Global Partnership
for Sustainable Development Data

What Do We Know About the Value of Data?

Author
Jenna Slotin, Senior Director of Policy, Global Partnership for Sustainable Development Data

Acknowledgements
This paper was produced as an input to More and Better Financing for Data, an initiative coordinated by the Global Partnership for Sustainable Development Data (GPSDD). We are grateful for the technical and advisory support of Open Data Watch and the Value of Data Inventory they produced. The initiative is supported by an expert group of GPSDD partners, including PARIS21, World Bank, SDSN TReNDS, AidData, Development Gateway, Sustainable Development Policy Institute (Pakistan), Local Development Research Institute (Kenya), Centre of Excellence for CRVS at IDRC, and DFID. Any questions should be directed to **info@data4sdgs.org**.

IF DEVELOPMENT DATA IS SO POWERFUL, WHY DOES NO ONE WANT TO PAY FOR IT?

When it comes to the development data needed to support the 2030 Agenda for Sustainable Development, there is a deep disconnect between what policymakers say and what they do. There is strong rhetorical consensus that better data are a prerequisite to achieving the Sustainable Development Goals (SDGs) and leaving no one behind, but the demand for better data has not translated into corresponding growth in funding. The excitement around new data sources and disruptive technology has not been matched by support for the national systems, capacities, and infrastructure that are needed to take advantage of new data and technology, particularly in low- and middle-income countries.

One reason for this is that data and statistics are often de-prioritized in budgetary decisions. They are typically viewed as long-term investments in processes and systems in contrast to competing priorities that are viewed as having a more immediate impact on outcomes. This speaks to policymakers' and politicians' short-term horizons combined with impatience to see some kind of result, even when that result may not lead to systemic change.

By the latest numbers, 0.3% of official development assistance (ODA) is spent on data and statistics (Paris21, 2017). Does this paltry resourcing suggest that contrary to public discourse, policymakers do not genuinely value data? Or, more likely, that they do not know how to assess its value relative to other investment priorities?

To confront this challenge, this paper seeks to better understand what we know about the value of data. How have economists and other researchers attempted to measure the value of data and how well do they do at painting the picture of what data are really worth?

This paper builds on the great work of the team at Open Data Watch who recently produced a Value of Data Inventory[1] comprised of articles and reports on the topic. The inventory shows the different ways that analysts and thinkers across disciplines are attempting to conceptualize and measure the value of data. It is striking that among these diverse perspectives, each author is grappling in their own way with the implications of data as a new economic asset, and yet there appears to be little consensus on how best to measure its value. One thing they can agree on is that measuring the value of data – and making the case for investing in data – is very difficult.

Drawing on the inventory, this paper unpacks the different methods for measuring the value of data that we found when scanning the literature. It aims to understand the benefits and drawbacks of these methods, and to determine which approaches would be most effective to convince policymakers of the need to increase investments in data.

[1] Accessible at: goo.gl/hdcVLF

LANDSCAPE OF METHODS

An important starting point for all of these methods is the recognition that data are an intangible asset with several unique features that make measuring their value extremely difficult. Data are, in large part, a public good and therefore do not have a market price. Data are non-rival, meaning one person's use does not take away from others' ability to use it or diminish its value. This means it can be difficult to track and measure the value to different users over time and for different purposes. These features make it particularly difficult to pinpoint value and to monetize the benefits of particular data products. Nevertheless, we identified five approaches that have been used to measure the value of data.

1 COST-BASED APPROACHES (UNECE, 2017; Klein 2016)

Value is determined based on the full cost to produce data and statistics either through pure cash recording annually or the full accruals method where costs are allocated as much as possible to the years in which they add to the production of statistics so that major expenditures (e.g. census) are smoothed over time. These costs have been employed to make the case that data and statistics investments are relatively inexpensive by looking at them relative to population, GDP, government budgets, or ODA.

This is a valuable exercise in and of itself, but it provides little indication of the benefits in terms of whether the investment (no matter how small) was worthwhile.

2 MARKET-BASED APPROACHES

Value is determined based on the market price of equivalent products or users' willingness to pay.

Market equivalent pricing

(UNECE, 2017; Klein 2016)

This is the market price that should have been received if the statistical or data outputs had been sold in a market environment. This approach approximates market prices by looking at the market prices of similar products, such as those from companies that offer data on prices, or business trends drawn from a range of sources including open government data.

The drawback of this method is that it relies on the similarity between the products produced by the public sector and those provided on the information market. Such a comparison is problematic because companies often add value to statistical products by bundling them with analytical services and tools and by exploiting niches that public sector information cannot fill in terms of timeliness. This method is also limited by the price consumers are willing to pay and may underestimate the value of publicly (and often freely) provided data and statistics for public policy and research.

Stated/revealed preferences

(UNECE, 2017; Klein, 2016; Ballivian and Maret, 2015)

Stated preference methods use a survey to determine the amount users are willing to pay or the amount they are willing to accept for not having the data. This is a relatively straightforward and flexible method, and it is considered relatively easy to explain and analyze. However, it is difficult to validate externally and relies on the assumption that the users being surveyed really understand the product in question and have a good basis for articulating its value.

Revealed preference methods aim to understand preferences for attributes when choosing among a bundle of goods or services when their choices are constrained by their available budget. This approach examines the trade-off between choices and their associated costs from observed real world behavior. An example of revealed preferences used by the Mexican and Spanish statistical offices assesses media coverage of official statistics by examining the trade-off made by publishers when deciding whether to place statistical content or an advertisement. Statistical content is assumed to be at least as valuable to readers as the revenue that would have been generated by an advertisement.

While revealed preference methods are widely used in cost-benefit analysis and may be viewed as quite credible, they are limited by their focus on specific products and specific users. For example, the media valuation approach used by Mexico and Spain highlights the potential value of statistical content to the general public but not to policymakers.

3 INCOME-BASED APPROACHES

Value is defined by estimating the productivity improvements and future cash flows that can be derived from the data. This approach has been applied most commonly to assess the fiscal benefits of open government data.

For example, a 2014 U.S. Department of Commerce study cited by Ballivian and Maret (2015) measured the commercial value of government data by measuring the size of businesses in the data industry in the U.S. The study identifies all businesses for which government data is an important input to their business model and then estimates their revenues. The Department of Commerce found that since 2004, data-driven businesses created a $17-trillion dollar economy, according to estimates of their revenues. The expenditure on data collection, processing, and dissemination amounted to $3.7 billion annually – 0.02% of the value created.

In 2013, a McKinsey study found that open data can help unlock $3.2 trillion to $5.4 trillion in economic value per year across seven “domains” of the economy – education, transportation, consumer products, electricity, oil and gas, healthcare, and consumer finance. Their analysis identifies five levers through which the value of open data is realized:

- Creating transparency to unearth information to make better personal, business, and government decisions;
- Exposing variability and enabling experimentation to identify areas for improvement;
- Segmenting populations to tailor actions such as creating custom marketing offers;
- Augmenting or automating human decision-making;
- Defining new products, services, and business models.

McKinsey applies these levers across the seven domains of the economy and makes a range of assumptions in each case around the potential economic impacts of open data.

A 2014 study by Lateral Economics reviewed several quantitative studies of the value of government data. They highlight the difficulties inherent in these methods and the wide range of estimates arising from different approaches. Broadly, they argue that top-down macro-economic methods tend to over-estimate the value of open data because they ignore the possible substitutes that could be used. Bottom-up approaches built from micro-economic analysis tend toward under-estimation because they miss a range of positive, often serendipitous, impacts. They also point out that economic structures vary from country to country making direct conversions difficult. Nevertheless, building on this review, they estimate the aggregate direct and indirect value of open data in the G20 countries to be between USD 700 billion and USD 950 billion per annum.

This is a seductive stream in the literature because the studies estimate high-value impacts and point towards the strong economic, social, and governance benefits from open data. However, as the Lateral Economics study notes, the methods vary widely and build-in numerous different assumptions that are difficult to verify and unpack. Further, the studies have predominantly been conducted in developed economies where the quality and reliability of public sector data are not in question and the information market is better developed than in low-income countries.

4 BENEFIT MONETIZATION

Value is estimated by defining the benefits of particular data products, such as a census, and then monetizing the benefits.

To determine the value of the New Zealand census, Bakker (2014) estimated reasonable ranges of potential impacts in key benefit areas, including resource allocation; capital investment planning; policymaking and monitoring; service planning; academic and market research; statistical benchmarks; and, electoral boundaries and representation.

The study notes that there are many other direct and indirect benefits that could not be quantified leading to an inevitable underestimation of the value.

The study identified the expected improvements to accuracy derived as a result of census related information being available. This involved establishing the counterfactual by determining the likely range of alternate data sources that could be used, and the impacts on accuracy from relying on the counterfactual data source. Following this approach, Bakker estimated a net present value of close to $1 billion for the benefits to New Zealand from the use of census and population statistics using the most generally applicable discount rate of 8%, or a net return of about five dollars for every dollar spent.

The UK Office of National Statistics (ONS) conducted a valuation of the benefits of the 2011 census as part of building a business case for that investment beginning in 2006. The business case describes the value of the census across six main uses: resource allocation; targeting investment; service planning; policymaking and monitoring; academic and market research, and as a benchmark for other national statistics. However, only a sub-set of these benefits were quantified. In 2013, the ONS redid the benefits calculations based on the known uses of the 2011 census and determined that the benefits were £490m per year. The 2011 census cost £482m in cash prices giving a payback period of just over a year.

5 IMPACT-BASED APPROACHES

Value is determined by assessing the causal effect of data availability on economic and social outcomes, or the costs in terms of inefficiencies or poor policy decisions due to limited or poor-quality data. There are many case studies showing these impacts, from deliberate experiments such as randomized control trials to retrospective assessments of impact. Below are a few examples, but there are many more.[2]

- One frequently cited study (UNECE 2017; Klein 2016) builds from a rare policy change that resulted in the provision of school statistics in England but not in Wales. Burgess et al (2013) found a significant and sizeable negative effect on pupil progress in Wales compared to England following the policy change in 2001. Based on the results of this research, other analysis estimates that every £1 invested in the examination system and the subsequent production and dissemination of school league tables results in academic improvements equivalent to a £16 increase in GPD. Another approach found cost savings of £19 compared to cutting class size to achieve equivalent improvements.

[2] For example, see www.dataimpacts.org

- Assessing the effects of the IMF's data standards initiatives, Cady and Pellechio (2006) found that participation in these initiatives can decrease sovereign borrowing costs for countries. Discounts amount to about 8% for General Data Dissemination System (GDDS) participants and 20% for Special Data Dissemination Standard (SDDS) subscribers.

The impact-based approaches show the strong role of storytelling in promoting value. Ultimately, demonstrating the relationship between data investments and outcomes that affect people's lives is what resonates best. Their most significant drawback is context specificity, which can limit their influence. However, strong illustrative cases that policymakers can relate to on a human level may still be most influential. If it is possible to show both the human impact and the return on investment, we may have a winning combination.

Across these five approaches, we found few attempts to measure the value of data in low- and middle-income settings. The most common are impact-based case studies where social or economic returns have been documented after specific investments in data.

Whether applied in high or low-income settings, the methodologies are complex and can be quite theoretical making them potentially inaccessible and less compelling for policymakers. However, some methods like the income-based approaches have the advantage of providing headline numbers backed up by a methodology, which may be convincing to some actors. In the case of impact-based approaches, the examples and findings are clearer and more straightforward to grasp and the benefits for data users are clear.

By linking to real-life outcomes, impact-based case studies show how investments in data systems can translate into meaningful outcomes for people.

CONCLUSION

Getting a good measure of the value of data is not at all straightforward. In popular and academic literature, many thinkers are grappling with the role that data plays in our economies and societies, and we do not yet have a consensus on how best to measure its value. Our landscaping found five approaches to valuation, all of which were found wanting in some way. The most promising appears to be the impact-based approaches, where case studies are used to show the concrete impacts of data on improving people's lives, or how a failure to invest in data holds back progress.

The impact-based approaches are easier to understand and communicate. By linking to real-life outcomes, they show how investments in systems can translate into meaningful impacts. They can also show that while investments in systems may take longer to bear fruit, they are foundational to the realization of so many other goals.

Going forward, the Global Partnership for Sustainable Development Data will support efforts to compile stories of data impact that make a clear link to return on investment. We will build on the work of our partners and encourage anyone to contact us with stories they would like to share. We will package these stories and use them in our advocacy as we make the case for more and better financing for data.

REFERENCES

ACIL Tasman (2009) Spatial Information in the New Zealand Economy. Available at: https://www.linz.govt.nz/about-linz/publications/spatial-information-new-zealand-economy

Australian Government Productivity Commission (2017). Data Availability and Use, Inquiry Report. Available at: https://www.pc.gov.au/inquiries/completed/data-access#report

Bakker, C (2014). Valuing the census. Available at: https://drive.google.com/open?id=1soaAxnIpzD3zGiwqm8SAquMtEnK3oL3y (Available in our google folder because the report is no longer available on the Stats NZ site.)

Ballivian, Amparo and Fenohasina Rakotondrazaka Maret (2015) Measuring the Value of Data. Available at: https://statswiki.unece.org/download/attachments/117772954/World%20Bank_Ballivian_Maret_MeasuringtheValueofData_20151202.pdf?version=1&modificationDate=1473158675433&api=v2

Burgess, S. et al. (2013) "A natural experiment in school accountability: The impact of school performance information on pupil progress." Journal of Public Economics, Vol. 106, pp. 57-67. Available at https://doi.org/10.1016/j.jpubeco.2013.06.005

Cady, John and Anthony Pellechio (2006) "Sovereign Borrowing Cost and the IMF's Data Standards Initiatives", IMF Working Paper WP/06/78, IMF. Available at: https://www.imf.org/external/pubs/ft/wp/2006/wp0678.pdf

Deloitte (2013) Market Assessment of Public Sector Information. Available at: https://www.gov.uk/government/publications/public-sector-information-market-assessment

Klein, Thilo (2016) Monetizing the Value of Official Statistics: A case study. Available at: http://www.ksh.hu/cess2016/prezi/cess2016_a3_thilo_klein.pdf

Lateral Economics (2014) Open for Business: How Open Data Can Help Achieve the G20 Growth Target, available at: https://www.omidyar.com/sites/default/files/file_archive/insights/ON%20Report_061114_FNL.pdf

Manyika, James et al. (2013) Open Data: Unlocking innovation and performance with liquid information, McKinsey Global Institute, McKinsey Center for Government, McKinsey Business Technology Office, available at: https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information

Philips, et al. (2015) Are Well Functioning Civil Registration and Vital Statistics Systems Associated with Better Health Outcomes? Lancet; 386: 1386-94. Available at: http://dx.doi.org/10.1016/S0140-6736(15)60172-6

UK Office for National Statistics. The Value of Census Statistics. Available at: https://www.ukdataservice.ac.uk/media/455474/cope.pdf

UN-ECE Task Force on the Value of Official Statistics (2017) Value of official statistics: Recommendations on promoting, measuring and communicating the value of official statistics. Available at: https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2017/CES_4-Value_of_Official_Statistics_for_endorsement_for_upload.pdf

Global Partnership
for Sustainable Development Data

www.data4sdgs.org