The Data Value Chain:
Moving from Production to Impact

Prepared for Data2X by Open Data Watch

> The data value chain describes the evolution of data from collection to analysis, dissemination, and the final impact of data on decision making. While the value chain can be applied to all types of data, we hope it will be particularly useful for a better understanding of the gaps in gender data.

Data are the information we use as the basis for reasoning, analysis, and debate. They are the factual currency for evidence-based policy making. In a data-driven utopia, data would be highly valued and demanded and used ethically and effectively. But data travel a long journey, gaining value as they go, before they achieve their highest purpose. The data value chain provides a framework through which to visualize the life cycle of data, from defining a need to using them for impact.

The concept of a value chain was first used by John Porter to describe how a firm receives raw materials, turns them into products, distributes, markets and sells them, and subsequently provides services, adding value and hence, increasing the firm's revenues with each step.[1] Value chain analysis has since been applied to sequential activities at a larger scale: a global value chain has been used to characterize the movement of goods in various stages of production between firms and across national boundaries.[2]

The value chain describes connections between each step that change low-value inputs into high-value outputs. Although it has a logical flow, from start to finish, a value chain has no theory: it is a pragmatic construct. Value chain analysis can help to identify impediments toward a final goal. It can also focus attention on high-value steps where more effort is needed or suggest ways to reduce resources spent on low-value activities. The model of the value chain is just as applicable to the production and use of intangible goods, such as data and statistics, as it is to physical products.

The data value chain has been discussed in the context of big data and the private sector, where data sources are discovered, ingested, processed, stored, analyzed, and ultimately exploited by the organization to add value.[3] The falling cost of electronic storage and the exponential growth of data have been well documented, but what has spurred the data revolution is not the volume of data but the recognition that the data have valuable uses. Finding high-value uses and creating a process to transform raw data into actionable information is the essence of the data value chain. In an increasingly data-driven world, data have been described as the new oil.[4] Ultimately, when data are put to use, they have an impact: a decision is made, a condition is altered, and someone's well-being is affected.

In this paper, we describe a value chain for data that are usually described as official statistics, produced by governments or other public agencies. At each stage we include examples of important, value-adding activities. The examples come from desk research on the potential impact of gender data.

The production and interpretation of statistics derived from large datasets is a complex process. For the data user at any stage of the value chain, there must be confidence – trust – that the data are fit for their intended purpose. This is an example of the classic principal-agent problem, in which the agent – the data producer – has information that the principal – the user – needs to assess the value of the data. Another way to describe this problem is "asymmetric information." Data quality assessment frameworks help to overcome information asymmetries by providing metadata about each stage of the data production process.[5] Trust in data and the interpretation of data are essential at each step of the data value chain.

1 Porter, Michael E. (1985). Competitive Advantage: Creating and Sustaining Superior Performance. New York: Simon and Schuster.

2 Gurría, Angel. (2012). The Emergence of Global Value Chains: What Do They Mean for Business. G20 Trade and Investment Promotion Summit. Mexico City: OECD.

3 Dumbill, Edd. (2014). "Understanding the Data Value Chain." IBM Big Data & Analytics Hub.

4 The Economist. (2017). "The world's most valuable resource is no longer oil, but data." The Economist.

5 See, for example, the International Monetary Fund's generic framework: http://dsbb.imf.org/images/pdfs/dqrs_factsheet.pdf.

Without trust, there can be no value. Therefore, producers must be as transparent as possible about their work, adhering to appropriate methodologies and making metadata readily available. Users, for their part, must undertake to understand the data and utilize data in a trustworthy manner. Thus, building trust in data is a collaborative activity of producers and users.

In the private and for-profit sector, the value of data can be measured by their impact on the bottom line: do more data or more data analysis or better use of data lead to greater revenues or lower costs or both? In the public sector, there is often no discernible bottom line. Although there is continual pressure to reduce the costs of data collection, processing, and storage, it is harder to measure the returns to investments in data. As public goods, data can and should be freely available and widely used both for public and private benefit, but it may be hard to capture those values by conventional methods of cost-benefit analysis. That is why impact stories are so important: they help to answer the question, "What use are all these data?"

In the past, there has been a tendency to focus too narrowly on the collection and production side of data under the assumption that whatever is produced will be used. Data dissemination still requires attention. Hans Rosling used to refer to "data huggers," people or organizations that wanted to hold onto data and prevent others from using them. But data are inherently a public good.[6] Once they have been produced they can be used and reused at little cost, generating value each time. And so, our goal should be to use and reuse data to their maximum effect. However, reports such as AidData's *Avoiding Data Graveyards*[7] and Development Gateway's *Results Data Initiative*[8], show that a "build it and they will come" mentality is obsolete in today's data age. Collecting and publishing data alone does not ensure they will be used or lead to positive impacts. More attention is needed on the uptake and use of data, which the data value chain unpacks and illustrates.

Although we focus on impact as the final stage of the value chain, we should keep in mind that data are often used to monitor progress along the value chains of other processes. Some data humbly measure inputs: How many children should be in school? How many teachers and classrooms are available? Others measure outputs: How many students progressed from the primary to the secondary stage? Others measure outcomes: How much did the students learn? And finally, some measure headline impacts: How did education change peoples' lives? Data are of value at every stage of this process. And the transformation of raw data into actionable information is just as important at the input stage as it is when measuring outcomes and impacts, such as career choices and incomes.
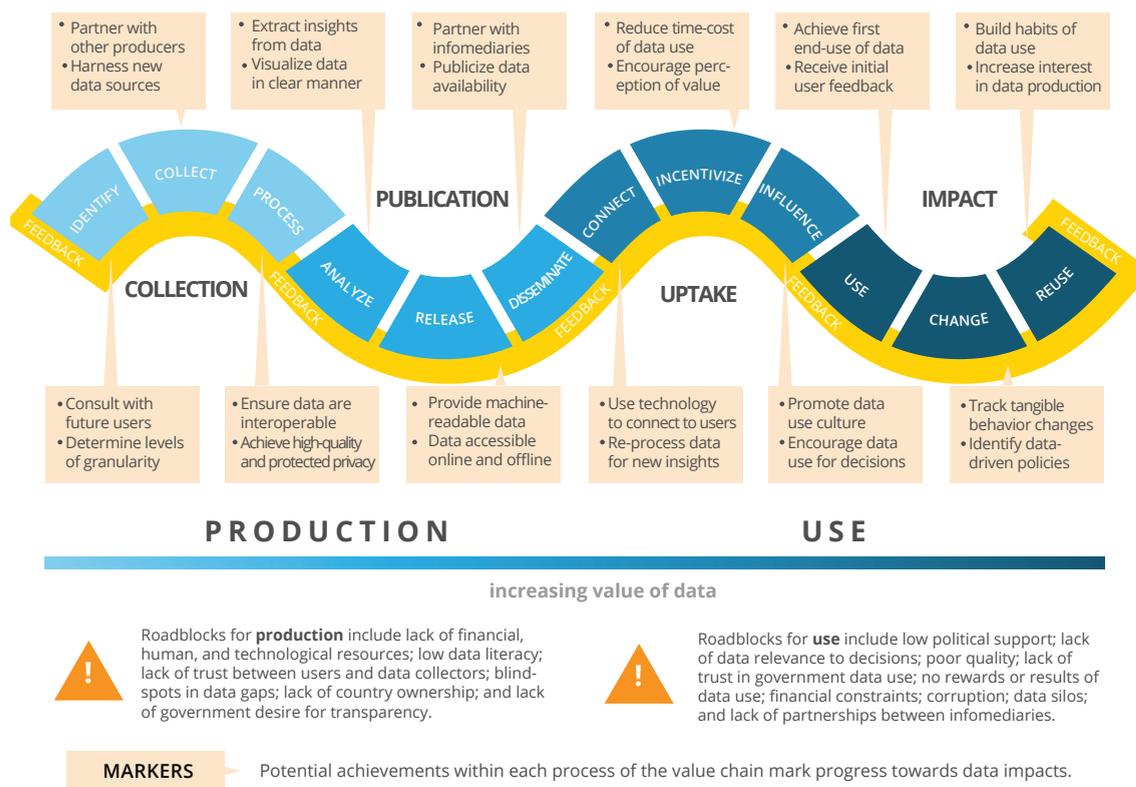
The data value chain can be used to analyze the process that data undergo from first identifying a need for the data to using them for impact and change. The data value chain has four major stages: collection, publication, uptake, and impact. These four stages are further separated into twelve steps: identify, collect, process, analyze, release, disseminate, connect, incentivize, influence, use, change, and reuse. Throughout the process, from one end of the value chain to another and back again, there should be constant feedback between producers and stakeholders. The data value chain can be used as a teaching tool to show the complex set of steps from data creation to use and impact or as a management tool to monitor and evaluate the data production process. Like any empirical model, applications of the data value chain may differ from one instance to another.

---

6 While data are a public good, there are instances where data are proprietary such as data in the private sector where data may be restricted because it provides a revenue stream.

7 Custer, S. and Sethi T. (Eds.). (2017). Avoiding Data Graveyards: Insights from Data Producers and Users in Three Countries. Williamsburg, VA: AidData at the College of William & Mary.

8 Development Gateway. (2017). Results Data Initiative.

DATA VALUE CHAIN

- Partner with other producers
- Harness new data sources

- Extract insights from data
- Visualize data in clear manner

- Partner with infomediaries
- Publicize data availability

- Reduce time-cost of data use
- Encourage perception of value

- Achieve first end-use of data
- Receive initial user feedback

- Build habits of data use
- Increase interest in data production

PUBLICATION

IMPACT

IDENTIFY · COLLECT · PROCESS · ANALYZE · RELEASE · DISSEMINATE · CONNECT · INCENTIVIZE · INFLUENCE · USE · CHANGE · REUSE · FEEDBACK

COLLECTION

UPTAKE

- Consult with future users
- Determine levels of granularity

- Ensure data are interoperable
- Achieve high-quality and protected privacy

- Provide machine-readable data
- Data accessible online and offline

- Use technology to connect to users
- Re-process data for new insights

- Promote data use culture
- Encourage data use for decisions

- Track tangible behavior changes
- Identify data-driven policies

PRODUCTION   **USE**

increasing value of data

⚠ Roadblocks for **production** include lack of financial, human, and technological resources; low data literacy; lack of trust between users and data collectors; blindspots in data gaps; lack of country ownership; and lack of government desire for transparency.

⚠ Roadblocks for **use** include low political support; lack of data relevance to decisions; poor quality; lack of trust in government data use; no rewards or results of data use; financial constraints; corruption; data silos; and lack of partnerships between infomediaries.

MARKERS ▸ Potential achievements within each process of the value chain mark progress towards data impacts.

Collection

The first stage of the data value chain is *collection*. Even if data fell from trees, we would have to ask where to find the trees, how to gather their fruit, and how to clean them and prepare them for consumption. Just so, in the collection phase we begin by asking what kind of data we need to solve a problem, answer a question, or monitor a process. In the first phase of data collection, we are **identifying** what data to collect and how we will use them. In the next phase, we establish a process for **collecting** the data. This may involve surveys or retrieval of administrative data or the use of remote sensing methods. Decisions made here will affect the quality and usability of data over their whole life-cycle. The last phase of data collection involves **processing** data to ensure they are correctly recorded, classified, and stored in formats that allow further use.

Collecting new data on time spent on household chores by boys and girls

SDG target 5.4 calls for recognizing and valuing time spent on unpaid household services (UHS) – such as house cleaning, caring for siblings, and other chores – and draws attention to disparities in the burden of household chores between the sexes and their impact on the welfare of women and girls. How much time do girls spend on household chores compared to boys? How does an unequal distribution impact a girl's health, her ability to stay in school, or to be involved in sports or other activities that accumulate human capital?

The UNICEF publication, *Harnessing the Power of Data for Girls*[9], pulls together the first global estimates of household chores by young girls and boys. Based on new data from MICS surveys, the report finds that girls spend 50 percent more time on chores than boys. Worldwide, girls aged 5–9 and 10–14 spend, respectively, 30 percent and 50 percent more of their time helping around the house than boys of the same age. In some regions, the gender disparities can be even more severe:

9 United Nations Children's Fund (2016), Harnessing the Power of Data for Girls: Taking stock and looking ahead to 2030. New York: UNICEF.

In the Middle East and North Africa and South Asia regions, girls aged 5–14 spend nearly twice as many hours per week on household chores as boys of the same age. Having these data are important to understand the constraints girls face in, among others, accessing and completing schooling, and advocate for more equal opportunities for them.

To better understand the time spent on household chores, UNICEF revised the child labor module for its Multiple Indicator Cluster Survey (MICS) program in 2013. The new child labor module carefully differentiates between economic activities and household tasks, providing data that better measure the roles of girls and boys.

A methodological study commissioned by UNICEF[10], using data from ILO labor force surveys and MICS, attempts to quantify the harm done to children through UHS through their effect on children's schooling. While the study finds only a weak association between time spent on UHS and school attendance, in depth country studies point to significant risks associated with hazardous UHS activities, such as exposure to fire or gas or use of dangerous tools. The study concludes that far more detailed data are needed to properly measure the hazards faced by children: more information on the conditions under which UHS are performed, time-use surveys that provide information on the activities carried out by children and their exact timing in a typical day, and other purposefully collected data that measure the link between UHS and leaning outcomes. To protect the lives of children of both sexes, data must be collected, analyzed, and presented in a manner that allows appropriate laws and regulations to be developed.

This stage highlights some of the most basic yet important statistical activities. It involves specifying the evidence needed to answer a question or understand a problem. What questions will the data be used to answer? Who will use them? It also involves calculating sample sizes, efficient modes of data collection, and ensuring the integrity of the data. Following standards like the IMF's Data Quality Assessment Framework (DQAF)[11] helps producers ensure that they provide high quality data that their users can trust – a critical component to value and use. The last fundamental activity in this stage is cleaning, coding, and storing the data. The value of data begins to grow in this stage.

Publication

The second stage on the value chain is *publication*. Once data have been collected, the data and the accompanying metadata must be published in such a way that data users can access them. The publication stage involves three activities: **publishing** data with appropriate documentation in online and offline formats; **disseminating** the data to prospective users; and **analyzing** data to extract useful information.

Using high-resolution mapping to explore sex-disaggregated development indicators

Bosco and others explored the use of high-resolution mapping of sex-disaggregated development indicators.[12] The aim of the study was to produce maps and visualize the spatial distribution of SDG indicators affecting women using geo-located Demographic and Health Survey (DHS) data. Their study looked at sex-disaggregated indicators for literacy, stunting as well as the use of modern contraceptive methods among women in Nigeria, Kenya, Tanzania, and Bangladesh. They created maps like the ones below showing the correlation between the indicators and characteristics of their geographic locations, such as topography, climate, population density, and ethnicity. The published maps – a higher value form of the original data – are a more friendly and understandable translation of these analyses and are available for use by policymakers or as an advocacy tool. The study also published detailed directions for constructing the dataset for use by other researchers.

---

10 Dayıoğlu, Meltem. Impact of Unpaid Household services on the Measurement of Child Labour. MICS Methodological Papers. Paper No. 2, 2013. https://data.unicef.org/wp-content/uploads/2015/12/Child_labour_paper_No.2_FINAL_163.pdf.

11 United Nations Trade Statistics Knowledgebase. IMF Data Quality Assessment Framework

12 Bosco, C., Alegana, V., et al. (2017). Exploring the high-resolution mapping of gender-disaggregated development indicators. Journal of the Royal Society Interface, Volume 14, Issue 129, April 2017.
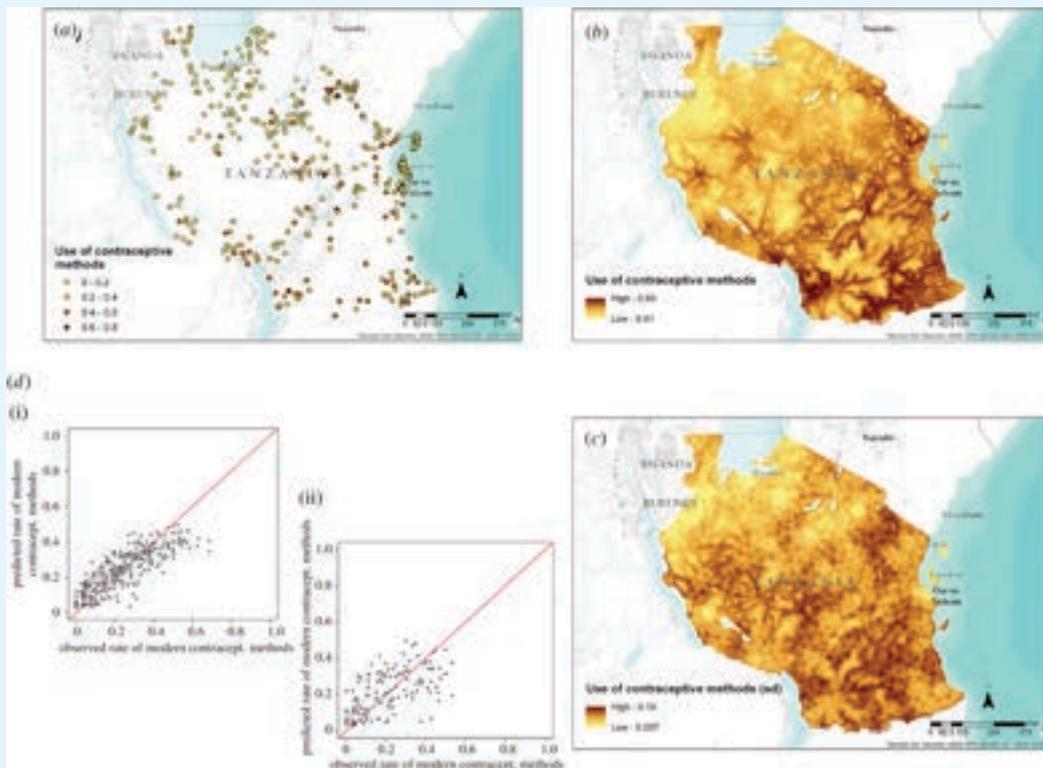
The figures[13] show input data, scatterplots, and output maps related to the use of modern contraception methods in women, with low levels of uncertainty in the predictions across the country. Key covariates across the different variables modelled were the distance from roads, accessibility, aridity index, and precipitation.

As data move through the steps of the publication stage, they increase in value. To optimize their value, it is important to strategically and carefully think through how data will be analyzed, released, and disseminated to encourage use for future impact. Are the data reaching their intended users through appropriate dissemination channels? Are they in open and accessible formats? Are metadata provided and do they conform to international standards?[14] Editing and compiling aggregates, creating tables and visualizations, and disseminating the results with the end user in mind are critical activities on the road to creating impact.

Uptake

The *uptake* stage involves three activities: **connecting** data to users; **incentivizing** users to incorporate data into the decision-making process; and **influencing** them to value data. Connections to users can be made in many ways: through press releases and online dissemination; by holding trainings, seminars or other educational events; and by improving the user experience offered by websites, data portals, and archives. Incentives can take many forms. Within government, data producing agencies may be ordered to publish data or offered additional budget to do so. A national statistical office may sponsor trainings for their staff. Training initiatives increase the value of data by preparing users with the competency to use the data.

Operating agencies can be required to incorporate data into their decision making and management practices; they can also be rewarded for informative infographics and other uses of statistical information. Beyond bureaucratic incentives to produce and use data, politicians, decision-makers, and program managers may need to be further influenced before they will adopt new habits of data use. Advocates

---

13 (a) The distribution of cluster-level data from the DHS household survey in Tanzania showing the proportion of women aged 15–49 using modern contraceptive methods. (b,c) Map of the mean predicted proportion of women using modern contraceptive methods at 1 km2 resolution (b) and related uncertainty map (c) showing its standard deviation. (d) Scatter plot of the predicted proportion of women using modern contraceptive methods (y-axis) by observed data (x-axis) for the training (i) and validation (ii) subset of data.

14 The Official Site of the SDMX. (2016). SDMX Technical Specifications.

within governments and agencies can promote a culture that values evidence-based policies and accurate accounting of outcomes. It is also the role of civil society advocates and academics to put pressure on decision-makers to demonstrate that their actions are evidence-based and accountable.

As users inside and outside of government value and need development data, their engagement with data producers is likely to increase, further improving the quality and relevance of the data. It is important to note that trust in the quality of the data is a crucial pre-requisite to incentivize users to use data. To increase credibility, producers should be as transparent as possible about data collection methods and quality controls. It is also important to separate commentary or interpretation of the data from metadata, especially where partisan, political interests may be involved. Uptake of data will depend on perceptions of autonomy or lack of political interference, combined with trust, relevance, and quality.

The perception of data as trust-worthy and of high-quality can be thought of as its brand recognition. Building the brand of data and gaining awareness of its value can increase impact. As the United Nations Economic Commission for Europe (UNECE) notes,

> "Excessive modesty about official statistics is dangerous. We can realize the potential value of what we produce only if users recognize what we have to offer and turn to us to meet their needs. Official statistics have a strong comparative advantage, its unique selling point noted above. Explicit brand recognition and promotion strategies for each national statistics office would advertise our strengths."[15]

Increasing and encouraging data use through a Gender Statistics Toolkit

The Gender Statistics Toolkit designed by United Nations System Staff College (UNSSC) and the African Centre for Statistics of the United Nations Economic Commission for Africa[16] (UNECA) educates statistics users and producers about the development of gender statistics at the country level. The course provides orientation in best practices and issues in planning, gathering, and sharing unbiased gender statistics. This training provides a foundation of knowledge that will reduce barriers to producing high-quality gender statistics and using them as a basis for informed decision-making, policy and program formulation, and monitoring.

The toolkit is structured around four modules. The first introduces gender statistics and covers the importance of gender statistics in the African context[17], key gender terms and indicators, and locating additional educational resources. The second module discusses planning a gender statistics initiative, including the importance of producer-user dialogue, conducting a gender-sensitive needs analysis, and communicating the need for new data collection. The third module discusses producing gender statistics: integrating gender perspectives into the data production process, identifying gender issues and gender-biased language in surveys, and how to provide feedback to on bringing a gender perspective into statistical activities in the offices of course participants. The fourth module discusses communicating and using gender statistics as a tool for change, including ways to disseminate gender statistics, prepare reader-friendly gender data tables, choose data visualizations, and make policy recommendations based on gender-data analysis.

Training programs such as UNECA's Gender Statistics Toolkit not only help connect users to the data but also increase their capacity and literacy to use the data. Trainings or toolkits may serve as evidence of a national statistics office's or international organization's recognition of the importance of data uptake and use through investments in its data users.

---

15 United Nations Economic Commission for Europe. (2017). Value of official statistics: Recommendations on promoting, measuring and communicating the value of official statistics. Geneva: UNECE.

16 United Nations Economic Commission for Africa (UNECA). http://uneca.unssc.org/course/view.php?id=2.

17 The participants page for this course in the Gender Statistics Toolkit shows that a total of 185 participants (not including the test-users) have registered to take this course. Of these participants, 109 are from African countries.

Impact

The *impact* stage involves three activities: **using** the data to understand a problem or make a decision; **changing** the outcome of a project or improving a situation; and **reusing** the data by combining them with other data and sharing them freely. The use of development data is often seen as the end goal, but use is something that occurs throughout many different stages between the publication of data and encouraging a decision-maker to modify decisions based on the data. Thus, at times the path from raw data to use is quick, other times it is a long process with many stages of analysis. To encourage support for data production, users and the general public should see clear positive changes as a result of using the data. Positive change will also increase the trust in the data, incentivizing further use. Data must also be open: freely licensed for use and reuse. The argument for providing additional resources and capacity building and the reason for advocacy and incentivizing users is to reinforce the use of information to improve people's lives. Once these habits are created, the channels connecting data and users will become more entrenched. With each cycle of production and use, collaboration between actors involved in the supply and demand sides will further improve the process, making it more efficient, sustainable, and effective.

Using sex-disaggregated data to improve financial inclusion of women

Women are still disproportionately excluded from the formal financial system and make up more than half of the world's unbanked population. According to the 2014 Global Findex survey, 50 percent of women had a bank account compared to 59 percent of men.[18] Sex-disaggregated data are essential to addressing the gender gap in financial inclusion. They both inform evidence-based financial inclusion policies and track the effectiveness of efforts to address barriers facing women. In 2012, the Banque de la République du Burundi conducted a demand-side financial inclusion survey.[19] The sex disaggregated results showed the bank that women were less likely to participate in financial services and highlighted the lack of traditional forms of collateral as a key constraint to women's access and inclusion. These data contributed to the design of the National Financial Inclusion Strategy 2015-2020 – specifically the new law entitled "Loi sur les Surêtés Mobilières," which will enable diverse types of collateral to be used to access credit, thus increasing women's access to credit. While the data had a direct impact on the design of policy, the next step would be to measure changes in women's access to credit and its impact on their economic opportunities.

Maternal and Perinatal Death Review (MPDR): Experiences in Bangladesh

The Maternal and Perinatal Death Review (MPDR), which collected data on maternal and neonatal deaths to inform remedial action plans at the community and facility level, helped reduce maternal and infant deaths in the Thakurgaon district of Bangladesh where it was piloted.[20] The project was a collaboration between a number of government and international organizations.[21] Data collection and dissemination included death notification followed by verbal and social autopsies, which obtain information about medical and social factors contributing to death. Mapping of deaths showed that the Kashipur sub-district had a high death density. Further investigation found that it was a hard-to-reach area, 30 kilometers from the nearest health complex. Policy-makers implemented a remedial

---

16 United Nations Economic Commission for Africa (UNECA). http://uneca.unssc.org/course/view.php?id=2.

17 The participants page for this course in the Gender Statistics Toolkit shows that a total of 185 participants (not including the test-users) have registered to take this course. Of these participants, 109 are from African countries.

18 The World Bank. (2016). "Infographic: Global Findex, Financial Inclusion." Washington, DC: World Bank.

19 Banque de la République du Burundi and the AFI Financial Inclusion Data Working Group. (2014). The Use of Financial Inclusion Data Country Case Study: Burundi.

20 World Health Organization. Maternal and Perinatal Death Review (MPDR): Experiences in Bangladesh.

21 This included the Line Director, Maternal Newborn Child, and Adolescent Health (MNCAH) of the Directorate General of Health Services (DGHS) in collaboration with the Line Director of the Health Monitoring Information System (HMIS) and e-Health, the Line Director of Maternal, Child, Reproductive and Adolescent Health (MCRAH) of the Directorate General of Family Planning (DGFP) within the scope of the Joint Government of Bangladesh (GoB) and United Nations (UN) Maternal and Newborn Health (MNH) Initiatives in Thakurgaon district. UNICEF in Bangladesh and Centre for Injury Prevention and Research, Bangladesh (CIPRB) provided technical support in the implementation process.

action plan to improve poor care-seeking behavior and use of health services, resulting in a reduction of deaths from 29 in 2011 to 22 in 2012. Other initiatives throughout the district included health camps for pregnant women to ensure pregnancy registration and antenatal care with birth planning. Managers also strengthened the referral system of high risk pregnancies. Immediate referrals for four mothers in the Kashipur sub-district mitigated pregnancy complications, allowing for timely interventions to save their lives. Because of this initiative maternal deaths in the Thakurgaon district fell from 59 in 2010 to 47 in 2014, while neonatal deaths fell from 739 to 683, and stillbirths fell from 633 to 535. This initiative showed that death notification in the community was possible and achievable, and the Government of Bangladesh agreed to expand the initiative to all 64 districts.

Conclusion

The example from Bangladesh highlights a clear link between data, policy, and impact. It demonstrates the direct value of investing in data and the returns on those investments. In a world where data and statistics receive inadequate funding, examples of impact are a means of advocating for data to be a higher priority in government budget allocations or donor funding decisions. And while it is important to collect and document these cases, the data value chain shows the need to expand focus to all stages of data's life cycle and recognize the importance of each stage.

The journey data travel from their creation to their use is more complex than simply moving from one stage to the next. However, with a better conceptual understanding of the value added at each stage, data can be more effectively and efficiently managed to create impact. This is the first iteration of the data value chain. We encourage data users and producers to provide feedback based on their own experiences. If you would like to share your input, please contact **info@opendatawatch.com**.