Workshop Report

Franco-British Workshop: Evaluation of research capacity strengthening in Sub-Saharan Africa

Friday 20 March 2015

Contents

Executive Summary

- There are numerous different approaches to evaluating efforts to strengthen research capacity in low and middle income countries. The approach and method used depends on the complexity of the capacity strengthening activity, the scale of the intervention and the rationale for evaluation. This makes it difficult to devise generalisable indicator frameworks or toolkits that are also context specific and can drive genuine change.

- However, key principles enshrined in frameworks such as LSTM's 5-Step Pathway[1] and the ESSENCE "Seven Principles" document[2] have helped to encourage better evaluation. They urge a precise definition of capacity strengthening goals at the start of the programme to help define the parameters of impact and success. Both these frameworks advise the embedding of evaluation into research capacity strengthening programmes from the very start and the use of evaluation as a tool for continual learning.

- Even evaluation of research capacity strengthening efforts at the level of individual researchers, through career tracking for example, presents complex challenges. However this was proposed as an area where French and UK funders can begin to collaborate more closely to share experiences and learning.

- A potential self-assessment tool, learning from the Access to Medicines Index, was proposed as a means of encouraging participatory evaluation by institutions. This would encourage the institution to evaluate its own progress towards addressing key capacity gaps, thus holding funders to account and helping guide future capacity strengthening investments. This approach presents several conceptual and implementation challenges, and participants agreed to collaborate to explore some of these.

- Workshop participants from both the UK and France agreed to work with the ESSENCE on Health Research partnership on the planned update to ESSENCE's research capacity strengthening evaluation framework.

- Although the workshop was confined to UK and French funders and practitioners, participants stressed the need to involve research institutions and agencies in Sub-Saharan Africa as further collaboration develops.

[1] LSTM Capacity Research Unit. "A 5-step Pathway". [Online] Available at: http://www.lstmliverpool.ac.uk/media/325345/cru-comms-link-5-steps-to-strengthen-existing-capacity.pdf
[2] ESSENCE on Health Research. 2014. *Seven principles for strengthening research capacity in low- and middle-income countries.* [Online] Available at: http://www.who.int/tdr/publications/seven-principles/en/

1. Background and aims of the workshop

This workshop was focussed on the theme of evaluation of research capacity strengthening (hereafter referred to as RCS). It was co-organised by the UK Science and Innovation Network (through staff based at the UK Embassy in Paris and the French Embassy in London), IRD, LSTM, Royal Society and UKCDS.

Previous bilateral Franco-British discussions around closer cooperation on research for international development identified RCS as a priority topic. A meeting was organised by the Science and Technology Department of the French Embassy in London in April 2014. This identified two sub-topics within RCS for further exploration: improved communication between UK and French RCS organisations, and the evaluation of RCS, which became the focus of the March 2015 workshop.

The workshop was based upon a shared understanding of RCS as a complex, messy endeavour which is subject to extensive monitoring and reporting but less rigorous evaluation. RCS involves stakeholders from across the science, international development and policy spheres, who employ a range of approaches and activities in a variety of different LMIC contexts. Thus evaluations tend to be tailored to the many different kinds of RCS activity taking place (from training individuals to setting up new research funding bodies) and to the motivating rationale for evaluation (e.g. accountability to the funder or advocacy for a certain RCS approach).

As such, there is no formal, harmonised system or framework for evaluating the success of RCS initiatives beyond intuitive recognition of whether an RCS initiative has worked or not. While there is a significant literature on RCS evaluation, and various tools and indicators, few of these have been consistently adopted by RCS funders or practitioners, or LMIC institutions.

The workshop therefore aimed to:

- Explore the feasibility of a more standardised approach to RCS evaluation given the numerous constraints on evaluators and the particularly context-specific nature of RCS activities.
- Consider the potential for a generalisable framework or toolkit to harmonise approaches to evaluation of RCS.
- Promote lesson sharing and improved collaboration on RCS between key French and UK RCS funders and practitioners.

2. Insights from the workshop

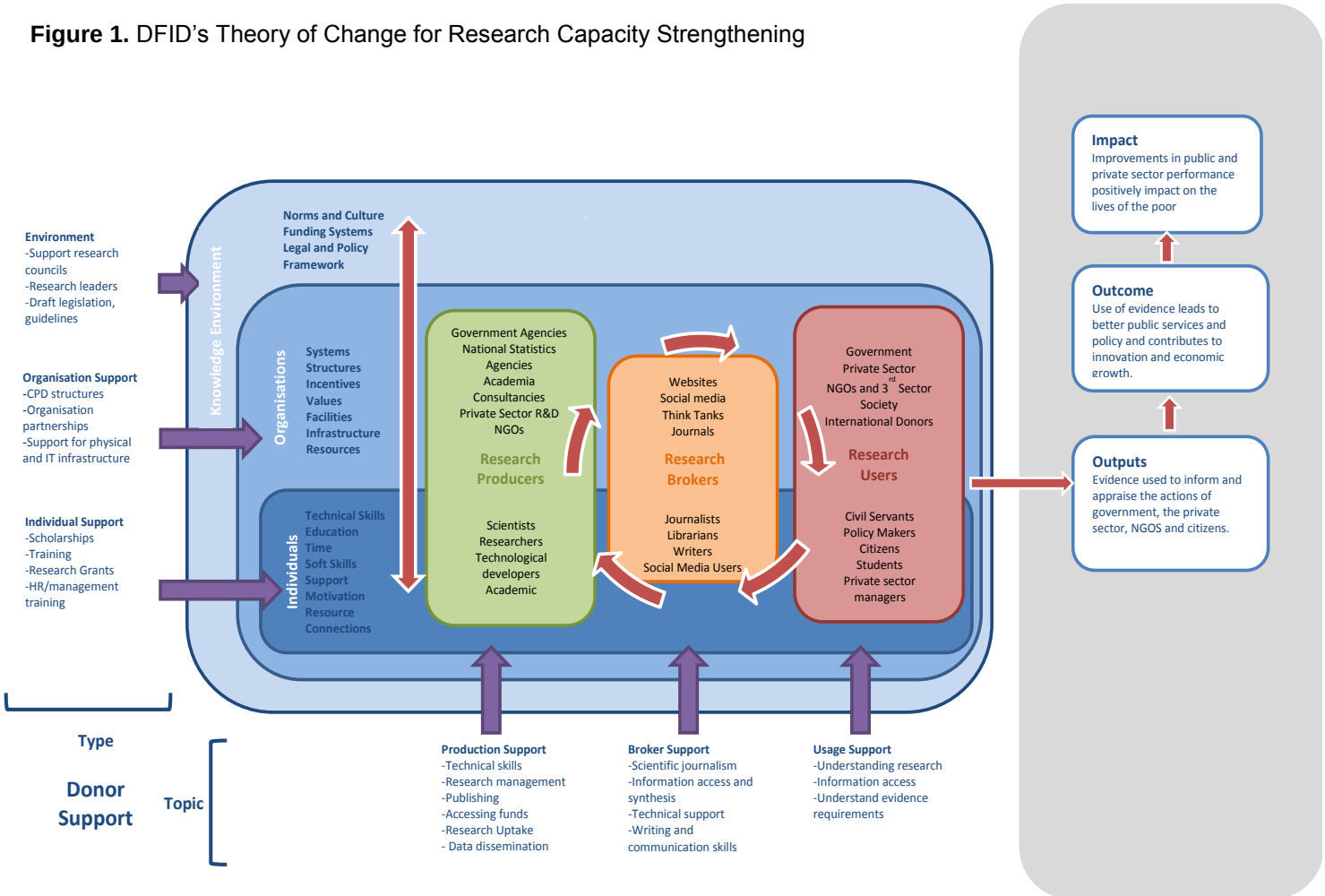
2.1. Introduction

This brief introduction aimed to define some core concepts in RCS evaluation and establish common understanding of the key terminology and ideas.

RCS was defined as: “enhancing the ability and resources of **individuals, institutions** and/or **systems** to **undertake, communicate** and/or **use** high quality research efficiently, effectively and sustainably”. RCS can therefore occur at three levels of the research and innovation system (individual, institutional or systems level) and at three stages along the ‘research into use’ chain (research supply, research communicating/brokering and demand for/use of research).

This multiplicity of approaches is usefully summarised in a Theory of Change for RCS produced by DFID (see Figure 1). From left to right, we see a conceptualisation of: the process of research production; brokering to encourage uptake of this research; and finally use of research. Going from the bottom of the diagram to the top, we see the type of research support broadening out, going from support of individual researchers, to whole institutions and finally to the overall national or regional research environment or ‘system’.

Figure 1. DFID’s Theory of Change for Research Capacity Strengthening



Evaluation was defined as “the independent and systematic collection, analysis, and interpretation of data”, which moves beyond reporting and monitoring to investigate causality, and capture longer-term (and often unforeseen) consequences and externalities. The nature of the evaluation is dependent upon the motivating rationale for evaluation. Using a typology developed by RAND Europe[3], the main purpose of evaluation (in the context of research) can be conceived as one of:

- **Advocacy**: to demonstrate the benefits of supporting research; enhance understanding of research and its processes among policymakers and the public; and to make the case for policy and practice change.
- **Accountability**: to show that money has been used efficiently and effectively, and hold researchers to account.
- **Analysis**: to understand how and why research is effective and how it can be better supported, feeding into research strategy and decision-making by providing a stronger evidence base.
- **Allocation**: to determine where best to allocate funds in the future, making the best use possible of a limited funding pot.

Whatever the purpose or rationale of the evaluation, there are certain challenges common to all RCS evaluations. For example:

- The impacts of RCS initiatives can be diffuse. For example at the systems level, how do you attribute incremental improvements in a country’s publication output to a particular RCS activity? And even at the less complex individual level, how do you necessarily demonstrate that a certain number of PhD trainees has had a transformative impact on the research culture?
- Evaluating change takes time, a fact which can be difficult to reconcile with short-term reporting demands and programme cycles.
- Robust, good quality data may be lacking. For example it can be particularly difficult to track the careers of trainees and showcase how support from a funder at one stage of their career can be linked to their current status. The effort, cost and time required to source necessary data may not be feasible with the evaluation resources available.
- Evaluation capacity in LMICs may be limited. External evaluators may provide greater rigour but equally may lack profound understanding of the local context and miss certain nuances.

Participants were invited to see some of these problems as inherent, and move forward to share what they did in response to the challenge to undertake ‘good enough’ evaluation.

Some key questions were posed for the day including how well previous indicator frameworks have worked[4], and whether different kinds of evaluation are needed when RCS is integrated into research, in contrast to RCS as a standalone intervention. Participants were encouraged to be realistic about evaluation, and consider options for improved RCS evaluation that are practical, flexible and proportionate.

[3] Guthrie, S., Wamae, W., Diepeveen, S., Wooding, S. and Grant, J. 2013. *Developing a Research Evaluation Framework.* [Online] Available at: http://www.rand.org/pubs/research_briefs/RB9716.html Page 1

[4] For example: ESSENCE on Health Research. 2011. *Planning, Monitoring and Evaluation Framework for Capacity Strengthening in Health Research.* [Online] Available at: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@sf_central_grants_admin/documents/web_document/wtp057117.pdf

2.2. Evaluation Case Study 1: The Malaria Capacity Development Consortium

Imelda Bates from the LSTM Capacity Research Unit (CRU) discussed experiences of planning, tracking and evaluating RCS programmes. (LSTM in this role is situated between funders and those on the ground who are implementing and undertaking the research.) RCS presents a challenge for funders, especially when capacity strengthening is embedded as a small element in larger research programmes. While reporting on outputs such as the number of PhD and MSc fellowships is relatively simple, it is difficult for funders to know if a programme is autonomous, sustainable and durable in the long-term.

The CRU involves multidisciplinary teams who take a rigorous, systematic approach to researching and designing RCS. Learning has to be built in from the very start of the programme. The team has developed a five step pathway for designing RCS programmes in health, but also applicable to other research fields:

1. Define the goal of the project and the pathway for change.
2. Use literature and evidence available to define the optimal capacity needed to achieve the goal, as a benchmark.
3. Determine existing research capacity and identify gaps in relation to the optimal capacity defined in stage 2.
4. Devise and implement an action plan to fill the gaps.
5. “Learn by doing” and revise the plan and indicators regularly. Indicators will change, and ideally become more sophisticated as capacity is strengthened.

The five step pathway was illustrated through two examples:

- The first was the LSTM CRU’s involvement in the design, monitoring and evaluation of a Malaria Capacity Development Consortium (MCDC) programme which established doctoral programmes in five African universities (in Ghana, Malawi, Senegal, Tanzania and Uganda). 18 researchers were supported by MCDC, and provided with joint supervision by local and EU-based supervisors. The implicit assumption was that RCS at the individual level would strengthen their host institutions by proxy. The CRU’s role was to identify capacity gaps in universities’ doctoral programmes at the start of the programme in 2009, and report on progress in filling the gaps in 2012 and 2014.
- The second was LSTM’s recent review of four universities’ research support systems (in Ghana, Malawi, Senegal and Tanzania) to aid with the MCDC’s implementation of their institutional support work through identification of strengths and weaknesses.

The way that activities within these examples match up with the five step pathway is shown in Figure 2 overleaf (p7).

The five step “active management” programme has had great success, allowing institutions to create their own wish list for funders which helps identify future areas for joint funding (for example further support for PhD supervisor training was provided by SIDA and THRIVE). Participation by universities in the process encouraged genuine change and empowered teams in the university. The 5 step approach is also transferrable to non-research contexts, being applied for example to the National Blood Service in Zimbabwe. It is also being employed in management of the Royal Society-DFID Africa Capacity Building Initiative.

Both examples illustrated the key point that the goal of RCS programmes is frequently left quite vague or taken for granted. There is no standard definition of “optimal” research capacity, and so stage (2) requires a lot of time but facilitates the rest of the pathway. Identifying current capacity under stage (3) required buy-in from senior administrators, and often site visits of around three days. The action plan under (4) requires input from across the university, not just the department of focus. A learning orientation is crucial throughout.

Figure 2. Illustrations of the LSTM CRU 5-Step Pathway

5-Step Pathway phase	**Example 1:** Doctoral programme in 5 African universities	**Example 2:** Research support systems review in 4 African universities
Define the goal	CRU helped the implementing team define a precise, explicit programme goal: “Universities are able to provide an international quality PhD programme”.	CRU helped the implementing team define a precise, explicit programme goal: “MCDC partner African universities are recognised as internationally excellent research centres”.
Define “optimal” research capacity	CRU conducted an evidence review of the necessary components for a successful PhD programme using qualitative research methodologies as well as observation of facilities. This helped define the core components, classified into eight overarching categories.	A literature review helped defined the benchmark “optimal” capacity and synthesised all university research system components into 8 categories.
Define gaps in existing capacity compared to “optimal” capacity	A review of existing capacity was conducted through interviews with 83 individuals (including deans, supervisors, librarians, technicians and students). This process produced a confidential report for the institutions, as well as an overview report highlighting common issues. Some of these included: a lack of up-to-date handbooks and guidelines for students; lack of provision for group critical thinking; problems with degree documentation, with occasionally no formal degree completion date; supervisors being too junior (lacking experience) or too senior (lacking time); student assessments being driven by Northern partners’ regulations; limited student representation and welfare especially for those with disabilities.	Data covering the criteria defined under stage 2 was collected through interviews conducted during 3-day site visits. This brought out examples of good practice and innovative thinking. It also provided an avenue for institutions to flag up difficulties and risks in the system such as financial management; informal recruitment practices; the need for support when integrating multidisciplinary inputs into proposals; and the lack of sign-off or institutional oversight for project contracts.
Actions to fill capacity gaps	Universities defined their own action plans. The CRU team initially conducted monthly or quarterly visit to provide guidance for universities and resolve preliminary difficulties.	Written reports will be produced for each university detailing the findings and identifying strengths and weaknesses.
“Learning by doing” and revising the plan	CRU conducted follow up visits with the 5 universities in 2012, and 4 in 2014 (as one university had a new evaluation programme). This guided improvements to the PhD programmes and helped galvanise change.	Plans for regular contact with PIs, with progress documented against - and revisions made to - the action plan.

In the discussion following the presentation, the expense of the interventions required to address capacity gaps (identified through the 5 step pathway approach) was considered. Roughly two-thirds of the needs gaps identified in both examples can be achieved internally without extra resources. The process is however time-intensive, and requires buy-in from the focus institutions. If the institutions are to take on the recommendations and integrate them across the university, this involves liaison with senior representatives of the institution to effect change on the ground. The extent to which universities should be involved in the needs assessment under stages (2) and (3) depends on whether the programme is focussed on capacity strengthening alone; if the initiative is research with subsidiary capacity strengthening elements, the 5 stage pathway may not be proportionate. However the general principles of the approach, including incorporating evaluation into project management from the very beginning and explicit definition of goals, can be useful for a variety of programmes.

2.3. Evaluation Case Study 2: Assessing capacity building outputs of the JEAI programme

Christian Devaux from IRD presented on IRD's RCS strategy, which aligns with its mission to "work in the South, with the South and for the South". Around 150 doctoral students are trained each year by IRD as a whole, while the JEAI programme specifically trains about 25 young postdoctoral scientists each year through a three year programme of support.

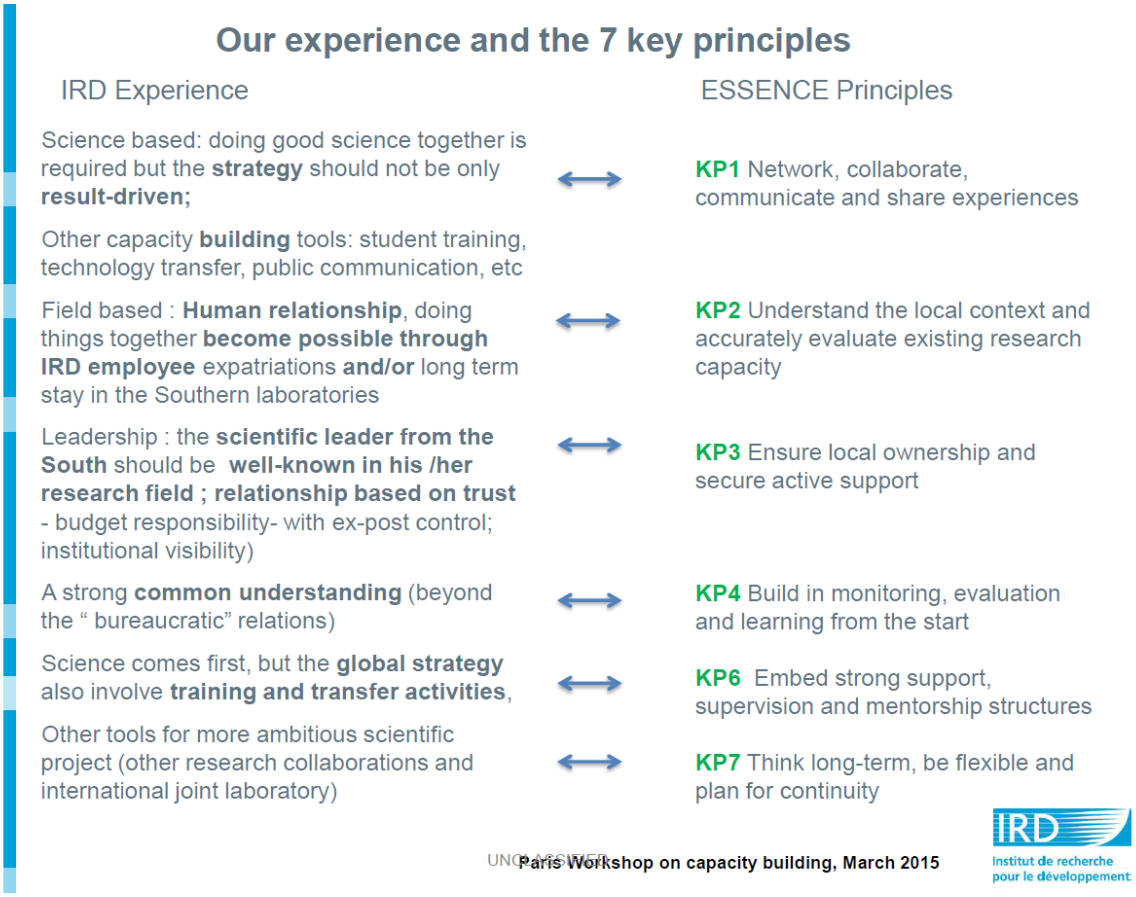
JEAI is based on the principle that RCS is crucial for research; it is difficult to build strong scientific partnerships with LMICs that can result in high-quality joint publications when there are few well trained researchers, limited dedicated infrastructures, poor infrastructure (e.g. laboratories, ICT) and little funding. The programme aims to create a critical mass of scientists in LMICs, and can also provide funding for the team to establish start-ups.

Each team under JEAI is based in an LMIC, with at least three full-time researchers per team and linked to a local IRD research unit which helps provide a favourable institutional environment for research. €50,000 is provided in financial support over three years, and short training courses (in publishing, project writing etc.) are provided, as well as support for research communication. A potential indicator of success can be if a JEAI researcher wins a prize or establishes successful joint projects.

An independent expert committee (from both high income countries and LMICs) of 16 qualified specialists is assembled to select proposals. For evaluation purposes, there are yearly meetings for teams to report their progress and discuss steps forward. Teams are assessed on the basis of the evolution of the scientific project according to publications and symposia; feedback from the field; and reports from researchers, units, heads of department and capacity building project managers.

The evaluation is science based, but IRD's approach goes beyond assessment of results alone. They consider, for example, the added value of the partnership between the LMIC team and IRD, and evaluate the degree of balance and equity in the partnership. The evaluation attempts to understand how the JEAI project is supporting the scientific community and integrating the scientists into local networks and civil society.

Figure 3. Coherence between IRD's evaluation principles for JEAI and the ESSENCE "Seven Principles"



IRD's evaluation criteria are aligned with the principles outlined in the ESSENCE network's best practice document on RCS[5] (Figure 3 above). IRD's programmes place a strong emphasis on networking and exchange, which links into ESSENCE principle 1. Furthermore, Principle 3 in the ESSENCE document is "Ensure local ownership and active support"; the corresponding criterion in IRD's RCS framework concerns leadership, with the stipulation that "the scientific leader from the South should be well-known in his or her research field", with relationships between the LMIC and IRD scientists based on a strong foundation of trust. Part of this trust comes from placing responsibility for financial management and administration of the programme in the LMIC university.

Overall, IRD's emphasis is on the science; however the global strategy also involves training and knowledge transfer activities that can support the scientist in the long term, for example by providing skills to write effective grant applications or to critique the feasibility of projects.

Two examples of this approach, helping individuals in LMICs to undertake quality science in long-term partnerships, are the JEAI Kenyan Wetlands Biodiversity (Kenweb)[6] and Volcanology in Peru (VIP)[7] projects. They both illustrate IRD's belief in adopting a pragmatic outlook to capacity strengthening, based on some key transferrable principles around quality science and the building of equitable relationships.

[5] ESSENCE on Health Research. 2014. *Seven principles for strengthening research capacity in low- and middle-income countries.* [Online] Available at: http://www.who.int/tdr/publications/seven-principles/en/
[6] http://www.ird.fr/layout/set/print/climat/recherches-sur-le-changement-climatique/renforcement-des-capacites/jeai/jeai-kenweb-kenya-2011-2014
[7] http://www.perou.ird.fr/nos-activites/programmes-de-recherche/sciences-de-la-terre/jeai-vip-equipe-de-volcanologie-de-l-ingemmet-perou

2.4. Plenary discussion: The definition of impact, and the challenge of measuring it

In this session, workshop delegates provided their own experiences of measuring impact.

Agreeing a common definition of impact was acknowledged to be impossible given all the possible manifestations of impact. Aid agencies and science funders have differing perceptions of what constitute positive outcomes and impacts, and a range of impact pathways exist. As an example, WHO-TDR suggested that if a grantee becomes a policymaker that is just as much impact as if they remained in research. The impact can thus only be determined when the precise goal of the initial intervention is made explicit. For example, if retaining trained scientists is the goal of a programme then a PhD trainee becoming a minister would not be captured as an impact even if it clearly is an impact of sorts. Honest understanding of the goal of the *evaluation* (e.g. accountability, or organisational learning) is also crucial in helping funders define a clear impact pathway.

Career tracking was discussed as one method of evaluating the impact of RCS schemes focussed at the individual researcher level. Some examples of activity in this area:

- WHO-TDR are embarking on a new programme with the European Science Foundation to assess the contribution of their grants to scientists' careers five years on. In a pilot they worked with 5 institutions to survey about 300 MSc and PhD trainees, to obtain qualitative data on where they are now working and how the TDR grant transformed their careers. One challenge is that email addresses have often changed which complicates attempts to contact former grantees.
- MCDC are also attempting to track careers on a small scale. They highlighted the importance of strengthening institutions and the opportunities institutional RCS presents to help retain individual researchers.
- GSK also highlighted the importance of retention, and suggested there may be learning from GSK's tracking of healthcare workers which seeks to analyse whether the latter remain in the communities where they were trained.
- Fondation Mérieux has conducted a survey of 400 alumni from its Advanced Vaccinology Course for decision makers and researchers. Those admitted to the course are often already highly skilled, so it can be hard to trace causality and attribute participants' career success to the course.
- CNRS are trying to fight against brain drain in their project in Zimbabwe, and are considering how they can best support those already employed in institutions.
- INASP suggested longitudinal studies could help to shape better markers of impact.
- Wellcome Trust has data on UK fellowships which are not always granular but provide insight into macro trends, for example about gender disparities in science. For their African Institutions Initiative, there was no direct career tracking, with the Trust relying instead on indirect information. Information on careers could be improved through close partnership with the LMIC consortia from the start of the programme.

RCS activities at the systems level were also a focus of discussion.

- INASP related their experience of looking across the whole research ecosystem, and classifying the different constituent groups (e.g. donors, coordinators, communicators and users). Mapping the research system and the interlinkages helps them decide where evaluation efforts can best be applied.
- CIRAD is developing a new task force to move towards evaluation at the systems level. When they previously evaluated at the individual level, they were able to report adequately on outputs but capturing outcomes and impacts was much more complex. Understanding the system improves the capacity to innovate and solve key bottlenecks.

- Fondation Mérieux agreed that impact is often indirect, involving externalities that are often beyond the scope of project teams to assess. Outputs are easily measurable, and thus may be more honest and robust than outcomes and impacts, the measuring of which is far more challenging.
- Royal Society highlighted the lack of knowledge about how research ecosystems function and a need to acknowledge the messy, iterative nature of capacity strengthening. Funding RCS activities can be seen as a means of engaging with partner universities in LMICs rather than purely an end in itself.

While funders establish their own definitions of impact, the discussion considered how LMIC institutions themselves define and seek to measure impact. INASP often see their partners defining impact in terms of increasing finances, and attracting new students and grants, which may provide a proxy for increasing research strength. The LSTM CRU method, by defining a clear route for filling institutional capacity gaps, has helped individuals within their partner institutions to perceive longer term routes to change and impact.

The session also highlighted the array of tools and methods that do already exist to evaluate impact that are generalizable to many interventions, depending on the purpose of the evaluation. Participants highlighted the SIAMPI (Social Impact Assessment Methods through Productive Interactions) methods[8] and the Lives Saved Tool (LiST) which GSK use in their Save the Children partnership[9].

2.5. Break-out discussions: Reconciling conflicting demands for evaluations – contextual, motivational and environmental factors

Participants separated into two groups to discuss the difficulties of doing evaluation well in light of the conflicting demands on evaluators, the various rationales for evaluating and subjectivities around impact. There was a fair degree of consensus between both groups, and highlights of their discussions are summarised below:

- **Complexity of evaluation motives:** Evaluation is a complex area, and donors, recipients and stakeholders have different needs and objectives. These also depend on whether RCS is the main focus or a supplementary element of programming.
    - Many of the French funders embed RCS and do not have explicit standalone RCS schemes. A study may therefore be useful to assess the needs and define the objectives of specific RCS activities at the start of the programme.
    - There are also different requirements depending on whether RCS projects are funded by aid agencies or scientific research funders. Aid agencies or foreign ministries have more flexibility to provide institutional support than scientific funders where the focus is more on excellent research.
    - There is rarely one key overarching motive for evaluation – learning and accountability are two of the most significant. Nevertheless donors almost always want to know how to improve future projects as well as understand how effectively funded organisations have carried out the project.
    - Multiple parallel evaluations can be an option where stakeholders have different requirements. For example, LSTM's project on strengthening research capacity in blood transfusion services uses both the EU's reporting/evaluation templates and the LSTM CRU's five step pathway (described on p6-7).

[8] http://www.siampi.eu/Content/SIAMPI_Final%20report.pdf
[9] http://www.jhsph.edu/research/centers-and-institutes/institute-for-international-programs/current-projects/lives-saved-tool/

- **Value for money:** there is no standardised means of monetising the return on investment in RCS which can often influence funding decisions. For some projects, patents can be a useful indicator but generally it is difficult to translate RCS investments into a measure of tangible economic impact.

- **Who gets evaluated?** The question of who to evaluate, in terms of individuals or the institutions that host them, can be difficult to disentangle. Individuals can only be supported in the context of their institution, yet institutional improvements that come through capacity strengthening of individuals can be hard to track beyond anecdotal evidence. For scientific funders like CNRS, the focus will remain on scientific outputs rather than institutional strengthening for its own sake. However LSTM related that even where the focus is on individual RCS with secondary institutional aims, designing a clear monitoring and evaluation plan up front can help leverage additional impacts. For example if a programme is training 20 PhD students in an institution, a defined plan for using the cohort to drive institutional improvements in PhD programmes in general, with goals and indicators, can significantly enhance the positive impact of these trainees on the institution.

- **Using existing frameworks for evaluation:** In terms of operationalising existing frameworks, there is no off-the-shelf toolbox. Instead evaluators have to take aspects of a framework and adapt them. Measures and indicators can be specific to the nature of the project or for RCS. For example, CNRS' platform in Zimbabwe evaluates the science first and employs metrics for other kinds of success later. However funders may want to have a set of basic common indicators (along the lines of OECD DAC criteria for evaluating development effectiveness[10]) which provide flexibility for unique attributes of a programme. Indicators can then be adapted to the LMIC partners' needs.

- **Evidence on successful RCS:** While lots of toolkits are available, research on research capacity strengthening is an evolving field and one that requires robust evidence of what works (and does not). The problems with existing frameworks are acknowledged by funders and practitioners, but perhaps need to be better formalised in terms of what the "known problems" are. This is especially the case in RCS where the effectiveness of the impact pathway is not self-evident, and where RCS remains a more risky and complex venture than many other interventions.

- **Methods:** Quantitative and qualitative measures both have a role to play, but qualitative research in particular demands buy-in from the institution being evaluated. Securing this buy-in becomes more challenging in instances where partnerships between Northern and LMIC organisations are asymmetrical.
    - It is important to differentiate between regular monitoring as part of the programme through quantitative feedback and in-depth evaluation using qualitative indicators.
    - The group debated the merits of on-going evaluation against more thorough end of project evaluation. Indicators that teams themselves help establish and which are tweaked throughout the project may be less costly.

- **Auditing:** Audit, and helping organisations establish standard operating procedures (SOPs), is important to ensure progress can be checked against a baseline assessment. Internal audit may provide more depth, although external audit with a participatory methodology can be powerful.

[10] http://www.oecd.org/development/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm

- **Evaluation as a vehicle for change:** It is crucial to consider how evaluation will catalyse real change. A high degree of trust is required to ensure that there is more than merely superficial engagement from stakeholders. Furthermore, if the evaluation does not provide the anticipated positive results, people will frequently critique the methodology of the evaluation and may claim it is not strong enough to encourage a departure from business-as-usual activity.

2.6. Break-out discussions: Reconciling conflicting demands for evaluations – what would the key principles/indicators of an evaluation toolkit look like?

In contrast to the session above where there was a large degree of commonality between the two groups, this session was more exploratory and the two groups came up with different ideas. These are summarised below:

Group 1

Group 1 considered drawing on the Access to Medicines Index[11] (AMI) as a potential model for institutions who are the target of RCS initiatives to assess themselves, thus creating a ranking for how well institutions' capacity gaps are being filled.

The concept was debated and further refined down, and the idea of a competitive ranking was quickly placed to one side. Delegates moved on to consider instead how the AMI methodology, which relies on quality assured self-assessment by pharmaceutical companies, could be extended to RCS. The potential RCS index could build on the ESSENCE evaluation framework[12] and use ESSENCE networks to coordinate buy-in into the self-assessment concept. This could help established a participatory, peer-reviewed self-assessment process which could help enhance evaluation, auditing and reporting capacity in LMIC institutions.

Rather than creating a competitive ranking or "league table" which might affect funding decisions, it was proposed that the self-assessment tool would function as a kind of "maturity" index, showing progress in certain areas over time. These "areas" could build on Imelda Bates' work, and use 8 categories with 4 indicators for each. The tool would not be a finished product, but encourage continuous learning. It would help gather evidence on how institutions supported by donors are evolving, and would not necessarily be particularly resource intensive.

While this would not replace reports to funders, it would help rationalise the various reporting processes and provide a more holistic picture of how donor support is helping improve the institution. Recipients would therefore be indirectly assessing the utility of their various combined funding sources. Some, as yet undefined, system of external validation would provide quality assurance. Participants noted some of the conceptual and implementation challenges, not least that this is an idea being mooted by Northern donors rather than requested by LMIC institutions themselves.

An immediate step could be to run a pilot study of the self-assessment prototype, for example focussing on one institution in one of the eight chosen categories, in order to translate the idea from theory to practice and explore certain fundamental issues. In order to fund this, the group suggested writing a concept note to present the ideas more clearly to funders (see Section 2.7, p15).

[11] http://www.accesstomedicineindex.org/what-index In short, the AMI is an industry funded tool which ranks pharmaceutical companies in their efforts to improve access to medicines for people in LMICs.
[12] The ESSENCE evaluation framework in this document (http://www.wellcome.ac.uk/stellent/groups/corporatesite/@sf_central_grants_admin/documents/web_document/wtp057117.pdf) is being refined over the coming months in the light of new research, from LSTM among others.

Group 2

This group's discussion focussed on the ultimate aims of evaluation, considering how funders, implementers and users can capitalise on and benefit from evaluation. The question that framed their session was: How can one approach integrate these different needs?

The group highlighted that changes in the context of funding, particularly with respect to policy, can lead to significant shifts in the modalities of RCS and therefore the kind of evaluation you need. The example given (from outside the RCS sphere) was of environmental issues being overshadowed by economic ones in the transition from the European Commission's FP7 programme to Horizon 2020.

Rankings were also discussed by this group, in terms of how they may often only help to evaluate one dimension of success, which may not always be applicable across the board. For example, the Shanghai ranking of universities[13] is weighted towards publications which can shift productivity and incentives in unintended ways. The group underlined that establishment of core principles should precede metrics, with the incremental process of formulating these metrics having its own value by drawing together different evaluation approaches and methodologies. Terminology and language barriers were also highlighted as issues that can complicate RCS evaluation, as the language around 'outcomes' and 'impacts' can be highly nuanced.

The idea of emulating some kind of IPCC-style mechanism, which is effectively a large-scale repository of all available scientific evidence, could be considered for RCS. This would require open sharing of experiences in RCS on the merits of different processes, and sharing of "negative results", when evaluations do not return positive views of programmes.

Another option suggested was that metrics could be conceived for each of the 9 boxes in the three by three matrix below that builds on DFID's theory of change for RCS (Figure 4).

Figure 4. The DFID Theory of Change diagram (from Figure 1) simplified into a 3x3 matrix, with select examples

Topic / **Level**	Research Production	Research Brokering	Research Use
Environmental/ System/National	ANRS/CIRAD/Fondation Mérieux/INSERM/IRD/Pasteur Asia Regional Research Platform (PRR-Asie)		Australia DFAT's Knowledge Sector Indonesia programme
Institutional	LSTM's Research support systems review in 4 African universities		DFID's Building Capacity to Use Research Evidence programme
Individual	IRD's JEAI programme	DFID/INASP's Strengthening Research Knowledge Systems programme	

[13] http://www.shanghairanking.com/ARWU-Methodology-2013.html

2.7. Feedback on the discussions and next steps

A number of delegates highlighted their key takeaway points from the event. This included, for example:

- The need to make a better case for evaluation of long-term impacts in RCS.
- The importance of breaking down barriers between organisations and improving communications on RCS and evaluation.
- GSK, embarking on their new Africa 2020 capacity strengthening programme, stated that they will embed evaluation into their initiative from the start.

A number of concrete next steps were agreed upon as follows:

- **Career tracking.** A number of organisations agreed to form a working group to collaborate on career tracking. This could involve sharing methodologies and data from ongoing bibliometric analyses, and considering the role of social media and other online technologies in mapping career evolution of trainees.
    - The core group would involve Hervé Fritz (CNRS), Beatrice Halpaap (WHO-TDR), Hazel McCullough (LSHTM), Christophe Longuet (Fondation Mérieux), Sophie Mathewson (Wellcome Trust) and Philip Horgan (INASP).

- **Adapt the Access to Medicines Index self-assessment tool for RCS evaluation.** All participants were interested in pursuing this idea, and remaining involved in deciding on indicators which might be involved in the self-assessment.
    - LSTM will discuss the idea of a feasibility of the self-assessment idea with partner universities.
    - Once a prototype of the indicators is conceived, CNRS can share it with French institutions and think about broadening the business case.
    - INASP suggested exploring potential for sharing the self-assessment via universities' online Moodle platforms.
    - WHO-TDR will liaise with WHO's Regional Training Centres to discuss the idea.
    - LSTM and UKCDS will draft a 2 page document to put the idea on paper and have a structure to engage those who did not attend the workshop.

- **ESSENCE's evaluation framework to be adapted.** Garry Aslanyan from ESSENCE/WHO-TDR will lead on this, in consultation with Fondation Mérieux, Wellcome Trust, UKCDS and potentially Pasteur and IRD.

This report was written by Jamie Enoch at UKCDS, and aims to provide as accurate an account as possible of the workshop proceedings. It does not constitute any institutional policy of the organising committee (made up of the UK Science and Innovation Network, Liverpool School of Tropical Medicine, Royal Society, Institut de recherche pour le développement and UKCDS).

Annex 1: Acronyms

AllEnvi	Alliance Nationale de recherche pour l'environnement / National Alliance for Research on the Environment
CIRAD	Centre de coopération internationale en recherche agronomique pour le développement / International Centre of Cooperation in Agricultural Research for Development
CNRS	Centre national de la recherche scientifique / National Centre for Scientific Research
CRU	Capacity Research Unit (at LSTM)
DFID	UK Department for International Development
ESSENCE	Enhancing Support for Strengthening the Effectiveness of National Capacity Efforts
GSK	GlaxoSmithKline
ICT	Information and Communications Technology
INASP	International Network for the Availability of Scientific Publications
IRD	Institut de recherche pour le développement / Institute of Research for Development
JEAI	Jeunes équipes associées à l'IRD / Emerging Research Teams Associated with IRD
LSHTM	London School of Hygiene and Tropical Medicine
LMIC(s)	Low and Middle Income Country(/ies)
LSTM	Liverpool School of Tropical Medicine
MCDC	Malaria Capacity Development Consortium
RCS	Research Capacity Strengthening
SIN	UK Science and Innovation Network
UKCDS	UK Collaborative on Development Sciences
WHO-TDR	World Health Organisation Special Programme for Research and Training in Tropical Diseases

Annex 2: Workshop Participants

Garry Aslanyan	WHO-TDR
Danielle Barret	CIRAD
Imelda Bates	Liverpool School of Tropical Medicine
Mariana Beija	French Embassy London
Lisa Bonadonna	GSK
Denis Despréaux	AllEnvi
Christian Devaux	IRD
Ann Dufton	GSK
Jamie Enoch	UKCDS
Hervé Fritz	CNRS
Hans Hagen	Royal Society
Beatrice Halpaap	WHO-TDR
Philip Horgan	INASP
Arnaud Lalo	CNRS (Institut des sciences biologiques)
Christophe Longuet	Fondation Mérieux
Alison MacEwen	UK Embassy Paris
Sophie Mathewson	Wellcome Trust
Hazel McCullough	LSHTM
Edouard Michel	CNRS (Institut Ecologie et Environnement)
Gláucia Paranhos-Baccalà	Fondation Mérieux
Ian Thornton	UKCDS

Annex 3: Workshop Agenda

Franco-British workshop: Evaluation of research capacity strengthening in sub-Saharan Africa

09:30 – 09:40	Welcome and introduction *Ian Thornton, UKCDS*
	Two evaluation case studies
09:30 - 10:15	The Malaria Capacity Development Consortium *Prof Imelda Bates, Liverpool School of Tropical Medicine*
10:15 - 11:00	Assessing capacity building outputs toward the "emerging research team (JEAI) program *Christian Devaux, IRD*
11:00 - 12:00	Plenary discussion: **The definition of impact, and the challenge of measuring it**
12:00- 13:00	Lunch
13:00 – 13:45	Break-out discussions: **Reconciling conflicting demands for evaluations - contextual, motivational and environmental factors**
13:45 – 14:30	Feedback and discussion
14:30- 15:00	Coffee break
15:00 – 15:45	Break-out discussions: **Reconciling conflicting demands for evaluations - what would the key principles/indicators of an evaluation toolkit look like?**
15:45 - 16:30	Feedback and discussion
16:30 - 17:00	Discussion on future (practical) steps
17:00	Close