TOWARD THE NEXT IDRC CORPORATE STRATEGY:

METRICS FOR DEVELOPMENT

A contribution to an Environmental Scan carried out by the
International Development Research Centre, Ottawa, Canada

CENTRE FOR SCIENCE, TECHNOLOGY AND INNOVATION INDICATORS

HUMAN SCIENCES RESEARCH COUNCIL

Private Bag X9182
Cape Town
South Africa

27 October 2008

Principal Investigator: Michael Kahn

mkahn@hsrc.ac.za

EXECUTIVE SUMMARY

This paper 'Metrics for Development' is a contribution to the environmental scan that informs the scoping of the IDRC Corporate Strategy 2010-2015.

Funders that promote development research need to know that they are achieving what they set out to achieve and use measurement and evaluation methods to determine this. These methods may be both quantitative (indicators or metrics) and qualitative.

The paper was informed by both primary and secondary research. Telephonic and e mail interviews and short discussions were conducted with staff of the World Bank, Ford Foundation, Google Foundation, Bill and Melinda Gates Foundation, Sida and Acumen Fund. Desk research covered journals, book and web sources.

The paper commences with consideration of the problems of measuring socio-economic development. It then drills down to the level of the innovation system and its measurement pointing to the important unifying role that the OECD has played in providing guidelines and setting standards for such measurement and pushing its boundaries.

Next follows the program and project micro level where the measurement and evaluation practices of the major donors, donor networks and philanthropic organizations are considered. Consideration is then given to the measurement gaps and the alternative metrics marketplace especially the work of the 'new kids on the block.'

The review of Jones and Young (2007) for DFID found a broad diversity of definitions of research, research themes, and research processes and no evidence of common measurement approaches. Qualitative evaluation methodology remains the method of choice to assess donor interventions making the population of metrics difficult. Among the established bilateral and multilateral donors the use of metrics as a scoreboard for impact assessment is largely notable by its absence. Despite the general recognition of the value of evaluation, and the willingness of grantees to do it, funding and infrastructure are inadequate.

It is shown that the methodological frontier of impact assessment of social innovation and the management of the associated information is defined by the new kids on the block, the

Google Foundation, Bill and Melinda Gates Foundation, Skoll Foundation and Acumen Fund. Their work on alternative metrics, visualization and portfolio management tools is path breaking.

This path breaking work in the United States is complemented by that of the NESTI community of practice that is developing the measurement framework for the OECD Innovation Strategy. NESTI is driving projects on the measurement of technological and non-technological innovation.

In the immediate term this work does not offer an easy way out for the construction of metrics of development research - the NESTI measurement toolbox is best suited to industrialized countries though it can be used in adapted form in developing countries. As yet it does not speak to the matters of public sector or social innovation. Conventional scientometrics may be applied to development research initiatives subject to the caution that translation of basic research moves at a different pace to the funding cycles of the donor community.

Within agencies such as DFID[1], NSF, and the European Union pressure appears to be growing for greater attention to be given to the quantitative since such variables are essential inputs for performance modeling of impact assessment. The EU demand for regulatory impact assessment has had a spillover into impact assessment of the Framework Programmes with particular stress on micro-economic modeling. Such modeling is of course dependent upon the availability of the appropriate variables, for which read metrics. The need for better quantitative data is inescapable.

The main conclusions of the paper are as follows:

1. Donors, grantmakers and foundations operate according to their own theory of change and there is little evidence of common practice in the use of evaluation methods let alone metrics. They all express commitment to the importance of evaluation, but they fund and institutionalize this to varying degrees.

[1] DFID (2008), NSF (2008), NWO (2007)

2. On both sides of the Atlantic pressure is growing for quantitative assessment and modeling. In the US this is found in the call for a Science of Science and Innovation Policy and in Europe for *ex ante* impact assessment of the Framework Program 7.
3. The methodological frontier of social impact assessment is to be found in the approaches of US foundations and social venture capital funds.
4. The methodological frontier for measuring technological and non-technological innovation in firms is defined by the work of NESTI. The Frascati family of manuals for measuring R&D and innovation may travel with some adaptation to developing country contexts. There are no NESTI guidelines for the measurement of public sector or social innovation at present.
5. The various approaches to metrics of research for development – as research, as development, as social change- are potentially incommensurable and suggest a need for a pluralistic approach. A dialogue across the divide of communities of practice is overdue.

A broad observation is that there is no single or simple answer to determining 'value for money' or deciding 'who benefits?' Metrics for development are but one tool for addressing their needs. They are always subject to interpretation in the context they were generated. Pluralistic approaches may be expected to continue, and may be desirable. For IDRC with its mandate of promoting research and research capacity development this suggests that a dual measurement system might be a pragmatic choice. Metrics for (and of) development remain a work in progress.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

This study was carried out over a five month period and involved the contribution of many people spread across two continents.

In Canada thanks are due to Bruce Currie-Alder and the anonymous Reviewer as well as staff of the IDRC.

In Europe appreciation is due to Nina Lejonhud at Sida. In North America thanks are due to Michael Ehst at the World Bank, John Colborn at Ford Foundation, Faye Twersky of the Bill and Melinda Gates Foundation, Brad Pevsner of Acumen Fund, and Aleem Walji at Google Foundation.

My gratitude is expressed to all CeSTII staff who assisted with this work with a special word of thanks to Michael Gastrow for assisting with the review of high-level metrics, and to William Blankley for a critical reading and help with some references.

Any errors of omission or commission remain those of the principal investigator.

Michael Kahn

ACRONYMS

3IE	International Initiative for Impact Evaluation
BACO	Best Available Charitable Option
BMGF	Bill and Melinda Gates Foundation
DAC	Development Assistance Committee
DCD	Development Cooperation Directorate
DFID	Department for International Development
GCI	Global Competitiveness Index
GDP	Gross Domestic Product
GERD	Gross Expenditure on Research and Development
GNH	Gross National Happiness
GRDI	Gender Related Development Index
HDI	Human Development Index
HPI	Human Poverty Index
IDRC	International Development Research Centre
IEG	Independent Evaluation Group
KEI	Knowledge Economy Index
LFA	Logical Framework Approach
MDG	Millennium Development Goals
NESTI	National Experts on Science and Technology Indicators
NONIE	Network of Networks for Impact Evaluation
NSF	National Science Foundation
NWO	Nederlandse Organisatie voor Wetenschappelijk Onderzoek
OECD	Organization for Economic Co-operation and Development
PDMS	Portfolio Data Management System
R&D	Research and Experimental Development
RBM	Results Based Management
RCT	Randomized controlled trial
RCUK	Research Councils UK
Sida	Swedish International Development Cooperation Agency
STI	Science, Technology and Innovation
TAI	Technology Achievement Index
UNDP	United Nations Development Programme

TOWARD THE NEXT IDRC CORPORATE STRATEGY:
METRICS FOR DEVELOPMENT

I. METRICS AND THE ENVIRONMENTAL SCAN

Funders of development research require measurement and evaluation to monitor the use of resources and to meet accountability obligations

IDRC has embarked on an environmental scan to inform the development of its future corporate strategy. This paper is part of that environmental scan and addresses the topic of 'metrics for development.'

As a donor IDRC promotes development research utilizing public funds. It is part of a wider community of donor organizations, bilateral and multilateral, as well influential 'philanthropic organizations such as foundations or social venture funds active in promotion development research. These organizations (Table 1) need to know if they are delivering on their mandates and thus carry out program and project monitoring and evaluation, and most importantly, impact assessment of their efforts. These evaluative actions are essential for demonstrating effective and efficient use of resources and for organizational accountability to beneficiaries and resource providers.

Table1: The Top 10 Development Research Donors

Agency (Country)	**Annual Spending**	**Year**
Gates Foundation (USA)	$450m	2006
USAID (USA)	$282m	2002
European Union	$254m	2007/08
IRD (France)	$220m	2005
DFID (UK)	$174m	2005
Wellcome Trust (UK)	$143m	2005/06
SIDA (Sweden)	$135m	2006
Medical Research Council (UK)	$120-160m	2006
IDRC (Canada)	$110m	2006
World Bank	$ >100m	2005

Source: Jones and Young (2007)

For more than half a century considerable effort has gone into developing measurement and evaluation methodologies that seek to understand the relationship between project inputs and outputs. This effort is driven by many considerations: value for money, political accountability, the desire to effect positive change, the creation of enabling environments that empower recipient communities, and the promotion of wellbeing. Demonstrating the effect of such change initiatives is as important for funding organizations operating in their home environments as it is when they engage in development aid be this bi-lateral or multi-lateral.

In some instances change is directly measurable as in levels of participation, improved health or nutritional status, improved quality of water supply and so on. So to take a well-known example the Green Revolution had a clear base in scientific research in part funded by the Ford and Rockefeller Foundations – their inputs underpinned the development of high-yield hybrid food crops that eventually impacted on food security in developing countries.

In most instances change is more subtle, difficult to detect and attribute especially where it is aimed at new (desirable) behaviors whose measurement is more subjective than the 'objective' measurements of scientific research.

The possibility of determining with precision how inputs are related to observed changes and thereby ascribing a causal relationship is of central interest in development aid.

Funding organizations see possibilities for change in different ways

Development praxis is a fuzzy concept and the impact that donor and foundation funding has on the wellbeing of recipients remains a work in progress. As McGregor (2007: 349) reminds, 'international development is fundamentally about competing visions of what wellbeing is or should be … (and) … what is meant by desirable and socially feasible.' With such differences of expectation it might be expected that what donors and foundations value and measure will also show wide variation.

Each donor and philanthropic organization operates within the confines of their own theory of change (Box 1). Such theory is informed by how they characterize their base society and economy, and how they understand the behaviour of the intended first-line beneficiary, be this a local or participating foreign organization, firm or country.

Box 1: The importance of a 'theory of change'
To understand system impact we need a means of assessing how the things we achieve directly may be likely to lead to it. Many people now use a theory of change for this purpose. At Keystone, we promote a participatory approach to building a theory of change in which those who are meant to benefit are central to creating the vision of success. (Bonbright, 2007: 30)

There may even be confusion regarding the term 'theory of change' since this term has been appropriated to characterize an approach to change management advocated by Carol Weiss and co-workers. Understanding of the change process is critical in that it explicitly or implicitly shapes the relationship between the donor and 'partner.' Causal attribution is made yet more difficult as funding agencies differ in their conception of the development process. This necessarily influences how they transact their business and seek evidence for its success. Donors that operate within the assumptions of 'lean government' are likely to drive a different set of emphases to those that see developing countries as having to break out of the centre-periphery relationship that characterizes historic underdevelopment. The emerging digital and knowledge divides (Chattaway et al, 2003) speak to the continuing problem of underdevelopment.

Given the textured nature of the development process, narrow concentration on the 'financial' will fail to capture the essence of change so that a broad armory of performance measures including qualitative evaluation, and quantitative and qualitative indicators is needed.

Evaluation and impact assessment methodologies are generally qualitative. Disappointment with the success in translating qualitative evaluations into quantitative indicators has led to interest in constructing 'metrics' that might be better suited to inform accountability reporting and decision-making in regard to development aid in general and development research in particular. Development research faces these problems as it seeks to make a difference.

The paper addresses three questions

IDRC seeks to be informed of current donor practice with metrics, and possible alternative metrics for showing that investments in development research make a difference. Three broad questions are investigated through this paper:

1. How do research donors know they are doing what they set out to do?
2. Are alternative metrics needed?
3. What is the market for alternative metrics?

The paper is informed by both primary and secondary research. Telephonic and e mail interviews and short discussions were conducted with staff of the World Bank, Ford Foundation, Google Foundation, Gates Foundation, Sida and Acumen Fund. Desk research covered journals, book and web sources.

It commences by examining macro-level metrics of socio-economic development, namely the macro level. Next, given the focus on development research, it moves to the meso level of innovation systems within which research activities are situated. Donors to development research are actors at this meso level. This leads to discussion of measurement of research and innovation and the difficulties of applying methodologies suited to industrialized economy systems to developing countries. Consideration is given to the suggestions for new indicators that emanated from the OECD Blue Sky II conference and the call for the development of a 'science of science policy.'

This is followed by an overview of the project evaluation practices of DFID, IDRC, the World Bank and Sida, the Ford Foundation, Bill and Melinda Gates Foundation, and Google.org. Reference is also made to the evaluation practice of three domestic agencies – the British RCUK, NSF and Dutch NWO[2]. The discussion then turns to the matter of alternative and emerging metrics and in particular the tensions between peer review and metrics as well the rising influence of regulatory impact assessment. The funding community has been joined by large philanthropic organizations with origins in the information technology revolution. These 'new kids on the block' define the cutting edge of practice with metrics and their management and are developing creative approaches to output, outcome and impact assessment.

The overview suggests that pluralistic approaches may be expected to continue, and are in fact desirable. A set of five findings is offered for further discussion.

For a funder such as IDRC that supports both research and research capacity development it is suggested that a dual measurement system might be a pragmatic choice. A single set of metrics will be insufficient.

[2] A more detailed account of the donor and philanthropic organization evaluation practices is available as a separate annex.

II. METRICS AND SOCIO-ECONOMIC DEVELOPMENT - MACRO LEVEL

Metrics mean different things in different contexts

Metrics entail quantitative measurement according to procedures specific to a domain of interest. They may be constructed from historic data in which case they provide review information. When used to specify current parameters they are of operational type, and when used to set future levels of performance they may be termed synoptic (Kahn and Swanborough, 1999). Quantitative indicators are metrics, and many organizations and writers frequently use the terms metric and indicator interchangeably.

The term 'metric' has different meanings according to the context in which it is specified and used. Metrics as used in software development or in business performance measurement are not the same as those used to measure national development.

Metrics used to capture national development include the targets of the Millennium Development Goals (MDGs), GDP, GDP/capita, the Gini coefficient and Robin Hood index, statistics of education, health and demographics, and composite indicators such as the UNDP Human Development Index. The indicators for the Millennium Development Goals are synoptic; the others are generally of review type.

The set of available metrics is under constant development, a case in point being the work of the World Bank Institute on the Knowledge Economy Index (KEI) a composite indicator straddling knowledge (the science system) and development. The Human Development Index (HDI) [3] is arguably one of the most widely used composite metrics. The first Human Development Report (UNDP, 1990) introduced this proxy for development by combining indicators of life expectancy, educational attainment and income into a composite index, the HDI.

Given the imperfect nature of wealth as a gauge of human development, the HDI offers a powerful alternative to GDP/capita for measuring relative socio-economic progress. An alternative metric that gained few adherents is that of Gross National Happiness (GNH) as proposed by the King of Bhutan in 1972. The GNH is a construct of 'objective' and subjective measures with resulting lack of comparability.

[3] http://hdr.undp.org/en/statistics/indices/hdi/

The Gender Related Development Index (GRDI) is a composite index, also developed by the UNDP that measures human development on the same dimensions as the HDI while adjusting for gender inequality in those basic dimensions. The final example is the Technology Achievement Index (TAI) developed by the UNDP to measure a country's ability to participate in the network age. The index aims to capture how well a country is creating and diffusing technology and building a human skill base - reflecting capacity to participate in the technological innovations of the network age. The TAI index has shown a high correlation with the HDI – more so than with income.

The above are macro-level metrics generally beyond the influence of individual funding agency interventions. For development research in particular it makes sense to understand impact at the next level down, the meso level, that of innovation systems.

III. INNOVATION SYSTEMS – THE MESO LEVEL

Development research is an activity that should be understood within the context of each unique innovation system

Loosely put development research (Box 2) is a scientific activity that seeks to promote socio-cultural and economic change in developing countries. This research activity does not occur in isolation but takes place within individual country innovation systems. This assertion requires explication. What is meant by the term 'innovation system' and how are the activities of that system measured?

> Box 2: Development research is …
> 'Issue-driven research concerning the analysis of global and local processes of cultural, demographic, economic, environmental, political, technological and social change in low and middle income parts of the world, with particular reference to structures and institutions; the changing relationships between developed and developing countries; and the critical interrogation of theories of these processes and relationships, and of development policy.' (UK Research Assessment Exercise, quoted in Jones and Young, 2007)

The innovation systems approach is a recent attempt to better understand the relationship between knowledge production and the emergence of new goods, processes and services (Freeman, Clark and Soete, 1982). It demarcates a rupture with the previous linear model of innovation that drew on the pivotal role that government scientific research and development had played in the Second World War. The idea was that in peace time corporate R&D would naturally lead to marketable inventions. It was expected there would be a straight line relationship between research and the appearance of new products in the market place. However the Western economic slowdown of the late 1970s and 1980s as opposed to the rise of Japan and the Asian Tigers persuaded that a new conception of the relationship between research and economic growth was needed.

The innovation systems approach argues that innovation is fundamentally a non-linear process. This entails mutually reinforcing knowledge production and exchange activities among a range of actors that acquire, generate and transform knowledge into useful processes and products. The effectiveness and efficiency of these activities depend on factors such as the knowledge infrastructure, knowledge stocks, prior technological learning, absorptive capacity, and framework conditions. The theoretical basis of the approach springs from

evolutionary economics (Nelson and Winter, 1982) in which knowledge is understood to be a non-rival good that shows increasing returns over time.

The role of governments in innovation systems (at least in the OECD countries prior to the crash of 2008) was primarily to ensure the framework conditions that would promote innovation activities in the private sector and research in universities rather than attempting to correct for market failure. In this sense the innovation systems approach resonates with the notion of lean government.

Funders of development research provide funds and expertise to the actors of innovation systems by working with universities, organs of civil society, government laboratories, and increasingly with the private sector.

The innovation systems approach has served as an organizing device that promotes the development of standardized metrics of innovation

Governments undertake the measurement of innovation activity through R&D[4] and innovation[5] surveys, research evaluations, cost benefit studies, bibliometric analysis, monitoring patenting and licensing, the technology balance of payments, and personnel mobility. However it remains impossible generally to predict when and if basic research will translate into a new product or process. Investing in basic research is inherently risky.

The demonstration of the future benefit of investment of public funds in R&D is demanding of a paper in its own right (see e.g. Martin and Salter, 1996; OECD, 2007a; DFID, 2007; Commonwealth of Australia, 2007; National Audit Office, 2007). The rate of return on R&D investment is calculated separately for private and social returns. This is necessary since it is virtually impossible for the initiator (the private beneficiary) fully to appropriate the knowledge inherent in the research. The collective wisdom has it that investments in agricultural and health research offer the highest social rates of return that may be as high as

[4] Research is creative work and original investigation undertaken on a systematic basis to gain new knowledge, including knowledge of humanity, culture and society; Development is the application of research findings or other scientific knowledge for the creation of new or significantly improved products, applications or processes (OECD, 2002).

[5] A product innovation is the market introduction of a new good or service or a significantly improved good or service with respect to its capabilities; A process innovation is the implementation of a new or significantly improved production process, distribution method, or support activity for your goods or services (OECD/Eurostat, 2005).

200% while that for industrial research is in the order of 30% (Nadiri, 1993; Jones and Williams, 1997; Sen and Hoare, 2005).

Systematic country measurement of R&D may be regarded as commencing with the production of the guidelines codified in the Frascati Manual of 1963. Since then the OECD National Experts on Science and Technology Indicators (NESTI), a Working Party of the Committee on Science and Technology Policy has harmonized the standards for the collection and interpretation of such measures codified in the Frascati family of manuals (OECD, 2002: 3). The NESTI work on innovation was done jointly with the European statistical office, Eurostat, and the resulting manual is a joint publication (OECD/Eurostat, 2005). The standardized metrics of the OECD Science and Industry Directorate provide an aggregate view of the activities of R&D and innovation attributes of innovation systems.

At least three limitations characterize these metrics. The first is that the metrics do not adequately speak to the quite different institutional composition of the countries being studied. The 'path dependence' of evolutionary economics is a reality not only at firm level but also at country level. The second difficulty is that of the scale dependence of many of metrics (Katz, 2000). So for example India's world share of scientific publications places it at 14th position but normalized by R&D expenditure, India rises to the first rank (Government of India, 2008). The third is that when applied to a firm or an organizational process one can in principle understand how inputs lead to outputs. When scaled up to the level of the innovation system, complexity limits such attribution.

The toolbox of metrics constitutes a technology that is evolving along with the unfolding information technology revolution

The innovation systems approach is pre-paradigmatic, and its measurement is thus imperfect and evolving. As Gault (2007: 12) explains: '… the manuals, the tacit knowledge held by the experts, and the formal language used to discuss the measurement and interpretation issues are equivalent to a technology and, like machine-based technologies, they do not always behave, or diffuse, as expected'. The NESTI technology has also informed the approach to measurement of other OECD Working Parties dealing with indicators for the nascent technological revolutions of the information society, biotechnology, and more recently nanotechnology. These will all change with time.

Drawing on the perceived value of the 1996 OECD Blue Sky Conference, OECD together with Statistics Canada organized the 2006 Blue Sky II Conference to contemplate what indicators would be needed to measure and describe innovation systems a decade into the future. In his keynote address US Presidential Science Adviser John Marburger reiterated his previous call for what he has termed 'a science of science policy.' Such a science could in principle provide the theoretical basis for future STI indicator development.

The main thrusts of the 'Blue Sky II' conference have been collated in an OECD publication (OECD, 2007b). One theme was the need for improved understanding of the actors in the system of innovation at the micro level. A second was for the construction of internationally comparable composite indicators, for indicators of knowledge exchange and indicators of commercialization. Unsurprisingly the need for impact indicators was advocated for politically-charged areas: biotechnology, health research, sustainable development, climate change, energy, and food and biodiversity security. This serves to remind that STI indicators and metrics and their translation into useful devices to inform policy are a work in progress (Godin, 2005) with deep roots.

As important as the technologies of measurement are those that enable communication with the intended audience. Those to whom the indicators 'tell a story.' An excellent example is the dynamic web-based depiction of correlations between the macro level indicators of development provided by means of Trendanalyzer on www.gapminder.org. Another example is the work on the connectedness of science (Börner and Scharnhorst, 2008; Igami and Saka, 2007) that utilizes social network analysis and information technology visualization tools.

The innovation systems approach is an evolving body of knowledge

A further complexity arises in respect of the specification of the actors and boundaries of innovation systems. For some industrialized and open economies, the concept of a sovereign national system of innovation is now moot. For example much of the business sector R&D in Ireland and China is performed in the R&D laboratories of foreign multi-national corporations as part of inward FDI transfers - at least to October 2008! Where should this R&D activity be recorded? In the host country or the country that appropriates the knowledge? IBM for example claims that 'the sun never sets at IBM research' (Kaiserswerth, 2007). How does one count the R&D effort of IBM – in the United States, Switzerland or globally? Should one account for R&D where the researchers are paid, where the work is

done or where the funds originate? Has the concept of the national system of innovation outlived its usefulness for industrialized countries?

A second challenge to the usefulness of the national system of innovation comes from studies of agglomeration effects leading to ideas such as regional and sectoral systems of innovation Malerba (2004). Innovation systems thinking is an approach, not a paradigm.

Does the concept of the innovation system apply uniformly to developing countries? From an epistemological standpoint the answer must be 'no' since the very evolutionary economics that informs innovation systems thinking takes as starting point the unique histories of institutions. If institutions exhibit path dependence then so do countries.

So Albuquerque (2003) categorizes Brazil South Africa, India and Mexico as 'immature systems of innovation.' Mouton and Waast (2008) on the other hand believe that innovation systems have been de-institutionalized in Africa, and do not really exist in many of its countries. But Mytelka (2004) is catholic arguing that all countries have elements of innovation systems. From this standpoint it follows that the metrics used to measure innovation system activity in industrialized countries are applicable in emerging and developing economies. Of course their estimation requires adaptation of methods as for example through the 'Bogota Manual' (RICYT/OEA/CYTED, 2001) for Latin American innovation surveys and the work of the UNESCO Institute for Statistics on measuring R&D in developing countries (Fernandez Polcuch, 2008).

IV. MEASUREMENT AND EVALUATION OF DEVELOPMENT RESEARCH – MICRO LEVEL

The donor community ascribes to different theories of change but shares common approaches to project management and assessment, notably the LFA and RBM

In their attempt to organize and manage the projects they fund donors have come to follow two main approaches - the Logical Framework Approach (LFA) and Results Based Management (RBM), the latter sometimes incorporating the former. LFA has been adopted by many development agencies including the UN system, GTZ, CIDA, Sida, NORAD and USAID (Sida, 2004).

The logical framework approach involves specification of project objectives and activities whose attainment may be measured through objectively verifiable indicators (OVIs). In principle the specification of objectives, activities, indicators and targets provide the basis for evaluation of project relevance, effectiveness and efficiency and the construction of associated metrics.

The usefulness of LFA resides in the extent to which the OVIs enable outcome and impact measurement. However 'most organisations can find indicators for outputs … but … are struggling to find (impact) indicators' (Bakewell and Garbutt, 2005: 9). The difficulty may be that indicators are frequently presented as targets, rather than as an 'observable change or event' or 'milestones' (Anderson, Grude and Haug, 2004).

The Results Based Management (RBM) doctrine of many OECD countries donors is congruent with those countries' general approach to improving the efficient and effective use of public resources. In the case of the United States and United Kingdom this is consistent with domestic policy instruments, the Performance and Results Act and the Financial Management Initiative respectively.

Evaluation methodologies are the tool of choice to determine if the objectives of development funding have been met

There is a vast literature dealing with the merits of different evaluation approaches and as might be expected contestation between different schools of thought. The purpose of evaluations varies: so these may be formative or summative, focusing on outputs, outcomes

or impacts[6]. The choice of the actual evaluation methodology is pragmatic so that in practice this turns on what best fits the situation at hand. In terms of the popularity, utilization focused evaluation (Patton, 1996) has gained many adherents.

Donor forums have emerged to share and harmonize approaches, but there is still insufficient quantitative data for the population of a common of common metrics even if such existed

The Development Assistance Committee (DAC) working party of the OECD Development Co-operation Directorate (DCD) promotes evaluation activity through the DAC Network on Development Evaluation and the associated DAC Evaluation Resource Centre (DEREC). A second means of knowledge sharing is the Network of Networks on Impact Evaluation (NONIE) that operates from the premises of the World Bank. NONIE comprises the DAC Evaluation Network, the UN Evaluation Group, the Asian Development Bank Evaluation Cooperation Group (ECG) and regional evaluation associations. 3IE is a new organization intended to improve the impact evaluation praxis and usage. Its formation follows the Centre for Global Development (CGD) report '[When Will We Ever Learn? Improving Lives through Impact Evaluation](#).' Savedoff and Levine (2006: 3), note that 'demand for knowledge about impact is intensifying because of commitments to substantially increase aid flows… International commitments, including the Millennium Development Goals, and the Paris Declaration on Aid Effectiveness, create both a challenge for impact evaluation work and an opportunity to learn.'

The conduct of evaluation varies from donor to donor, and from project to project. Some evaluations use classical survey technique; others scrutinize records or conduct interviews with key informants; another project might lend itself to the Delphi method.

The World Bank for example is a strong advocate of impact assessment that ideally requires estimation of the status of beneficiaries in the absence of the donor-sponsored intervention i.e. an opportunity cost determination. The Department for International Development (DFID) subscribes to RBM and its approach to monitoring and evaluation[7] is codified in its *Guidance on Evaluation and Review for DFID Staff* (DFID, 2005). That document explicates

[6] The term 'evaluation' is often used interchangeably with the term 'assessment.'

[7] 'Evaluation' in DFID terminology refers to external evaluation, whilst internal evaluation is termed 'Review.'

the link between evaluation (deductive or inductive), the intervention strategy or theory of change and the LFA. The Swedish International Development Cooperation Agency (Sida) is a major source of development aid and an important supporter of development research. It has been a strong promoter of the use of LFA at project level, but a perceived overemphasis on inputs and activities has led it to champion 'managing for development results' (Sida, 2007a: 3). Sida advocates the use of a 'results chain' model to track the various steps and contributions that lead to the desired impact. Sida has a strong commitment to knowledge sharing and freely disseminates its in-house evaluation manual 'Looking Back, Moving Forward' (Sida, 2007b). Evaluation informed by the utilization-focused method of Patton (1996) is institutionalized at IDRC with actual methodology dependent on circumstance. The organization has made a strong commitment to 'Outcome Mapping' (Earl, Carden and Smutlyo, 2001). Earl, Carden and Smutlyo (2001: 6) argue that '… when donors and recipients try to be accountable for achieving impact, they are severely limiting their potential for understanding how and why impact occurs' with the danger of producing 'clueless feedback.' They claim that outcome mapping is a better means to tell the story.

At the Ford Foundation evaluation, both quantitative and qualitative is used for learning and external communication of results but no metrics of development are in place (Ford, 2008). The Rockefeller Foundation places monitoring and evaluation under the direct supervision of a Vice-President and 'fully integrates verifiable methods of assessing progress' (Rockefeller, 2007).

The Bill and Melinda Gates Foundation (BMGF), a relative newcomer is now the largest charitable research funder with interests in education, health and population, and equally diverse approaches to determining the results of this investment. It has no organization-wide approach to evaluation or impact measurement and follows best practice in each field of interest as e.g. randomized clinical trials (RCT) in health-related research. BMGF findings on the success or failure of its interventions are disseminated through case studies on their website. The development of common metrics is under discussion within BMGF.

Funders of development research all recognize the importance of evaluation but this is accorded varying status among the organizations. In some it is a senior corporate function with the expectation of regular evaluations cascaded from the top of the organization down to

project level. In others the commissioning of evaluation is a decision made by project officers.

It is thus theoretically possible to cascade backwards from intended impacts down to the inputs that Bonbright[8] describes as moving from the 'story of what success looks like' back to intent but harmonization of core standards, common terminology, indicators and benchmarks remains at an early stage, and there remains a dearth of quantitative measurement without which metrics cannot be populated. Showing causality remains elusive (Box 3).

> Box 3: Demonstrating causality
>
> It is not easy to measure the impact that research results have on development. In fact, it is even difficult to separate out those factors that actually result in change over time. This strategy accepts that individual research programmes rarely bring about change on their own. It is more likely that people will act on evidence that is built up over many years, in different contexts, once this has been communicated effectively. Even when research helps shape policy and practice, it can sometimes take years to see the benefits to poor people (DFID, 2005: 42)

While funding agencies evince some elements of a shared theory of change, and change management there is a diversity of approaches to measurement and evaluation. The Jones and Young (2007) review of the major donors also notes that limited knowledge management practices are in place alongside the absence of what they term 'good development research donorship.'

Application of the NESTI toolbox does not offer an easy way out

Given the above diversity of approaches and the difficulties associated with the extraction of indicators from project reports it is tempting to ask whether the NESTI toolbox offers a simple way forward, the more so where development research is supported in universities or other research institutes.

It is suggested that the direct use of the NESTI toolbox may not yield the anticipated measures. This arises from problems associated with institutional make up and their information systems. As mentioned above, innovation systems and institutions exhibit path

[8] Comments made by David Bonbright at the 2008 Skoll World Forum, Said Business School, 27 March 2008.

dependence. So in many developing countries (Uganda, Tanzania) and even in some emerging economies (e.g. Argentina, Mexico) the major site of research is in the universities. At first blush this is good news – one 'knows' where research is carried out, so it is straightforward to measure it. Not quite.

The institutional make up and the behaviours of developing country innovation systems may be quite different to what is encountered in the North. For some developing countries the three R&D sectors of industrialized country innovation systems – universities, government and the private sector are joined by a significant fourth, funders of development research. In some developing countries this fourth sector may be the largest, giving rise to problems of definition and sovereignty. The problem is exacerbated by the incompleteness of information systems in many (developing) countries. Not only are they incomplete but they are hosted by agencies that may not, cannot or will not exchange even the limited information they have.

Tracking and attributing the flows of funds for development research is particularly difficult the more so where a project includes both national and foreign staff that are paid in different jurisdictions. If there is no agreement on the attribution it becomes difficult accurately to determine gross national expenditure on R&D. In some cases the donor flows may be substantial enough to substitute for the role of national governments and other local actors, the more so where private sector R&D is limited, as in countries where resource extraction dominates. Suitable metrics that track donor funded research flows are needed.

A second problem facing developing countries is the determination of the nature of R&D. A standard way of detecting higher education R&D is by means of journal publication outputs. This will however not suffice where research staff undertake research consulting and other contract work that does not lead to publications. Such consulting tends to be in the applied research domain. It is unlikely to show up in country R&D statistics since it is often undeclared (for tax reasons). Donor agencies are often the source of these research consultancies. It may well be the case that the way to estimate research activity will be from headcount data since the use of academic diaries may be unacceptable.

A third issue is to do with the desire of funders of development research to demonstrate impact. This desire has a twist to it: impact is dependent on what is funded. If the funding goes to basic research, the impact will arise very far down the value chain, as for example in

the future work and value of the students that received their research training because that basic research project was funded. The best of these students trained in basic research may well land up as immigrants to foreign shores.

This observation raises a dilemma for funders. If a funder wishes to support university research and demonstrate impact, then support might better be given to applied research that is driven by local needs. In principle the impact should thus be more readily detectable. Ensuring that the donor research agenda is demand driven and thus locally contextual may also act to reduce brain drain.

V. ARE ALTERNATIVE METRICS NEEDED?

In the European Union the demand for regulatory impact assessment to demonstrate value for money is exerting pressure on scientific research to meet the same criteria

Hovland (2007) shows why it is that the evaluation of policy research requires different instruments to those used to evaluate academic research. Academic research relies on peer review and bibliometric analyses but these do not speak to issues such as impact, changes in behaviour, or the building of relationships. The same requirements arise in evaluating development research with its intention on making observable and verifiable change.

Impact assessment of scientific research is now being demanded. *Post facto* impact assessment in the European Union Framework programmes now goes back a decade and *ex ante* assessment is also on the agenda, the first such being carried out in 2005 (Delanghe and Muldur, 2007). The new demand for project impact assessment parallels that for regulatory impact assessment (RIA) as expressed by the Mandelkern Group.

The indicative format for RIA (European Union, 2002) includes quantitative measures that lend themselves to indicator construction and has created the pressure for similar measures of the Framework Programmes.

The *ex ante* assessment of Framework Programme 7 was based on project reports, bibliometric evidence and innovation survey data analysis. 'Nemesis', a general equilibrium econometric modeling tool was then used for the *ex ante* projections. The approach accords with the view that '… *ex ante* evaluation focuses primarily on value for money ...' (European Union, 2002: 3).

Delanghe and Muldur (ibid. 14) suggest that a standardized model would be needed across '… the field of S&T indicators, where the focus should shift from inputs (R&D intensity, HRST, etc) to flows (trans-national collaboration patterns, international co-publications, international co-patents, trans-national flows of R&D investment and people, etc. The discussion in section IV suggests that it would be even more difficult to construct such a model in developing countries.

Assessment of research in higher education swings back and forth between peer review and metrics

The October 2007 issue of *Science and Public Policy* was devoted to the controversy around new forms of research assessment in higher education. In the case of the UK this entails a move away from panel-based review to that of more quantitative metrics while Australia is shifting in the opposite direction. Introducing the special issue Donovan (2007: 539) posed the question: '…does the future of research evaluation rest with: metrics or peer review; the seemingly objective or the subjective; remote or embedded knowledge; serving disciplinary or societal ends?' The various articles draw out these tensions. Donovan concludes: '… the more broad, inclusive and democratic the vision of science policy, the more qualitative the appropriate evaluation process; and the more 'scientific' and quality-focused, the greater the need for quantitative methods (ibid. 542).

The perceived inadequacy of the evidence base for decision-making in science policy has generated a demand for a 'science of science policy'

The pressures for accountability and value are now joined by pressure for a theoretical rigour in science policy decision making as articulated Marburger (2005) with his claim that: 'Relating R&D to innovation in any but a general way is a tall order, but not a hopeless one. We need econometric models that encompass enough variables in a sufficient number of countries to produce reasonable simulations of the effect of specific policy choices. This need won't be satisfied by a few grants or workshops, but demands the attention of a specialist scholarly community. As more economists and social scientists turn to these issues, the effectiveness of science policy will grow, and of science advocacy too' (Marburger, 2005: 1087). This advocacy has persuaded the NSF to establish a program of research Science of Science and Innovation Policy that has already gone through three rounds of funding.

Innovation, change and evaluation

Feller (2007: 684), another Blue Sky II participant suggested there was a policy shift toward '… questions about which mechanism(s) directed to supporting which performer(s) in which field(s) of scientific or technological inquiry yield the highest societal returns, however characterized.' The tools to inform such decisions come from rate of return analysis and may not be strictly applicable to investments in basic science with their inherent uncertainty. As a caution he reminds of Cohen's remark (1985: 21) that the '… most brilliant scientists are not able to predict exactly the kind of revolution they themselves will be making.'

Feller concludes that present approaches to evaluation do not adequately address the complexities of the interactions within innovation systems and: '… touch only lightly on how the strategies, behavior, performance of the sectors or actors described in the national innovation taxonomy change as a result of the cumulative, long-term impact of a cluster of programs' (ibid. 687). He might also have been referring to the problem facing the evaluation of development research.

The measurement of scientific research, experimental development and industrial innovation has been codified in the OECD Frascati family of manuals but no equivalent harmonized and codified body of knowledge exists for measuring and evaluating societal change and innovation. Nor does OECD as yet have measures for public sector or social innovation.

This statement highlights a major conceptual difficulty that pertains to the different meanings attached to the terms innovation and change in different spheres of application and interest. The problem arises because of the seemingly precise meaning accorded to innovation as measured through the Community Innovation Surveys and the looser meaning of the word innovation in everyday usage, and its interchangeability with the word 'change.' The differing meanings constitute a gulf of practice across which the practitioners barely communicate. On the one side are statistics bureaus and other parties charged with the responsibility of measuring R&D and (technological) innovation. On the other are those supporting (administrators; funders) and those providing (clinics; schools; extension services) social delivery to the public that ideally meets specific standards. The discourses of the two groups include the term 'innovation' but they understand and measure it differently. For industry innovation is inextricably tied to a change in market share or sales of new or improved products. Educational innovation is less precise - a change in curriculum is supposed to lead to improved learning outcomes that are measured by what?

Though donors have agreed to broad guidelines regarding what attributes should be measured and evaluated there is as yet no standard tool box and they continue to use both metrics and qualitative evaluation. These remain separate but overlapping disciplines with bridges under construction to close the gap between them, an example being work at the Centre for

International Studies of Toronto University[9]. Both measurement and evaluation are dynamic fields and neither may claim to offer a standard way of impact accounting.

[9] http://www.utoronto.ca/cis/

VI. THE ALTERNATIVE METRICS MARKET PLACE

New, large philanthropic organizations are significant funders of development research and bring creativity in their approaches to measurement and evaluation

Conventional evaluation methods continue to dominate the approach of the established donor community dealing with development research. This community has been joined by large and influential new players from the private sector, in the form of philanthropic organizations, and social venture capital entrepreneurs.

Given the US philanthropic tradition, the arrival of these new kids on the block is unsurprising. What is perhaps surprising is their nimbleness of operation and their willingness to experiment. In short they are a source of creative energy in the funding community. Though some are modest in size and locally focused, their experiments are cogent. The new kids on the block include the Irvine Foundation, Bill and Melinda Gates Foundation (also underwritten by the Buffet Foundation), Google.org/Google Foundation, Acumen Fund, the Rockefeller ProVenEx fund and the Skoll Foundation.

The Irvine Foundation is concerned with the wellbeing of Californians rather than development research. The evaluation function is embedded at top level to capture program activity, outcomes, and lessons, using dashboard formats and other qualitative assessments (Irvine, 2008a). The Foundation promotes the sharing of evaluation tools and resources through its web site (Irvine, 2008b) and uniquely among donors Irvine has developed and uses a suite of indicators (Irvine, 2008c) that show similarity with the Kaplan Norton Balanced Scorecard (Kaplan and Norton, 1996).

The Acumen Fund is a social venture fund that is seeking to demonstrate that social value can be generated by venture capital methods: ‘We make financial investments in companies in the developing world who are explicitly targeting the poor as consumers of health, water, housing, and energy products and services … It’s using markets to solve problems of poverty. Our investing thesis is that there are innovative entrepreneurs out there who can find a way to provide these services to the poor in a cost-effective way, and that our "patient" capital can give them the critical financing they need to grow to scale’ (Acumen, 2008).

Acumen believes that metrics are essential management tools to guide and evaluate its investments and is seeking ways to move beyond those that grade only financial returns. Acumen, 2007: 1) notes that 'finding a standard metric to measure the success of a social investment has been a vexing challenge …'

The Fund has developed an approach to estimating social returns known as the BACO ratio – the Best Available Charitable Option. BACO seeks out the intervention that is the nearest equivalent donor or philanthropy funded equivalent project to the activity that Acumen as a hard-nosed investor is funding. It then compares the cost analysis and social impact projections that yield the intervention cost/person year. The ratio of the cost/person year for the Acumen and equivalent project provides the BACO ratio. The worked example in Acumen (2007) concerning the dissemination of insecticide impregnated mosquito bed nets provides a BACO ratio of 52:1. Acumen is open to the limitations of the method: it is experimental, does not capture long-term impact; is sensitive to the choice of equivalent and to the sector that is being addressed – health BACOs are much higher than in low-cost housing. Accordingly BACO is not used as the sole means of judging project impact. But it is without doubt an important item in the alternative metrics toolbox. This is but one of Acumen's contributions.

With the help of engineers from Google, Acumen Fund has commissioned a web application for tracking and reporting this data for its own portfolio - the PDMS (Portfolio Data Management System)[10]. PDMS tracks quantitative data (quarterly data; investment information) and some qualitative information as well (written status reports; documents; survey-style assessments). In November 2007 a Beta test version was shared with some two dozen peer intermediary organizations and fenders (including Google.org, the Skoll Foundation, Hewlett Foundation, and Root Capital, Echo) to assess whether this type of platform would be well received and used by others in this space. Having gathered a critical mass behind the project, Acumen has embarked on creating a platform to provide a user-friendly, globally accessible tool for managing a portfolio of social investments. The Portfolio Data Management System is being used to identify cross-cutting principles that may be used to gain understanding of how to bring successful interventions to the poorest of the poor, those at the bottom of the pyramid.

[10] Mail conversation with Acumen staff

Impact assessment is a work in progress whose boundaries are defined by social entrepreneurs

The longer term vision of the Acumen Fund is to aggregate the data from each investor's portfolio to create standard 'benchmarks' for social impact investing. Work is under way to think through what these 'standard' metrics might be, and how each sub-sector within the social investment space could work with the technology platform to enable this identification and subsequent tracking of common metrics. There have been ongoing conversations to find the right overarching metrics to track every social investment as well as what sector-specific metrics social investors could agree to track (e.g., 'clean energy' investments might have a few metrics tied to power production and carbon offsets; healthcare investments might have a different set of core metrics).

In June 2008 the Beta testers met in New York City. The Sales Force Foundation indicated an interest in working with Acumen to help think through the technology piece of the puzzle, and they joined the meeting to help think through the back-end implementation as a potential partner. The two day meeting was hosted by the Rockefeller Foundation.

The group is in the process of finalizing the funding necessary to enable the development of the next generation of the PDMS system and hope to have a marketable product by January 2009. Another initiative to watch is that of the Anne E Casey Foundation (2008) that is assembling a set of common indicators. These tools and others such as that of GuideStar International hold the promise of opening up the knowledge commons of 'what works in development research and what does not.' This might enable funders to break free of what Bonbright (2007) refers to as the one-to-one relationships between donors and grantseekers to a one-to-many social change marketplace that decides on what investment makes sense.

The distinction between social funding and financial investment by foundations is blurring. In this new context so-called 'double bottom line' investing has been studied by Clark et al (2006) and provides a catalogue of methods for social impact assessment. They conclude that a standard for social impact accounting does not yet exist, and identify at least nine promising methods.

VII. IMPLICATIONS AND CHALLENGES

This paper has provided insights into the meaning, construction and use of metrics for development and development research at the macro, meso and micro levels. At the macro level are the 'usual suspects,' namely review metrics such as HDI and GDP, and synoptic metrics such as the MDGs.

It was argued that externally-funded development research should be located in the framework of country systems of innovation systems, namely the meso level. Such location naturally prompts discussion of the measurement practices applicable to systems of innovation, the way this is evolving and its applicability to both industrialized and developing economies.

The next step was to look at the measurement and evaluation of development research at program and project, or micro level. This led to discussion of funders' approach to change and change management and their evaluation practices. The picture that emerges is that all donors, grantmakers and foundations commit to the importance of evaluation, and some institutionalize this at senior executive level. But despite the general recognition of the value of evaluation, and the willingness of grantees to do it, donors are not funding it adequately or making effective use of results, nor are they investing in a larger infrastructure to support it' (Bonbright, 2007: 31).

Patton (2008) claims that evaluation practice has become professionalized able to offer evaluations that are feasible, ethical, useful, and accurate. He disagrees that randomized control trials (RCT) are the now best way to assess impact, arguing that method must be appropriate to the situation at hand.

Jones and Young (2007) in their comparative review of the major funders of development research found a broad diversity of definitions of research, research themes, and research processes and no evidence of common measurement approaches. This study broadly concurs with their findings insofar as the bilateral and multilateral donors are concerned. Qualitative evaluation methodology remains the method of choice to assess donor interventions making the population of metrics difficult. Among the established bilateral and multilateral donors the use of metrics as a scoreboard for impact assessment is largely notable by its absence. A

case in point is that many funding organizations, other than with respect to the financial, do not to use metrics to assess their own performance.

The main contribution of this paper is to show that the methodological frontier concerning social innovation impact assessment and the management of the associated information is defined by the new kids on the block, the Google Foundation, Skoll Foundation and Acumen Fund, to name but three. Their work on alternative metrics, visualization and portfolio management tools is path breaking.

This path breaking work in the United States is complemented by that of the NESTI community of practice that is developing the measurement framework for the OECD Innovation Strategy. NESTI is driving projects on the measurement of technological and non-technological innovation.

In the immediate term this work does not offer an easy way out for the construction of metrics of development research - the NESTI measurement toolbox is best suited to industrialized countries; it can be used in adapted form in developing countries. As yet it does not speak to the matters of public sector or social innovation. Conventional scientometrics may be applied to development research initiatives subject to the caution that translation of basic research into moves at a different pace to the funding cycles of the donor community.

Within agencies such as DFID[11], NSF, and the European Union pressure appears to be growing for greater attention to be given to the quantitative since such variables are essential inputs for performance modeling of impact assessment. The EU demand for regulatory impact assessment has had a spillover into impact assessment of the Framework Programmes with particular stress on micro-economic modeling. Such modeling is of course dependent upon the availability of the appropriate variables, for which read metrics. The need for better quantitative data is inescapable.

The main conclusions of the paper may summarized as follows:

1. Donors, grantmakers and foundations operate according to their own theory of change and there is little evidence of common practice in the use of evaluation methods let

[11] DFID (2008), NSF (2008), NWO (2007)

alone metrics. They all express commitment to the importance of evaluation, but they fund and institutionalize this to varying degrees.

2. On both sides of the Atlantic pressure is growing for quantitative assessment and modeling. In the US this is found in the call for a Science of Science and Innovation Policy and in Europe for *ex ante* impact assessment of the Framework Program 7.
3. The methodological frontier of social impact assessment is to be found in the approaches of US foundations and social venture capital funds.
4. The methodological frontier for measuring technological and non-technological innovation in firms is defined by the work of NESTI. The Frascati family of manuals for measuring R&D and innovation may travel with some adaptation to developing country contexts. There are no NESTI guidelines for the measurement of public sector or social innovation at present.
5. The various approaches to metrics of research for development – as research, as development, as social change- are potentially incommensurable and suggest a need for a pluralistic approach. A dialogue across the divide of communities of practice is overdue.

A broad observation is that there is no single or simple answer to determining 'value for money' or deciding 'who benefits?' Metrics for development are but one tool for addressing their needs. They are always subject to interpretation in the context they were generated. Pluralistic approaches may be expected to continue, and may be desirable. For IDRC with its mandate of promoting research and research capacity development this suggests that a dual measurement system might be a pragmatic choice. Metrics for (and of) development remain a work in progress.

REFERENCES

Acumen 2007. *The Best Available Charitable Option*. New York: Acumen Fund.

Acumen 2008. http://www.acumenfund.org. Accessed 23 July 2008.

Albuquerque, E. (2003). Immature systems of innovation: Introductory notes about a comparison between South Africa, India, Mexico and Brazil based on Science and Technology Statistics. Discussion paper No. 221. Belo Hirozonte: CEDEPLAR/FACE/IFMG.ASGI-SA.

Andersen, E S, K V Grude and T Haug 2004. *Goal Directed Project Management* London: Kogan Page.

Anne E Casey Foundation 2008. http://www.pathwaystooutcomes.org

Bakewell, O and A Garbutt 2005. *The use and abuse of the logical framework approach*. Stockholm: SIDA .

Bonbright, D 2007. Making social investment decisions. What do we need to know? *Alliance* Volume 12 Number 4 December 2007

Börner, K and A Scharnhorst (Eds.) 2008 Special Issue on Science of Science: Conceptualizations and Models of Science. (2009). *Journal of Informetrics*, Vol. 3 No. 1

Chattaway, Joanna, Fred Gault, Paul Quintas and David Wield 2003. From Digital Divide to Knowledge Divide – A Primer, pp 89-120, in George Sciadas, ed., *Monitoring the Digital Divide…and Beyond*. Montreal: Orbicom

Clark, C, W Rosenzweig, D Long and S Olsen 2004. *Double bottom line project report: assessing social impact in double bottom line ventures methods catalog.* http://www.svtgroup.net/interact/svtresearch.html. Accessed 30 October 2008

Cohen, I. 1985. *Revolution in Science*. Cambridge MA: Harvard University Press.

Commonwealth of Australia 2007. *Public Support for Science and Innovation.* Melbourne: Australian Government Productivity Commission.

Delanghe, H and U Muldur 2007. Ex-ante impact assessment of research programmes: the experience of the European Union's 7th Framework Programme. *Science and Public Policy*, Vol. 34 No. 8 pp169-183.

DFID 2005. *Guidance on Evaluation and Review for DFID Staff.* London: Department for International Development.

DFID 2007. *DFID's 2005 Agricultural Policy: An interim evaluation*. EVSUM 672) London: Department for International Development.

Donovan, C 2007. Introduction: Future pathways for science policy and research assessment: metrics vs peer review, quality vs impact. *Science and Public Policy*, Vol. 34 No. 8 pp538-542.

Earl, S, F Carden and T Smutlyo 2001. *Outcome Mapping*. Ottawa: Iinternational Development Research Centre.

European Union 2002. Communication from the Commission on Impact Assessment. COM(2002) 276 Final. Brussels: Commission of the European Communities.

Feller, I 2007. Mapping the frontiers of evaluation in public sector R&D programs. *Science and Public Policy*, Vol. 34 No. 8 pp 681-690.

Fernandez Polcuch, E 2008. Measuring R&D in developing countries: unsettled issues and impact on the Frascati Manual. Unpublished presentation. NESTI, OECD. June 2008

Ford (2008) http://www.fordfound.org/about/history/overview Accessed 7 June 2008

Freeman, C, J Clark and L G Soete 1982. *Unemployment and Technical Innovation: A study of Long Waves and Economic Development,* London: Pinter.

Gault, F 2007 'Science, Technology and Innovation Indicators: The Context of Change' in *Science, Technology and Innovation Indicators in a Changing World Responding to Policy Needs,* Paris: Organization for Economic Cooperation and Development

Godin, B (2005) *Measurement and Statistics of Science and Technology: 1920 to the Present*. Routledge: London

Government of India 2008. http://dst.gov.in/admin_finance/ls_13/sq278.htm Accessed 5 August 2008

Hovland, I 2007. Working Paper 281. Making a difference: M&E of policy research. London: Overseas Development Institute.

Igami, M and A Saka (2007) Capturing the Evolving Nature of Science, the Development of new Scientific Indicators and the Mapping of Science. Working Paper 2007/1. Paris: Organization for Economic Cooperation and Development.

Irvine 2008a. *Evaluation Policies and Guidelines.* http://www.irvine.org/assets/pdf/evaluation/evaluation_policies.pdf Accessed 20 June 2008

Irvine 2008b. http://www.irvine.org/evaluation/resources.html Accessed 20 June 2008

Irvine 2008c. *Evaluation Policies and Guidelines*. Los Angeles: James Irvine Foundation.

Jones, C I and John C Williams 1997. *Measuring the Social Return to R&D.* Washington: Federal Reserve.

Jones, N and J Young 2007. Situating DFID's Research Funding Policy and Practice in an International Comparative Perspective. Overseas Development Institute paper for DFID; London: Overseas Development Institute.

Kahn, M J and R Swanborough 1999. IT in Public Sector Reform. University of Manchester Institute of Public and Development. www.sed.manchester.ac.uk/idpm/publications/wp/igov/
Kaplan, R S and D P Norton 1996. *The Balanced Scorecard.* Cambridge: Harvard Business School Press.
Kaiserswerth, M. (2007) The future of information technology. Zurich, Switzerland: IBM Research http://www.berno.polemb.net/files/doc/KAISERSWERTH_Swiss_Polish_Dialogue_Nov07_Final.pdf Accessed 30 October 2008
Katz, J S 2000. Scale-independent indicators and research evaluation. *Science and Public Policy,* Vol. 27 No. 1 pp. 23-36.
Malerba, F 2004. *Sectoral systems of innovation.* Cambridge UK: Cambridge University Press
Marburger, John H III 2005. Wanted: Better benchmarks. Science Vol. 308, No. 5725 p1087.
Martin, B. and Salter, A. (1996) The Relationship between publicly funded research and economic performance. Working Paper for HM Treasury, SPRU.
McGregor, J A 2007. Researching wellbeing: concepts to methodology, in Gough, I and McGregor, J A (Eds) *Wellbeing in developing countries: from theory to research 1st edition.* Cambridge UK: Cambridge University Press
Mouton, J and Waast, R 2008. Draft Global Synthesis Report. Study on National Research Systems: A Meta review. http://portal.unesco.org/education/en/files/55636/12000449845Draft_Global_Synthesis_Report.pdf/Draft%2BGlobal%2BSynthesis%2BReport.pdf. Accessed 20 October 2008
Nadiri, M I 1993. Innovations and Technological Spillovers. Working Paper 4423. Cambridge, MA: National Bureau of Economic Research.
National Audit Office 2007. *Big Science: Public Investment in Large Science Facilities*. London: The Stationery Office.
Nelson, Richard R. and Winter, Sydney. G. (1982) *An Evolutionary Model of Economic Change*. Cambridge, MA: Harvard University Press.
NSF 2008. http://www.nsf.gov/pubs/2008/nsf0803/measuring_performance6.pdf Accessed 5 June 2008
NWO 2007. *NWO Annual Report 2006-07.* Den Haag: Nederlands Organizasie voor Wetenskaplike Onderzoek.
OECD (2002) *Proposed Standard Practice for Surveys on Research and Experimental Development.* Paris: Organization for Economic Cooperation and Development.
OECD/Eurostat (2005) *Oslo Manual. Guidelines for collecting and interpreting innovation data. 3rd Edition.* Paris: Organization for Economic Cooperation and Development.
OECD 2007a. Proposal for Assessing the Socio-Economic Benefit of Public R&D. Working Paper DSTI/STP/TIP(2007)8. Paris: Organization for Economic Cooperation and Development.
OECD 2007b. *Science, Technology and Innovation Indicators in a Changing World Responding to Policy Needs,* Paris: Organization for Economic Cooperation and Development
Patton, M Q 1996. *Utilization-Focused Evaluation: The New Century Text*. *3rd Edition* Berkeley: Sage Publications Inc.
Patton, M 2008. State of the Art in Measuring Development Assistance. Address to the World Bank IEG Conference, 10 April 2008, Washington.
RICYT/OEA/CYTED 2001. www.ricyt.org/interior/difusion/pubs/bogota/bogota.pdf. Accessed 20 August 2008
Rockefeller 2007. http://www.rockfound.org/library/annual_reports/2007rf_ar.pdf. Accessed 20 August 2008
Savedoff, W D and R Levine 2006. *Learning from Development: the Case for an International Council to Catalyze Independent Impact Evaluations of Social Sector* Interventions Washington: Centre for Global Development.
Sen, K. and Hoare, K. (2005) *Rates of Return to Research. A Literature Review and Critique.* Reading: Enterplan Limited.
Sida 2004. *The Logical Framework Approach.* Stockholm: Swedish International Development Cooperation Agency.
Sida 2007a. *Strengthening Sida Management for Development Results.* Stockholm: Swedish International Development Cooperation Agency.
Sida 2007b. *Looking Back, Moving Forward. Sida Evaluation Manual 2nd revised edition.* Stockholm: Swedish International Development Cooperation Agency.
Skoll Foundation 2008. http://www.sbs.ox.ac.uk/skoll/forum/2008/dissemination/dissemination.htm
UNDP 1990. *Human Development Report.* Geneva: United Nations Development Programme.

Peer Review

The purpose of this paper is to 'frame a debate on evaluation, measurement related to evaluation, and alternative or more elaborate measurement systems'. As IDRC has a dual mandate, supporting research and building research capacity, the debate to be framed has to take this into account. The paper does this.

A question to be answered is the direction of the accountability, whether it is the Canadian government, the governing Board of IDRC, the countries in which the research and research capacity building are carried out, the community of practice that is being influenced by the IDRC intervention or, some combination of all of these. The assumption of this reviewer is that the principal client for the paper is the Board and that its interest in the recommendations of the paper relates to allocating the resources of IDRC to its project areas and to projects within those areas. The implication is that the debate is not to be about due dilligence and satisfying the Auditor General of Canada that resource allocation conforms to prevailing policy and the law of the land. Rather, it is about the effectiveness of the work of IDRC in supporting research for development and the building of research capacity. However, to hold this debate, the Board must know what IDRC is trying to achieve.

Both in the paper, and the Terms of Reference for the paper, is found the importance of research funders having a 'theory of change' which guides their actions. While 'theory of change' appears in the literature, it seems to be more of a 'view' or a 'vision' of the funding body, not a theory. While this could be dismissed as a semantic quibble, theories are based on fundamental principles and have predictive capacity. A 'view' or a 'vision' need have neither of these. As an example, Margaret Thatcher had a clear view of where the UK should be going when she gained power in 1979, but it was based on ideology, not theory.

If the view of IDRC is that it is important to fund research for development and to fund the building of capacity to do research for development, then that view suggests ways of confirming that what is wanted is being done. That leads to another question, found in the Terms of Reference, about whether the organization knows that it is doing what it set out to do. With a view that HIV/AIDS should be reduced, the logical next step would be to measure the occurrence of HIV in the population over time to observe the success of the intervention. If the view is supporting research for development, there are well established indicators to measure that research is being done, the costs and the outcomes. Promoting capacity building is more difficult to measure as the increased capacity can only be measured if it is used, and then it is a question of the means of measurement; time series of statistical indicators, case studies, peer review, or a mixture. The tools are found in the OECD Frascati family of manuals (OECD 2002), augmented by papers on bibliometrics, and are well covered by the author. However, collections of these tools are limited in their value when faced with complexity of the research system, with rapid change, and with the non-linearity of response to interventions resulting from IDRC support for projects.

The paper does adopt a systems approach and acknowledges the problems of measuring even short term outcomes, and especially longer term impacts. This is an important observation as it leads naturally to the call by John Marburger, the Science Advisor to the U.S. President, to develop a new social science that deals with the 'science' of science policy (Marburger 2007). He calls for new models that are at least as intimidating as those used by economists who provide advice to ministers of finance, but which would inform ministers of education, industry, trade, and human resource development. The models that Marburger wants to see are not necessarily econometric, but more like those proposed by Herbert Simon (1996). In such models, the physical and temporal boundary conditions are included and the activity of interest, constrained by the boundary conditions, is examined through scenario analysis. As an example, a policy objective that required doubling the R&D activity in Canada in 10 years would be shown to fail because of insufficient researchers in the system. This observation could then be followed by a more informed discussion of how to move towards the objective by persuading more students to enter the physical sciences, by encouraging the immigration of already qualified researchers, and by providing incentives for those already in the

system to remain and share their knowledge. This may be a distant objective, but it is one about which the Board might wish to become better informed. The Marburger proposal is being implemented by the U.S. National Science Foundation and there have now been three solicitation for work on the 'Science of Science and Innovation Policy (SciSIP)'. A community of practice is emerging, that involves scholars and policy analysts, in which IDRC could participate.

Following the Terms of Reference, the paper looks at conventional indicators and at the approach of the Development Asssistance Committee (DAC) in the Development Centre Directorate of the OECD. The objective is to look for tensions, but the Frascati family of tools are there to measure activities which may or may not promote development. Additional criteria are imposed by DAC to deal with interventions to promote development effectively and the tools and the criteria are complementary.

In discussing measurement, the author observes that 'a sound theoretical platform for science policy has not as yet emerged'. The same statement would hold if 'science' were deleted. Again, if theory is based on principles and is predictive, a theory of policy is far away. However, policy is a process, driven by a view of what is needed, and that does provide guidance to the development and use of indicators. The author presents some of the findings of the 2006 OECD Blue Sky II Forum (OECD 2007) which acknowledged that in dealing with policy intervention, a systems approach to indicators was a useful way of illustrating the linkages that tie the actors together and that more emphasis should be given to measureing outcomes and, in due course, impacts of the activities of the actors. The OECD Forum emphasized the importance of using indicators to tell a compelling story to the policy community and Marburger reinforced this by his call for intimidating models. This is well captured by the author.

However, there are debates that were not present at the Blue Sky II Forum that appear in the Terms of Reference of the paper, such as the need measure well being, gender empowerment, or S&T capacity. The author introduces a number of indicators and describes attempts to produce and use them. As with the more conventional indicators, the funding organization must have a clear view of what it is trying to achieve through the projects that it supports. The message here is that project proposals should be able to say what is likely to be their outcome and how those outcomes will be measured. This means clear guidelines on how to produce such a proposal and then the capacity to follow up with an independent assessement at some future time. For this to happen, IDRC must have a clear and coherent approach to interventions which move the participants in a desired direction. Considering such a statement is a task for the Board.

Missing from the paper, but easily added, is reference to the community affected by the intervention (Bonbright 2007. The reviewer is grateful to the IDRC commissioning officer for this reference). This goes beyond the holder of an IDRC grant and could include researchers who benefit from the building of research capacity, policy analysts who are better able to develop evidence-based policy as a result of development research, or other development agencies that are able to transfer best practices from IDRC work. Should representatives of such a wider community be consulted as part of developing an IDRC programme?

The paper provides a wealth of informationa and analysis and is a valuable resource. But, to frame a debate, the paper would benefit from greater emphasis on the findings and with the supporting material moved to appendices.

References

Bonbright, David (2007), Making Social Investment Decisions: What Do We Need to Know?, Alliance, Vol. 12, No. 4, December 2007, pp. 29-33.

Marburger, John (2007), The Science of Science and Innovation Policy, in OECD (2007), *Science, Technology and Innovation Indicators in a Changing World: Responding to Policy Needs*, Paris: OECD.

OECD (2002), Frascati Manual: Proposed Standard Practice for Surveys on Research and Experimental Development, Paris: OECD.

OECD (2007), Science, Technology and Innovation Indicators in a Changing World: Responding to Policy Needs, Paris: OECD.

OECD/Eurostat (2005), Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data, 3rd ed., Paris and Luxembourg: OECD/Eurostat.

Simon, Herbert (1996), The Sciences of the Artificial, 3rd edition, Cambridge, MA: The MIT Press.

Terms of Reference: *Metrics of Research for Development*

How do research donors know if they are delivering on their mandate? There are different metrics of success in both science and development. Investing in science can be described in numbers of scientists trained, papers published, patents issued or amount of R&D expenditures. Investing in development can be described in terms of economic growth, livelihoods, human development or happiness. Each of these categories of metrics has strengths and weakness, yet each is also imperfect for describing development research. Research donors are also concerned about getting new knowledge into use –to benefit the lives of poor women and men– and building research capacity –so people can propose their own solutions to their problems.

The purpose of this paper is to frame a debate on measuring research efforts. Development research needs to measure the effect of investments intended to enhance pro-poor innovation in developing regions. Thus, the metrics used by research donors must bridge both science and development. In part, the choice of metrics is informed by how each donor understands change and how they perceive their investments affect world. IDRC seeks to learn more about the variety of approaches used for showing that investments in development research make a difference. This paper will describe the tensions between the metrics used for scientific excellence, development outcomes, and those used to describe the 'knowledge divide'.

- *How do research donors know they are doing what they set out to do?* Survey the metrics used by a range of research donors (i.e. DFID, SAREC, IRD, Ford, Gates, Rockefeller) to measure their performance or report the results of their investments. What is the primary purpose and who is the primary audience of such metrics? Who is responsible for preparing them and how are they used? How do developing countries and southern thinkers perceive these metrics?

- *Are alternative metrics needed?* Describe a range of alternative metrics have been proposed for measuring development and science (i.e. gross happiness, human development indicators, S&T capacity indices, innovation systems, etc.). What weaknesses exist in the metrics currently used by research donors, and how do alternatives intend to address these weaknesses? How do developing countries and southern thinkers perceive them?

- *What is the market for alternative metrics?* Who prepares them, how are they used, and for what purpose? What are the forces driving the development of metrics? Which forums or organizations are most influential in establishing norms for such metrics (i.e. OECD, UNESCO, ICSU, IAC, NEPAD)?

References:

Donovan, C. (2007) Future pathways for science policy and research assessment: metrics vs peer review, quality vs impact. *Science and Public Policy* 34(8):538-542

B. Godin (2005) *Measurement and Statistics on Science and Technology: 1920 to the Present,* London: Routledge

Lepori, B.; R. Barré and G. Filliatreau (2008) New perspectives and challenges for the design and production of S&T indicators. *Research Evaluation* 17(1):33-44

OECD (2002) Frascati Manual – Standard Practice for Surveys of Research; (2007) DAC Statistical Reporting Directives; OECD/Eurostat (2005) Oslo Manual – Guidelines for Innovation Data