A discussion of issues raised at the July 23 conference CAN OPEN DATA BOOST ECONOMIC GROWTH AND PROSPERITY sponsored by the World Bank in Washington, D.C.

On Measuring Open Data Benefits in International Development Projects

Dennis D. McDonald, Ph.D.

On Measuring Open Data Benefits in International Development Projects

By Dennis D. McDonald, PhD.[1]

August 1, 2014

PART I

Back when my daughter was a Peace Corps volunteer she managed a project in a rural Dominican Republic village to convert kitchen floors from dirt to cement.

She raised the money via online networking and fundraising, organized the people and resources to do the work, recruited the homeowners and scheduled the work, then she helped manage the individual projects.

I was reminded of projects like that and the differences in infrastructure between developed and developing countries going into last week's World Bank-sponsored meeting on measuring the economic benefits of "open data" in developing countries. Jason Hare [pictured] and I attended on behalf of BaleFire Global, an open data consultancy he founded.

Having been a World Bank consultant in the past on planning and measuring adoption of collaborative technologies I was already familiar with the Bank's operations and how it plans and manages international development projects.

I admit to being somewhat skeptical regarding the ability to establish a linkage between "open data" -- the definition of which is still evolving -- and development activities like, well, converting dirt floors into cement.

Some of my skepticism was based on having spent almost a decade early in my career on trying to define and measure the linkage between information products and services and the usage and benefits derived from these products and services. It wasn't easy then and, based on what I heard it this World Bank conference, it isn't easy now.

But some things *have* changed since then and for me this conference was a valuable and eye-opening experience. I came away with a better understanding of the role open data can play in international development programs. Plus, I think the Bank is asking the right questions about open data.

PART II

This is how Balefire defines open data in alignment with the international Open Data Institute (ODI) of which they are a member:

> *Open data is information that is available for anyone to use, for any purpose, at no cost.*
>
> *Open data has to have a licence that says it is open data. Without a licence, the data can't be reused. The licence might also say:*
>
> - *that people who use the data must credit whoever is publishing it (this is called attribution)*
>
> - *that people who mix the data with other data have to also release the results as open data (this is called share-alike)*

This is a pretty broad definition and the key is in the first sentence.

It is worth considering how it might be adapted for use by the World Bank which already makes a huge amount of data available. There are many ways to describe data and the uses to which data are put. Attempting to measure the "benefits" of open data in the context of the World Bank's extensive development efforts requires specificity about what we're trying to measure.

Most of the all-day conference was devoted to discussing such distinctions. I found the discussions useful, meaningful, occasionally frustrating, but for the most part enlightening.

Some of what I heard echoed my own learning from several decades as a consultant. At other times I had to concede there are things (a) we still don't understand but (b) we're getting better at it all the time.

PART III

Here are some of the things I found myself nodding my head about.

One. Data doesn't have value. It's how data are used that generates value.

This has always been pretty obvious to me. Keep in mind my academic background is library and information science where, in pre-internet days, you had to make decisions about managing data and systems that helped you to locate information containers (for example, books and documents) that might in turn contain the information you might actually find useful for some application or decision.

Technology has changed since those early days. Storage and transmission costs have plummeted. It's now possible for the same system to help you locate as well as obtain access to documents as well as primary data.

What hasn't changed is that the value derived from using data delivered by the system often occurs *outside* the system in the "real world."

This demarcation between access and actual usage of information has significant measurement and cost implications. No matter how good we are at designing and managing systems for delivering data to users, we're still going to be challenged in linking system-delivered data to the value of its usage.

One reason is that usage may occur removed in time and place from where the information was obtained. One of the valuable things discussed by several conference participants was the importance of looking at open data by "sector" (for example, health, transportation, agriculture, tourism, etc.) Usage of data often varies across sector and can be influenced by availability of complementary or competing services.

Furthermore -- and this is based on my personally having design dozens of information product or service "usage surveys" -- you increase the likelihood of getting reliable and meaningful data if (a) you ask about usage in the context of an actual incident the user can remember and (c) you know how, numerically or statistically, this usage relates to a "total population" of uses.

The cost of doing the types of such analytical efforts may not be trivial, which brings us to the second take-away from last week's conference:

Two. Open data measurement efforts need to be built into development projects right from the start.

To understand the significance of this you need to understand a little about how World Bank development efforts operate. Bank programs are organized by sector, region, and country. Individual development projects may take years to plan as they go through a gated process of design and review that leads to initiation, project management, and eventual conclusion. Date of all types are useful at all stages. Projects can be viewed both as data production and data consumption engines with the processes and systems involved becoming more open and accessible over time; more "openness" is a stated goal of World Bank management.

Given the size and complexity variations of World Bank development projects -- some projects take many years from start to finish -- waiting till a project is underway is problematic for deciding on which data associated with the project can be made open and accessible. In many cases the "boots on the ground" in development projects are local contractors, individuals and companies hired to perform services of various kinds in conjunction with local government organizations.

Money and expertise that flow into a country for project work can be tracked through the standard World Bank management information infrastructure. How data generated by the projects are used may not be automatically captured. Building data-usage-tracking metrics into a project ideally requires involvement right from the start with one likely goal being to minimize the additional cost associated with measurement. After all, the less spent on management and administration, the more you have for the actual development work, right? Which brings us to point number three:

Three. Costs associated with open data projects are dropping.

I heard statements like this several times at the World Bank and each time I had to grit my teeth. "Isn't this true?" I hear you say, given cloud services, cheap Internet, the popularity of mobile services, etc. etc.?

Well, yes and no. Certainly, technology associated costs have plummeted. For example, early in my career I worked on a National Science Foundation project in Egypt that had to *lease a satellite line* to run bibliographic data searches of technical and scientific information in the U.S. from Cairo. Nowadays such communication functionality is *almost* free.

The other fact of the Internet age that we have all come to expect is that at some point we can shift some of our costs on to someone else either by shifting certain tasks onto users or by taking advantage of "free" services that are government or advertiser subsidized.

This is where it makes sense to bear in mind that the "business model" for sustaining an open data effort in a developing country may not be the same as in a developed country. For example, the capacity for a country's development sectors to make something of a data resource needs to be taken into account. Are the technical and analytical skills available to take World Bank supplied data and from it create a data "product" people are willing to pay for? Are ancillary data sources that can be combined with World Bank data available to create new or unique products? Would any jobs created directly or indirectly from open data efforts be limited to an elite group of traditional "haves" or will they be distributed more broadly to the "have nots"?

Four. Is it possible to predict the impact of how open data are used?

During our discussions around economic concepts applied to open data I raised my hand and asked the question, "Aren't we really just asking market research questions? If you're planning a new product or service isn't it just a natural thing to do that you research and test the market for your product or services before you spend significant time and money on development?"

I still think the analogy with market research makes sense. That's something I've done and I understand. But on reflection I don't think my question really takes into account a couple of important factors related to open data.

The first is that when researching information products and services **you really need to get an example of the product into the hands (or in front of the eyeballs) of your prospective customers** in order to get useful feedback on its usage and value.

One of the advantages of web delivered products is that the cost of doing updates and measuring responses to different variations has been simplified compared with physical products. That doesn't change the fact though that you still need to deliver products and services that are useful and you need to create or build in a mechanism for gaining feedback. If the feedback mechanism can be built into the

delivery mechanism, so much the better; that's one of the reasons you want to be involved in a project right from the start.

The second factor to be considered in predicting the impact of open data is that **impacts can't always be predicted or controlled**. Simply making raw data accessible about something in the economy with the "hope" this availability will encourage innovative and beneficial uses and API development to evolve is a high risk proposition unless you have already done your homework.

I'm not saying that wonderful and unanticipated products or services won't emerge from making raw data available; maybe they will. But *assuming* they will and using that assumption as the basis for a risky investment decision, is problematic.

I'm all for innovation and giving people a chance. Still, there is one of the reasons I have supported segmenting program costs and benefits among, at minimum, a program's internal and external users. There's no reason to limit uses of an "open data" program to external users, especially in situations – as is the case with the World Bank – a significant number of projects staff at the country level are actually contractors or consultants without a long-term affiliation with the Bank.

In the real world of development and distributed enterprise operations traditional notions of internal and external systems and users can be difficult to maintain. Measuring and promoting use of open data among staff members makes a lot of sense, especially if the Bank wants to break down the organizational silos that exist among different groups. Removing barriers to data sharing internally and externally might be one way to do that.

Five. What business is the World Bank in?

A lot of what I heard at the conference were the types of questions an information product developer might want to ask before making a development decision. Just as we need to ask questions about the skills and resources needed in the development project's target market to take advantage of open data, we also need to ask if the World Bank itself has the wherewithal among its own staff to address information product development questions.

I am certainly encouraged by the level of staff involvement in sophisticated discussions during the conference's morning and afternoon sessions. It's also clear that the younger staff members with "hands-on" development project experience "get it" when it comes to open data.

Still, I'm wondering if it might also make sense to consider what kind of "subject matter expertise" might be needed at the World Bank to ensure that open data opportunities, risks, and requirements are adequately addressed throughout the development project lifecycle (perhaps this has already been done).

Asking "What kind of business is the World Bank in?" is a reasonable thing to do. The Bank's development projects address a wide range of societal needs and industry sectors. The Bank over the years has developed expertise and contacts worldwide in all the development projects types it addresses. Raising data to the level of a development tool, though, suggests the need to assess the types of "data literacy" needed by bank staff. In [Management Needs Data Literacy To Run Open Data Programs](#) I outlined at a high-level what management "needs to know" when it comes to making open

data concepts and practices an integral and effective part of an organization's programs; perhaps some of these same concepts can be applied to the World Bank?

PART IV

I'm a project manager. Most of the projects I've managed or have been a part of have had various types of data as a key component either as a deliverable or as part of the management process.

What this experience has taught me is that, no matter what kind of project you're managing, you need to be able to look at the project from a variety of perspectives including the numbers describing the project's cost, the perspectives of the different people and stakeholders, the risks involved, and how users will react to the products you deliver.

Becoming more "data-driven" in how projects are managed and evaluated is a laudable goal as this implies greater objectivity and accuracy. That's good.

At the same time, the numbers have to *mean* something. That's where I come back to the kitchen floor projects my daughter managed in the Peace Corps. Ultimately we want to make sure that opening up the data associated with development projects actually does advance the goals of those projects.

At the end of the day are more people being employed productively? Are projects being run more efficiently? Are people finding new, innovative, and useful applications of the data? And -- perhaps most important -- are more concrete floors being delivered?