome 3.04 – we have weak baseline information on how many financial services providers (FSPs) were lending their services to the agroforestry sector... and so how will we measure progress at the end? We cannot define the percentage of change without a baseline. Similarly, the baseline of increase of lending of 25% - definition of Small and Medium Enterprise and of Small Holder is unclear so we are doing some work on this, some countries have their own definitions, so we can work on that. How do we track financing transactions? So, this baseline activity is what we are trying to track now – see the eDialogue ongoing in these weeks, as an example.

Discussion

Ramni – I am worried that we are still asking ourselves what was the baseline. Doesn't it concern you? We are still struggling with the baseline, this is frightening...

Michael – yes, we are trying to address this – we will not have a fully acceptable baseline – but we need to qualify it.

Ramni – we need to underline that we did what we could.

Michael – but we will need to address these SLOs, in some way.

Stefan – As we move from outcomes to impact – the blue arrows in the slide of Michael’s ppt – from the outcome to the impact, is this a passive effect or are we actually engaging in this? So maybe we need to put more energy to deliver a lot of impact, in the arrow... This question applies to all FPs.

Brian – the concept of end of project outputs (EoPO) is really important – within spheres of influence. Michael is asking the right questions here. FPs have defined EoPOs, let’s start being specific and try to measure.

Peter – we have strong baselines for some enterprises – and we are starting a similar process in Gambia. For me, when I listen to these discussions, I really think we need to clarify between policy adoption and technology adoption. Policy adoption means huge impact, the process might be complicated but it has an incredible effect. Alternatively, it can be done in smaller groups and set of projects, which then, if they create a critical mass, produce a change.

Karl – we also need to check whether the baseline is still relevant – after all this time and when confronting complex systems, which are inherently not compatible – baselines might not be so critical at this point.

Florencia – Questions such as what constitutes adoption – this has to be pinned down – planting does not mean adoption and does not necessarily lead to outcomes. How do you attribute that to your project?

ISC Chair: I am struck to hear a lot of FP leaders saying that all the ToC pathways are holding up the original assumptions. I would have expected quite the opposite. If there are no feedback loops, for sound science, it is actually very worrying. We need to be able to document changes in the paths of the ToC and in the paths of impact assessment – it’s a learning process. Science is always full of feedback loops.

GCS REDD+ ToC – Christopher Martius / FP5

Knowledge production is a big part of the FP – a lot of publications are broken down in briefs. This is not the only impact story, but the scientific publications provide the legitimacy for other tools which are then used and read, being more accessible. We go to actors in our countries and they have the books on the table, when we come at least...

Different modules have different influencing strategies. Interviews are also encouraged because we go back to the country and show them the actual results of the interviews they have done. Vietnam gave us an award. We also try to influence through journalists. In 2014 we had an external ODI assessment of GCS REDD+

There is a lot of value in these stories – because you can give a narrative. And counterfactuals are not always available. Evolution and Climate Change are two domains

where there is absolutely no counterfactual. So, what can we do? We triangulate! This is still solid; it can become a strong evidence. Sometimes you cannot anticipate the story that will make the biggest impact. We should embrace the idea that we might need to go in hindsight. There is no mechanics or automated IA generation – we need to acknowledge this.

Also: a big question is the permanence of impact along time in a politically volatile period. Quick and deep are not possible together – there are trade-offs. So, there is more work to do on transformational change, to be linked to IA.

The 2014 Evaluation cost approximatively 150/185 k.

Discussion

Karl – we should avoid the statement that the counterfactual is a gold standard – rather we need to differentiate between situations. Approaches vary depending on interventions. We need to aspire for more rigorous qualitative processes. We can have a number of questions for different steps.

Richard – In an FP5 paper there is a link between adaptation and mitigation – may be true at global level but how was it verified at local level? Are we sure stories are better than numbers? Might be true at some stage, but at some point for climate change we need numbers. Stories, if not translated from numbers, cannot enter the statistics so do not have an impact. I agree that there is no magic bullet but the link with facts on the ground is important. Commitment, building trust is ok but direct link with communities would be extremely important. As for ensuring permanence of impact in times of political volatility– another question would be how to ensure permanence of impact in times of climate change.

Christopher – Numbers are very important and FP5 also works on numbers. There is a sort of artificial differentiation between mitigation and adaptation. As climate change is advancing so much, it is becoming clear that without adaptation there will be no mitigation. For mitigation it is clear in terms of numbers so back to the dichotomy numbers vs stories – but for adaptation there is no end target: the end target is resilience – and being more resilient is hard to measure.

Florencia – within FTA adaptation and mitigation have to go together, this linkage has to be made clear and strong. I praise your conclusions on the recommendations, they are very realistic – I think that these could apply to other FPs. I wonder if other FPs experience the same challenges.

Marlène – we need to synergize numbers and stories together and how they reinforce each other; this could be a strong point. We need to advocate for mixed methods and influence

SPIA. Power of both is more than the power of the 2 methods alone. We are showing things moving linearly but we are not giving space to the back and forth and sideway movements in our ToC.

Peter – we could be looking at all the submissions that have gone into the UNFCCC process – it is an interesting way of demonstrating how FTA work gets into the process. We could also use a bibliometric study.

Chris FP5 – when you achieve a lot of policy change in the main tropical forest countries – you will see a huge change. It takes a lot of time. REDD+ is now flying and we should capitalize on it, but it took a lot of time. We should acknowledge that these pathways take a very long time. The big numbers are coming with this more systemic thinking.

PATSPO ex-ante impact study – Lars / FP1

The first results will be ready by end 2019. It is an 8 million \$ project in Ethiopia. The Provision of Adequate Tree Seed Portfolio in Ethiopia (PATSPPO) project is financed for four years (2017-2020) by the Norwegian International Climate and Forest Initiative (NICFI) through the Royal Norwegian Embassy in Ethiopia (RNE). Should contribute to increasing the impact of the FLR in Ethiopia – enable the seed sector in Ethiopia to serve this large restoration effort. The project is composed of 4 components:

1. Governance and Policies
2. Knowledge and provision of knowledge to users
3. Building seed sources for the future (gardens and orchards with a large number of species)
4. Capacity building

The design of the project matched FP1's ToC, but feedback loops are expected, which might alter the overall ToC. But as it stands it will contribute directly to all FP1 outcomes. The PATSPPO project also contributes to the SLOs. The ex-ante study will allow FTA to put some numbers on the baselines.

Conceptual tree breeding Impact Pathways are much more complicated than the traditional CGIAR variety development/adoption. 20 species will be covered for breeding, but we are analyzing many more species. We will probably have a Phase II of the programme, and it is also going to be institutionalized by the local authorities.

Among the arguments for donors were some projections:

- **PATSPO Outcomes**

- > PATSPO will ensure (...) tree planting actors having high quality seed of the most important tree species used for (...) all (...) tree planting activities in Ethiopia (ICRAF, 2017)
- > This involves (...) providing good seed and (...) providing standards, guidance and mechanisms to influence and monitor the use of seed (ICRAF, 2019)
- **Ex-ante study assignment objectives**
 - > To conduct an ex-ante impact assessment of the project, examining alternative project scenarios (...) and the expected impacts associated with each, i.e.,
 - > Mitigation: carbon, and soil and water conservation
 - > Adaptation: (bio)diversification; social and economic returns
- **PATSPO Impact Assessment Methodology steps:**
 1. Defining tree species
 2. Reviewing Tree-Crop-Environment Interactions
 3. Defining Landscape Restoration Options (LROs)
 4. Defining sphere of influence/interest Interventions
 5. Calculating Interventions' Impact / ha
 6. Comparing Tree Genetics Scenarios for each LRO
 7. Scaling Interventions by LRO acreage / % covered
 8. Adding up to estimate potential PATSPO impact

We are focusing on 2 species in this study (*Cordia Africana* and *Grevillea robusta*) – and will add more in the future. The model will generate forestry/agroforestry scenarios. The follow up is starting early next year: add species – of interest for Ethiopia, but also some used extensively by ICRAF in the past, like *Calliandra*. There is hope to put a lot of numbers associated to each of the species.

Discussion

Florencia – you say domestication in the project, but some species have been domestication many years ago. Why aren't you using species that have not really been domesticated?

Lars – we are targeting 150 species – but we will be able to work only on 20 – 90% of those are indigenous. We have included a number that are exotic. *Grevillea* is an excellent example: it is used all over, but in the wrong way – there is a knowledge gap that produces a low yield. We can close this yield gap by applying the knowledge we have. We can aim at having over 50% of improvement. The focus is of course on the indigenous species, but it will not happen overnight. So, we have to work also with exotic species – which are very popular and in high demand, as for the case of *Grevillea*.

Vincent – one of the scope of this workshop is to review existing methods and approaches for IA, with also a focus on ex-ante, because it is a numerical tool. Models have limitations but can be used on the fly to assess the translation into impact as the project develops.

Ramni – PATSPO is a project mainstreamed into development, for seed infrastructure and growth/restoration. Some of the partners and stakeholders want quick development.

Florencia – how did you come up with the 5% added cost of good quality planting material giving a positive impact...? Where are these numbers coming from?

Lars – it is based on past information: we know the costs of past programmes and their results – so we have an idea of the gross cost. Benefits are quantified based on an old analysis (1995). The estimation on the impact for growing trees with better material is also taken from previous data. We want to update this study with newer and more diverse data (these are currently mainly coming from industrial timber).

Peter – how are the assumptions linked to projections? It is important to link them to the ex-ante numbers.

Lars – one assumption is that we will close the yield gap by making more efficient delivery systems – better material and better use of it in the appropriate systems. What we do in the impact assessment is to link the levels to the outputs. On the gain side, it is based on realized gains, not pure fantasy, but it does assume that growers are enabled to be efficient independently.

Showing results using extrapolations – Fergus Sinclair / FP2

Valuing not only direct impact but also indirect and using models for impact over time. We have a big research portfolio – we place the accent on the options-by-context approach – a paradigm shift that we are promoting. We can document 3M people in households that are directly involved in some way in the projects we are deploying. These interact indirectly with around 100M people.

The Global Commission on Adaptation (GCA) report is also an example of how we extend our reach – it will make changes across other projects and organizations. The question is: what changes need to happen for agroforestry to be taken up? A diagram in the report shows how to implement ToC in another more positive way: Create a level playing field, embrace complexity and enable integration. In India in some areas the social movement is

really acting super quickly with the zero-budget farming initiative without any scientific underpinning... but why is it working? We want to learn from that type of development, where there is an active connection between social movements and science.

How do we predict the impact on livelihoods? Using simulation models to look at a whole network of interactions in agroforestry. From these simulations it is clear that agroforestry helps facilitating many interactions. Use Simile (see simulistics.com) modelling environment for exact assumptions and development. Ethiopia example: innovations are unlikely to lift a majority of farmers out of poverty on their own, but they have implications for whole livelihood systems, offering households greater options. Sometimes we are trying to evaluate ourselves as if we were development actors, whereas we are one step back from that. There are different ways of presenting data and information depending whom we are addressing.

Simulations also offer us ways to answer questions – e.g. How many trees does the household need to be self-sufficient for fuelwood? We can make some guesses, but the model based on numbers shows that we need many more trees than what we thought before. So, we need to model what the impact of tree growing would be in different households. We need to do this to 1) find out what will work (what will be transformational), and 2) also to show donors what the impact of the project is likely to be. Trees take years to grow; the project life is shorter; and it may take several years between investments and return therefore we need to quantify impact ahead. We can then multiply. Also model the areas of interventions using diffusion models.

Some results

- need substantial change of land use
- larger farms show greater potential to improve
- intra-household dynamics are important

Discussion

ISC Chair – do you use this approach in your research?

Fergus – yes, we model what the impact of tree growth would be in households to see which are transformational and which are not – and to show to donors what impact we will have. Some trees take 10-15 years to grow – so you can only forecast the real effect (ecological infrastructure) of this, cannot see it during the project which maybe only lasts 2-3 years. There was an external evaluation in Kenya which found that farmers not involved in the project were taking up the innovations because they saw the value from their peers. The

evaluator had decided not to visit the farm picks indicated by the project managers, but went for a random walk – s/he was amazed to see that also farms that were outside the project were taking on board the innovations, realizing the benefits. It should be clarified that agroforestry has to act as a contribution; cannot do everything alone. For stories – we can actually have personal cases with names and people telling their story – this is very compelling. In the folder for this meeting – there is a template for options-by-context approach and how it is adopted.

Plinio – does the model help the implementation of the project?

Fergus – yes. Both the examples are coming from the Dryland restoration IFAD funded project (5M USD) – it connects to an IFAD loan programme in each country (100 million). We use ICT and mobile infrastructure to understand contextual variables and outcomes; you can predict in some areas what will work. You also come across situations where you know nothing about that context; contexts for which none of your research options are viable. We are currently affecting about 3M people directly through this project (with over 5,000 farmers involved – from which data on performance is coming back).

Vincent – one of the reasons why we asked you to look at extrapolations of impact from different FPs is to learn from each other. Can a model be built to look at the FP's traceable indicators? Do you think that, given all you have accumulated, there is a way to use all of this to create a simplistic model for all the flagships so that we can track indicators for annual targets across FTA?

Fergus – we are developing dashboards for each co-located site, which could work. But the annual targets are wrong in my opinion – it's a fallacy to think that the outputs of one year are linked to the investments of the same year – so we use cumulative indicators which are more appropriate, but this brings the issue of double counting. For instance, PATSPO is certainly impacting some of the same people as Trees4People, etc. So, we need to have a broader approach. I hope that today we make a decision on how we cope for these bigger-reach-targets. We will not have details on the amount by which we have changed households. It will be a somewhat loose connection. We can have case studies to show transformation at the local level. FP2 is also bringing in things from other FPs and situating them in an appropriate way. Interactions with other FPs are extremely important. For example, I would like FP3 to take on more of the value chain work rather than it being done in FP2. So far there is a lack of consensus within FTA about what constitutes an appropriate claim. If we are too stringent, we will underplay our real impact, if we are too loose we risk losing credibility.

Florencia – you need to be careful because the FPs are very different. We were required to explain how we were contributing to the impact, but the way this should be done is not very well articulated from the central SRF. In the past there were thoughts of changing the funding schemes based on the impacts, not on the outputs.

Richard – it is not clear to me how innovations are not linked to alleviation of poverty etc.

Fergus – this was just about some specific issues and contexts – trees for firewood – will not generate the same benefits unless you couple it with the development of other activities so that the free time liberated from firewood collection can go into profit-oriented activities – but this requires enterprise investment along the agroforestry implementation. Trees contribute sensibly to transformation, but do not do the whole thing alone.

Stefan – we need rigor around the outcomes, numbers etc., but also on modeling. We need solid pieces of work for all FPs. Where we need more flexibility, when we take it to the next level, is in the realm/sphere of interests – they could be based on modeling, on the policy feedback, etc. For the first piece we need to be clear on what is the gold standard for FTA. Links have to be reasonable but also pragmatic.

Fergus – 1. Rigor – also for modeling – we are using apps and CSIRO crop modeling integrating it with SIMILE – so the crop yields we are making are credible; PIM and IFPRI will recognize these models as valid, etc. If we are not careful people will say that our models are constructed to match our aspirations, so it is good to rely on models which have been developed by others also.

Brian – 3-5 M people directly affected and 100 M indirectly. What are you talking about?

Fergus – it was people reached – if a programme includes an agroforestry programme for 60M households, we can then use that figure for “reached” – but people affected will be much more because households are larger – but unfortunately we cannot enter the intra-households dynamics. There are 3-5M people which we have documented in some way to adopt e.g. tree seedling or new planting method, etc. The 60M is an extension programme, from the German project referred to – there will be some statistics from the German Government, they say that they have a credible link, but I feel it is difficult to make clear-cut links. So now Germany wants us to train the GIZ people to have more impact on the field. There is a change in behavior in an organization – which means that they will be offering beneficiaries something different with respect to the past. And this has an (indirect) impact. This change of behavior is important.

Vincent – The ISPC was ok when it understood we were talking about reach and not adoption. We can assume that once we offer households more diverse options and a wider range of itineraries – it is already a positive impact. Having a wider palette of options can overall improve the beneficiaries. And you also mentioned cases where farmers decided to avoid some options. This also is good. Increasing the number of options would already be a good result (even if that leads farmers to understand that they should not change). We could improve the narrative about the contexts.

Fergus – yes one issue is that having only bright science will produce same errors in different places because the errors are not documented. We should mention limits as well as possibilities. As said, we have brought the Ministry to embed a scientist in their workgroup to strengthen the scientific base of their development programs.

Karl – we did an evaluation on biomass, we found that only 14% of the people targeted because the levels of biomass needed were too high.

Fergus – adoption is a wrong concept here. I totally fight this idea, because we need to see what people really take on board adapting it on their local necessities, not what the project had in mind for them to adopt. There is this question of local needs and context as well as a correct way to calculate the way they benefit.

Brian – All FPs work on co-generation of knowledge, yet this is not captured. We need to work on this.

A case study from FP1 - Ramni

It is based on the ToC of FTA on P19, Orphan tree crops.

Our vision was to get inputs from all the CRPs and platforms – it did not happen. It could be much better. Most of the links are with the gene bank, where Ramni works. Mapping of the outputs to regions, partners and targets. It is work in progress.

What is realistic to do with 27 projects? And this is only for one cluster of activities. What kind of survey can be done? We could run joint surveys with other FPs. In principle the adoption is captured but it is something that needs to be verified.

Uptake and use – we tried to assess output targets in each project. In project documentation (which is not complete – would need to be verified with those who implemented the projects). It is a starting point; it cannot be used at the moment.

Reviewers asked us to show the annual reads and downloads of our outputs which we did. It gave us also a way to show that targets have been achieved and in some cases we have also over performed. So, we can demonstrate linkages with outputs and impacts.

Discussion

Stefan – in terms of outcomes/outputs this is fine, but in terms of change in behavior (you mention cost-effective approaches adopted by people) – is this captured?

Ramni – no, this would be the last step: go back in the field and ask via surveys and through MELIA try to understand the change in practices from people who benefitted from the projects to link to impacts. Outcomes are inferred based on the context and what the outcome statement of the project is. Qualified guesses that need to be verified. Output targets are well beyond – we are confident we will reach them after verification.

Lars – in fact a bit of it is already captured.

Karl – can you elaborate a bit more on this? In the documents, does it say that this Minister is adopting this or that...? Or are you inferring this?

Lars – we are inferring it, but with qualified guesses.

Fergus – How can we try to do things that are not only for internal consumption? What if we publish a paper/article that backs up our statements, so they are not “internal” – so then it is a win-win situation: publication for authors and for FTA? (e.g. Emilie Smith on stakeholder engagement, showing changing behavior on indigenous species). We should be thinking about the most important things to evidence (which could be published).

Ramni – we are putting this forth as a publication, in fact.

Fergus – CATIE has a massive way of increasing the impact of our work, since they have country offices and centers, etc. we should exploit this much more.

D) Assessing progress: going from FP to program level

Assessing progress at programme level – Vincent Gitz

This is the exercise ahead of us. FTA is conceived as a sum of FPs, but at the end they need to add, with some care to avoid the duplications. We also need to illustrate the synergies.

This was already engineered in the beginning – with interdependencies. We should find that the sum of the impacts would be greater than the sum of the impacts of the individual FPs. The priorities reflect a common programmatic objective. The workplans of the priorities document how results will be achieved at the level of the program, rather than at the level of individual FPs.

We need to build common approaches in IA and measurements, and to make the best of programme research but also communication and outreach and how we look at partners. The task is complicated by the importance of policy dialogues and research. We need integrated approaches.

We need stories, numbers and learning loops (including from failures). We need better and more credible stories, better and more solid numbers, better methods and better evidence. Issue of learning loops is paramount – maybe there is no impact but at least we have provided advice on what works and what doesn't.

Discussion

Karl – evidence-based stories on how we tackle the big issues can be very attractive, so stories and numbers should be coupled – this is where the future lies.

ISC Chair – you may or not have numbers – the credibility of a story might not necessarily need them. If there is a good and credible impact claim, it should not be eliminated just because there are no numbers.

Karl – moving from IA towards stories is necessary because we are clustering and moving to a higher level of programmatic assessment.

Florencia – the individual differences of the FPs need to be acknowledged not only try to make a coherent picture in the scale up. We should not lose this wealth of differences. We should also use some excellent ex-ante modelling, because of the time lag we discussed.

Stefan – what will make us look good externally? Should we be weaving stories around priority areas? or staying at a FP level? The program may be ending but priorities areas are where we would want to continue the work. So, we want to show that we are making progress. Being entrenched in the FP model might not be a good strategy.

Plinio – we are only discussing reporting to donors and policy levels – but at a certain point we need to report to society. How can we have a dialogue with society? Look at the news – the debate is framed only around simplistic concepts such as good or bad, etc. Science

should bring a more informed debate, give more depth to this. The world is not getting better. It would be very interesting to show the FTA ToC to the normal people, see what they think. e.g. to show the Ethiopian farmers that our work has affected 60M people. FTA should reflect on communicating not only to the donors, but also to society.

Eduardo – CATIE is a different type of organization – we have the Governments on top of us, and the way it is structured all the impacts can be traced. From the Sentinel Landscapes – a project has immediately started after and is currently running while we speak. If we want to monitor the impact, CATIE could constitute an example.

Rene – how are we going to communicate within the CGIAR centers? The nature of the programme requires a particular approach to IA – but if other CRPs and centers are not facing the same challenges and do not recognize our approach – we have a problem. This should be included in the actions to follow this workshop. We should also get insights from CGIAR Centers and other CRPs who must be facing the same challenges as FTA.

Peter – there is a special issue and series of papers in agricultural systems laying out these approaches – and they do make reference to these methods. So, there is literature supporting these approaches.

Richard – a programme level approach needs to make sure we engage the stakeholders and recipients, ensuring sustainability.

Michael – from Rene’s comment – we have talked about refining ToCs, adjusting pathways, narratives, but we identified that some of our programme targets are completely unrealistic. How do we deal with that?

E) Conclusions and way forward

ISC Chair – concluding remarks:

I am delighted that we had the time to hear everything. From what was sent as background material, the prospect was not very promising. No one can walk out of this room, without realizing the richness and diversity of what is already available in FTA, in terms of IA. This is reassuring. Tomorrow we will talk about the ongoing reform, implications for CPRs and FTA. Certainly, IA will re-emerge. There is a certain urgency in this program being able to demonstrate a number of things regarding the role of trees in agriculture. I have appreciated the intellectual honesty of all presentations

Recommendations from the ISC

1. The ISC recommends that MELIA and all the FP leaders work together to come to a final agreement on the theoretical framework that Brian and Karl have developed – it seems that you are generically in agreement – but there is still a necessity to come to a rigorous agreement – recognizing that different methods need to be used to assess different types of innovations (in a broad sense) – agreeing on key terms on adoption (or better dis-adoption) – what is the temporal scale (different for policy or technological impact) – when does it click into adoption or dis-adoption and reasons why it does. Some concepts like adoption, dis-adoption need to be clarified.

Between now and 2021 FTA should focus on assessing impact, demonstrating that the type of work done by FTA makes a difference. We need to assess impact priority by priority and not by FPs. These are demand driven, problem-focused. FPs are supply-driven, in general. Up to you to make the strategic choices, not all 25 need to be there. All scientist leaders and MELIA need to make some strategic choices, starting from the ensemble of all priorities – ex-ante modeling has an interesting role to play, because work in different priorities is in different points of different pathways. In some cases, using ex-ante methods will be the most credible way to demonstrate that the type of work of FTA will most probably lead to significant changes. Produce credible evidence of impacts or outcomes on the roles of trees in agriculture. All to show that from now to 2021 to have credible assessment of impacts or outcomes (i.e. what they amount to in terms of roles of trees in agriculture) – in recognition of the fact, due to the reform, we are not in a mode of business as usual. Things will change. To protect the programme of work, the substance of what FTA has done – whatever the form it will take in the future. Evidence, under the form of figures/numbers and stories/narrative, increasing the probability that the CG system and the outside world better understand why trees and forests (trees of different densities) are absolutely essential to the world and to an organization working on agricultural systems.

2. We think it's important to communicate with other CRPs producing policy recommendations, mainly: PIM, WLE and CCAFS and others (note: CCAFS, WLE and FTA are the most influencers of international policy). Others are also interesting to approach to learn their impact assessment methods. We suggest a conference next year with all these CRPs to solidify the basis and foundations on which the impact on policy work is assessed, in a way that is not automatically quantitative. We recognize that the targets the CRPs were forced to sign off are not going to be necessarily the benchmark to which the CRPs will be held accountable to – also in the face of all these changes.

3. We want to stress why we are asking you to do all of this: we think that the results that can be produced in a more coordinated way on the impacts and outcomes (not the final impacts) should contribute to reinforce the international policy dialogue other and beyond

producing evidence in the CG system so as to protect the turf of trees in agriculture, producing evidence that should be extremely satisfying to the programme because it will allow you to see the results of the work you have done. Being the CG a learning institution, we want you to keep in mind that one learns more from ones' mistakes than from the successes. Don't keep in a drawer the cases when you have to change your hypothesis in the ToC. We didn't hear one case yesterday where this occurred, and this is worrisome. The CRP was constructed to create partnership, and the ToC to show linkages between partnerships and impacts and this should come out loud and clear from the research within FTA. We look forward to be surprised next time we look at your background documents.