Peter Burgess LinkedIn Blogs

The High Cost of Data

There is some concern that data about socio-economic progress and performance is very costly and I share that concern. However, in my view the reason that these data are costly is that the approach to the data and the analysis has been uninformed and inappropriate.

I have argued for years that the best management information is the least data that enables a correct decision to be made, and to track the results of the decision. The operative word is 'least'. Another aspect of data that reduces its cost is to collect the data once and use it many times.

Yet another aspect of data is to recognize that some data are fast moving, and some data change very slowly. For this latter point, in accounting, the balance sheet accounts change rather slowly, while the operating accounts or the transaction accounts are high volume and change rapidly.

None of these fundamental concepts seem to inform the current generation of experts in data processing.

Bjorn Lomborg has recently written a piece that was published by the Guardian.com alerting the reader to the incredible cost of gathering data to manage the new sustainable development goals. http://www.theguardian.com/global-development/poverty-matters/2014/sep/24/gathering-data-sustainable-development-crippling

Why am I not surprised. This is the way things have been going for a very long time. Even though the power of technology to improve the efficiency of data processing has increased, the actual efficiency of data systems has gone down and costs have gone up.

In the 1960s I was involved with the installation of mainframe computers. The hardware was massive and expensive, and the software was customized for every installation at another huge cost. Even so, when these systems were operated well, an organization was able to reduce its costs and increase it efficiency. By the 1980s Moore's Law had cut in, and the cost of computing was a fraction of what it had been in the 1960s. Around 1984 I bought an early version of Lotus123, a new spreadsheet program that would run on a PC (personal computer). It cost me around $500 and I remember telling a colleague that I was now effectively a millionaire. 20 years before, software with this performance would have cost a million dollars to program and I was able to own it for just $500. (Before Lotus 123 I had used Visicalc and MultiPlan and later used QuatroPro way before Excel came out!)

While it cost Lotus (later bought by IBM) several million dollars to create, each incremental unit they sold cost them not much more than one dollar … and I was able to buy one for $500 … a huge incremental profit for Lotus as each additional unit was sold. With competition, over time the price came down to around say $200 … while at the same time the incremental cost came down as well to almost zero (with Internet downloads).

This should have been very good news for society and the economy … but a new business model cut in and a few powerful business units became very successful, with society in general paying the bills and getting not so much of the benefit.

Fast forward another 30 years to now (2014) and the cost of doing computer calculation is tiny … yet the cost of data and information are enormous. From the perspective of the common man, society and the global economy are in a shambles and there seems to be no light at the end of the tunnel. What on earth is going on?

Maybe there is something fundamentally wrong with the whole institutional structure that is educating people and determining the policy framework for society and the economy and the methods used for data capture and data processing. Why are data and information costing so much when data processing … the computations … cost so little.

My conclusion is that those with some power and influence in this arena have got it all wrong and really don't understand what is possible. If the best management information is the 'least' amount of information that will enable good decisions to be made and then to track the subsequent performance effectively, then we certainly do not have an institutional ecosystem that embraces this idea. We have data processes that seem to see data and analysis as an end in itself (which it may be in some scientific situations) rather than merely being a component of a management system and specifically a feedback loop that helps to optimize performance. Data need to be part of a system that makes it possible to have feedback and corrective action in a timely way. The faster the feedback the better the performance is going to be.

Most of the data processes in place today and available for the public and for decision making for society have none of these characteristics. Most are ponderous and slow … and the data being used for processing badly designed, if designed at all. Actually doing anything that is meaningful based on these data is not easy and is probably impossible.

To much, decision making is based on some wishy-washy hope that there will favorable movement of the progress and performance indicators sometime in the future. A better data system would be one where there are good data to help make decisions, and then data that shows whether or not the decisions are getting good results … and subsequently use of the same data system to show what the cumulative impact of the decisions has been. Data should help to figure out 'cause' and 'effect' rather than merely being some sort of correlation, with or without causality.

The concerns being expressed (Bjorn Lomborg above) that getting data for the new UN Sustainable Development Goals is going to cost too much shows a complete misunderstanding of how data systems can work. The methodology still used are ones that first emerged in the 19th century, about 200 years before computers became possible. Why are the data systems not being rethought for the modern era. The data for decision making and the data for assessing performance should be essentially the same. If they are not, then there is something very wrong with the whole process. The data for summary reports and the data for detailed decision making should be part of a coherent architecture. There should be an understanding of 'state' and 'flow' as there is in accountancy but not found in a meaningful way where it would be useful in the analysis of social performance.

It is a great concern that there there is something very wrong with the whole process.

Maybe it is that too many people with powerful positions and important portfolios are fearful of data that will show in an independent and objective manner the actual performance of their portfolios. The same goes for big and powerful institutions … in fact almost all institutions, large and small. Fund flows into small independent NGOs are dependent on the story about their performance they tell their donors, which might well be disrupted if there are objective and independent data about the actual performance and impact.

Among the powerful there is no constituency for really good data that will enable accountability, and among common people who would like to see good data and accountability, there is no money to fund what is required. At least, this is the conventional wisdom.

I see light at the end of the tunnel because there is the possibility for the Internet to change everything. The stranglehold of the 'establishment' on the data about socio-economic performance and environmental performance in general, and the performance of specific actors is breaking down, and bit by bit a framework for accountability is emerging. It is all a bit chaotic at the moment, but bit by bit and byte by byte these data are being organized and maybe perhaps in the relatively near future there will be data that cost little and have a huge impact.

Peter Burgess - TrueValueMetrics - Multi Dimension Impact Accounting

---------- /// ----------

Contact information for Peter Burgess: Founder / CEO ... TrueValueMetrics

Website: http://truevaluemetrics,org
Email: peterbnyc@gmail.com
Skype: peterburgessnyc
Twitter: @truevaluemetric
LinkedIn for Peter Burgess : www.linkedin.com/in/peterburgess1/
LinkedIn for this blog: https://www.linkedin.com/pulse/article/20141013183253-27861-the-high-cost-of-data