QUALITY, COVERAGE AND OPENNESS: CHALLENGES FOR OFFICIAL STATISTICS*

Misha V. Belkindas,
Eric V. Swanson

National statistical offices are under pressure to provide open access to their official datasets. While many have announced their intention to provide open data, progress has been slow and gaps in quality and coverage persist in many countries. Several indexes of the openness of official data are now available, but they overlook much of the output of national statistical systems and ignore most developing countries. Countries must implement rigorous assessments of data quality and embed openness in the plans for development of their statistical systems. Increased donor support, based on prioritized national plans, is needed to ensure that countries are able to produce the data needed to monitor progress toward the proposed Sustainable Development goals

Key words: open data, national statistical systems, international development.
JEL: O19, C81, C82.

The Mandate for Open Data. The first Fundamental Principle of Official Statistics[1] says that «...official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies». Further recognizing the «citizen's entitlement to public information», Principle 1 becomes a strong statement committing national statistical offices (NSOs) to providing open access to official statistics. And, indeed, most NSOs publish statistical yearbooks and other compendiums of their major statistical series and maintain websites that offer online access to selected statistics and other information[2]. As the Internet reaches more people and the power of computers and mobile devices increases, there are more opportunities and more formats for publication of official statistics. But not all published data are readily accessible and usable. They may be published in formats that are difficult to access; they may lack clear definitions and other explanatory metadata; they may be incomplete or out of date or just poorly organized.

To make official statistics truly open and valuable to their users, they must meet the «test of practical utility» specified by the first Fundamental Principle. Openness is not a label that can be applied after the fact. It is not a switch that can be turned on or off. It must be built into the statistical system. When the High-Level Panel on the Post-2015 Sustainable Development Goals[3] called for a data revolution, they identified four «components» to the revolution: «data quality and timeliness»; «data gaps», by which they meant improving coverage of current statistical series and collecting data on new and emerging issues; transparency, accessibility, and availability, which are defining characteristics of open data; and «harnessing diverse sources of knowledge» through the use of new statistical methods and technologies.

In this paper we will discuss characteristics of statistical systems in similar terms: data quality, data coverage, and data openness. New methods for compiling statistics- big data, for example, or crowd sourcing - may come to compete with traditional methods, but standards, definitions, and common nomenclature - the hallmarks of data quality - are still required. A comprehensive assessment of a statistical system should take into account all three dimensions: quality, coverage, and openness, along with its capacity to adopt new methods and technologies.

Defining Open Data. The earliest definition of open data was published in October 2005 by the Open

Belkindas Misha V. (mishabelkindas@opendatawatch.com) - PhD Economics, Managing Director, Operations, Open Data Watch, Inc., Washington, DC USA.

Swanson Eric V. (ericswanson@opendatawatch.com) - PhD Economics, Managing Director, Research, Open Data Watch, Inc., Washington, DC USA.

*The article is published in author's original version.

[1] United Nations, Fundamental Principles of Official Statistics (revised preamble 2013). Available at: http://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx.

[2] Of 193 members of the United Nations all but 6 have a website listed for their national statistical websites. Available at: http://www.opendatawatch.com/Pages/NSO-List.aspx.

[3] See their report, available at: http://www.post2015hlp.org/the-report/.

Knowledge Foundation. The current version (1.1) of the «Open Definition» extends itself to knowledge in general, including music, film, and books, but refers specifically to «Data be it scientific, historical, geographic or otherwise» and «Government and other administrative information». The Open Definition sets out eleven necessary and sufficient conditions for data to be considered open[4]. These include access at reasonable cost; non-discrimination against persons, groups, or fields of endeavor; explicit licensing under some form of recognized open license; and absence of technological constraints to the use or reuse of data.Other versions of an open data definition include The Eight Principles of Open Government Data[5] and the recent G8 Open Data Charter[6], which sets out an action plan for achieving open government data based on five principles.

Taking the common elements of various open data definitions and considering their application to official statistics published on the Internet or World Wide Web, criteria for assessing the «openness» of the data dissemination programs of official statistical agencies emerge.

1. *Legal authority and independence.*

What is the legal authority of the agency to collect and publish data? Is there a statistical law that defines that authority and does the law guarantee the independence and impartiality of the statistical agency? Is the statistical law published and accessible through the website? Is the agency responsible for production and dissemination of the data clearly identified?

2. *Completeness.*

Do the data available on the website represent the full range of the agency's statistical holdings or only a subset? Are primary data available at the finest granularity consistent with protecting privacy? If the national statistical system is a federated system composed of multiple statistical offices, is there a convenient way to locate all of the important subsets of data?

3. *Timeliness and time span.*

Are data regularly updated? Is the updating schedule published? Are consistent historical series maintained and available?

4. *Adherence to standards.*

Does the responsible agency employ recognized international standards and definitions for the compilation and documentation of statistics? If exceptions are made to international standards, are they fully documented?

5. *Availability of metadata.*

Are metadata describing the relevant characteristics (including standards and definitions) of the data readily available with the data? Is other objective, interpretive commentary provided, as recommended by the Principals of Official Statistics? Is the commentary free of political or partisan considerations?

6. *Selectability.*

Can the user specify a unique selection of data from a larger data set? Is it possible to download a complete data set?

7. *Technical accessibility.*

Are the data published in commonly used formats that facilitate machine processing? Is there at least one non-proprietary option for retrieving data? Is there a published API?

8. *Licensing.*

What are the terms of the license under which the data published? Are they non-discriminatory? Does the license permit the use and reuse of data without restriction (except, possibly, requiring attribution)? Are data in standard formats provided free of charge? Do users have to register to access or download data?

Measuring Openness. Two recently published indexes of data openness - the Open Data Index (ODI)[7] and the Open Data Barometer (ODB)[8] have applied versions of these criteria to selected national and city-level datasets, but neither attempts to assess the full range of data produced by a national statistical office. Coverage of the indexes is, therefore, very limited.Nor do they explicitly assess data quality. Nevertheless they represent interesting approaches to evaluating the openness of government data. A third index, the World Bank's Statistical Capacity Indicator (SCI), evaluates the capacity of a national statistical system to produce major economic and social statistics against standards for statistical methodology, source data, and periodicity and timeliness,

---

[4] See The Open Definition, available at: http://opendefinition.org/od/..

[5] See «The Annotated Eight Principles of Open Government Data». Available at: http://opengovdata.org/. This website also includes a useful summary of other working definitions of open government and open data.

[5] «The G8 Open Data Charter». Available at: (https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex).

[7] «About the Open Data Index». Accessed at https://index.okfn.org/about/.

[8] The Web Foundation and The Open Data Institute, The Open Data Barometer 2013 Report. Available at: http://www.opendataresearch.org/barometer.

important elements of data quality[9] However, the SCI does not include openness among the evaluation criteria. The three indicators also differ in their country coverage. The ODB and ODI include a large number of high-income countries, while the SCI includes only countries classified by the World Bank as low- or middle-income. The ODB has 46 or 47 countries in common with the ODI and SCI. The ODI, which includes fewer developing countries, has only 28 countries in common with the SCI. Both the ODB and ODI omit more than half of the countries classified by the World Bank as low- or middle-income.

Despite their differences in construction and coverage, the three indicators have much in common. Their indicator scores and ranking of countries are, for the most part, highly correlated as shown in Table.

Table

Correlations of country scores and rankings
(percent)

	Open Data Barometer 2013	Open Data Index 2013	World Bank Statistical Capacity Index 2012
Correlation of scores			
Open Data Barometer	--	84.4	62.2
Open Data Index	84.4	--	78.2
World Bank Statistical Capacity Index	62.2	78.2	--
Rank correlation			
Open Data Barometer	--	84.4	69.5
Open Data Index	84.4	--	65.9
World Bank Statistical Capacity Index	69.5	65.9	--
*Correlation of scores - 18 low- and- middle-income countries**			
Open Data Barometer		30.5	47.0
Open Data Index	30.5		73.9
World Bank Statistical Capacity Index	47.0	73.9	

*Bangladesh, Brazil, Burkina Faso, China, Costa Rica, Ecuador, Hungary, India, Indonesia, Kenya, Mexico, Nepal, Nigeria, Russian Federation, Senegal, South Africa, Tunisia, Yemen.

When taking into account countries the indexes share in common by pairs, the correlations between the ODB and ODI indexes are somewhat higher than their correlations with the SCI, but they are all close to each other. This is, perhaps, unsurprising: countries with the capacity to produce reliable social and economic statistics are also more likely to have statistical systems capable of producing and disseminating the datasets evaluated by the ODB and ODI. However, the similarities disappear when we consider only the 18 low- and middle-income countries that are common across the three indexes. The ODB appears to be most affected: its correlation with the ODI falls to 30 percent, which is less than its correlation with the SCI at 47 percent. The correlation between the ODI and SCI remains at nearly the same value as in the larger, common subset. This suggests that the rating system used by the ODB may be less robust in developing countries.

Index measures such as the ODI, ODB, and SCI are a useful tool for identifying statistical systems that lack the capacity to meet expected performance standards. But to provide a robust assessment of a statistical system a more comprehensive approach is needed. First, the assessmentshould cover all of the statistical domains that are required for public policy making and for monitoring and evaluation of social, economic, and environmental conditions. Within each domain it should consider the timeliness and frequency of statistical series and the depth of disaggregation available. Then we can ask whether the data provided openly, applying the Open Definition to each domain. These characteristics can be quantified, at least on an ordinal basis. Assessing quality is more difficult: data that are adequate for one purpose may not be adequate for another. The provision of metadata - one of the requirements of the Open Definition - may help users decide whether data are fit for purpose. Ultimately the elements of data quality must be built into the statistical system.

Improving Data Quality. The past two decades have seen efforts on multiple fronts to improve the quality and availability of the statistical information needed for planning, monitoring, and assessing the social and economic development of a country. Much of the recent effort has focused on the indicators needed to monitor progress toward the Millennium Development Goals[10] (MDGs). The 2004 Marrakech Action Plan for Statistics[11], which set the agenda for international support for statistics, included improvements in MDG indicators among its six action items. The availability of the MDGs indicators has improved, but many of these

---

[9] See «Statistical Capacity Indicator». Available at: http://bbsc.worldbank.org.

[10] See http://www.un.org/millenniumgoals/.

[11] See http://www.worldbank.org/en/data/statistical-capacity-building/marrakech-action-plan-for-statistics.

improvements were the result of modeling, special surveys, or other data collection programs funded and carried out by bilateral donors and international agencies. But many gaps remain in fundamental statistical series: demographics, economic well-being, educational attainment, health status, and the economic and physical condition of the built and natural environment. The recent report by the Data2X project[12] of the UN Foundation documented extensive gaps in the availability of gender-disaggregated statistics needed to document the status of women and girls.

The challenges facing statistical offices are growing larger. The indicators proposed for the Sustainable Development Goals for 2030[13] are more numerous and complicated than the MDGs. The 2009 Sen-Stiglitz-Fitoussi report on the Measurement of Economic Performance and Social Progress[14] has laid out an ambitious program of fundamental changes in the measurement of economic, social, and environmental indicators. But six years after the publication of the 2008 revision to the System of National Accounts, many countries are still trying to bring their national accounts into conformity with the 1993 SNA.

Quality assurance frameworks. In 1996 the International Monetary Fund proposed a system of voluntary standards for certain categories of economic and financial data. The Special Data Dissemination Standard[15] (SDDS) was intended to apply to high- and middle-income countries participating in international financial markets. In 1997 the IMF launched a less stringent program, the General Data Dissemination System[16] (GDDS), which encouraged countries with less developed statistical systems to report on existing statistical practices and to set priorities for improving them. Both systems emphasized the importance of disseminating reliable information to final users, although neither specially mandated open data standards. In 2001 the IMF proposed to standardize the framework through which it documented the quality-related features of the governance of statistical systems, statistical processes, and statistical products. This became known as the Data Quality Assessment Framework[17] (DQAF). The IMF has produced DQAFs for important economic and financial datasets. In cooperation with the World Bank it has also produced a DQAF for poverty statistics and population statistics. UNESCO working with the World Bank has produced a DQAF for education statistics.

The GDDS and DQAF are a useful organizing framework for documenting statistical practices. But they are not complete in many countries; they do not cover all important statistical sectors; and many of those available are out of date. Other international, regional, and national statistical agencies have developed quality frameworks. Some are focused on specific sectors, others take a system-wide approach. There is no universally recognized standard for such frameworks, but any of the existing DQAFs could provide a serviceable template for documenting national statistical practices if applied consistently and regularly updated.

Although existing data quality frameworks do not insist on rigorous adherence to standards for open data, they encourage greater transparency at each stage of the production process from sample design and data collection through to the compilation and dissemination of statistics. This helps to establish a positive feedback loop leading to improvements in data quality when data users are able to verify the fitness of data and identify shortcomings. Truly open systems amplify the feedback effect by encouraging widespread use of data.

Planning for open statistics. The PARIS21 Guidelines for national strategies for the development of statistics[18] (NSDS) have encouraged a more rigorous approach to planning for improvements of national statistical systems. An NSDS should provide a prioritized program for building national statistical capacity, improving data quality, and addressing user requirements. It should also provide a realistic cost analysis, including estimates of the domestic and donor resources required. Many developing countries have now adopted the NSDS approach, although implementation of the resulting plans has lagged. However, the current NSDS guidelines have failed to keep up with the rapidly changing technical, institutional, and political environment

---

[12] See http://www.unfoundation.org/what-we-do/issues/women-and-population/data2x.html.
[13] See http://sustainabledevelopment.un.org/focussdgs.html.
[14] Available at: http://www.stiglitz-sen-fitoussi.fr/en/index.htm.
[15] See http://dsbb.imf.org/Pages/SDDS/Home.aspx.
[16] See http://dsbb.imf.org/Pages/GDDS/Home.aspx.
[17] See http://dsbb.imf.org/Pages/DQRS/DQAF.aspx.
[18] See http://nsdsguidelines.paris21.org/.

that is shaping the data revolution. Notably, existing NSDSs fail to take account of the demand for open data

National statistical systems in many developing countries will have to undertake significant reforms to integrate open data standards into revised NSDSs. With some well-thought through revisions to the data management and dissemination practices, along with legal, institutional, technical, and quality assurance frameworks, open data can be mainstreamed in the routine production of data. This will align data production systems with the demands of data users both inside and outside government.

An open NSDS should adopt the following features to develop a full commitment to open data:

• Make the case for open data and expected values to be gained.

• Seek inputs from all stakeholders in the development of the plan.

• Establish the legal framework to promote the dissemination of open data, including granting permission for free commercial use and reuse of data.

• Engage prime users in technical discussions about the design of major surveys and censuses.

• Articulate the step-by-step development of integrated databases and meta-databases.

• Make realistic estimates for staff training and other needs for proper open data dissemination

• Include realistic estimates of the resources needed for IT infrastructure (both hardware and software).

• Make unit record data collected or compiled by the national statistical office available (with appropriate safeguards for statistical confidentiality).

• Make macro indicators and social statistics available with accompanying meta-data in non-proprietary formats.

• Ensure regular updating of the web pages associated with country subscriptions to the IMF's SDDS or GDDS frameworks.

• Consult with data users in the formulation of changes to the subsequent annual work programs and provide a feedback mechanism for comments, complaints, and information requests.

• Conduct periodic audits employing the IMF's DQAF or other quality assessment frameworks.

Support for statistics. Before the adoption of the Marrakech Action Plan for Statistics (MAPS) in 2004, support for statistical systems in developing countries was generally limited to specific sectors, such as national accounts, or to surveys designed to collect data of interest to donors that were often managed by external consultants. Donors' projects were uncoordinated and often in conflict with each other. MAPS mobilized more than $100 million for partnerships to support the completion of the 2010 round of censuses, establisha national statistical planning process, implement a program to harvest indicators from existing surveys, and established the International Household Survey Network[19]. In the following years, donor support for statistics increased by 60 percent, reaching $1.6 billion in the period 2008-2010. In 2011 the five-point Busan Action Plan for Statistics[20] was adopted at the 4th High Level Forum on Aid Effectiveness. It called for fully integrating statistics in decision making; promoting open access to statistics; and increasing resources for statistical systems. According to the most recent PRESS report by PARIS21[21], donors provided $394 million for statistics in 2013, an increase over 2012 but 23 percent less than in 2011. The top five donors in 2013 were World Bank, United Kingdom, European Union, African Development Bank and United Nations Population Fund. It should be noted that the World Bank's share of donor support for statistics was 20 percent in 2006-2013; it rose to 56 percent in 2013. Such a concentration of ODA for statistics in one organization is worrisome because it leaves countries in need of assistance vulnerable to shifts in policies by one institution. It would be important, therefore,to have more donors increase their support for statistics.

The most active donors in the Commonwealth of Independent States in the years of transition were the European Union through the TACIS program and the World Bank through the Trust Fund for Statistical Capacity Building and the STATCAP lending program. The Russian Federation, Ukraine, Tajikistan and Kazakhstan went through comprehensive restructuring programs financed by World Bank loans. In recent years the Russian Federation and the World Bank have established the ECASTAT trust fund to support the least statistically developed countries in the region[22]. The fund supports region

---

[19] See http://www.ihsn.org/.
[20] See http://www.paris21.org/busan-action-plan.
[21] Available at: http://www.paris21.org/PRESS2013.
[22] More information is available at: http://search.worldbank.org/all?qterm=ECASTAT.

wide programs as well as country specific programs.

Despite these efforts, international support for statistics remains fragmented and lack a consensus on priorities, methods, and funding. In our experience, poor countries are unable or unwilling to borrow substantial sums of money to finance statistical projects and make do with limited grants. Donors too have been reluctant to finance comprehensive statistical capacity building programs. To make progress national statistical offices will need to demonstrate better uses of the existing funds and convince donors, other stakeholders, and their own governments to bring additional money to the table. Donors should support system-wide reform and capacity building, rather than buying «products» such as surveys or the delivery of indicators. Realistic plans embodied in NSDSs or similar documents should establish priorities and a framework for coordinating domestic and external support. And objectives assessments, based on recognized standards, should be used to measure progress toward establishing open statistical systems. The proposed Sustainable Development Goals (SDGs)[23] will create unprecedented demands on national statistical systems. Meeting those demands will require a renewed effort to build open and capable statistical systems.

References

1. International Household Survey Network. «Data archiving and dissemination». http://www.ihsn.org/home/archiving.

2. International Monetary Fund a. «Data Quality Assessment Framework». Dissemination Standards Bulletin Board. http://dsbb.imf.org/Pages/DQRS/DQAF.aspx.

3. International Monetary Fund b. «General Data Dissemination System». Dissemination Standards Bulletin Board. http://dsbb.imf.org/Pages/GDDS/Home.aspx.

4. International Monetary Fund c. «Special Data Dissemination Standard». Dissemination Standards Bulletin Board. http://dsbb.imf.org/Pages/SDDS/Home.aspx.

5. Open Knowledge Foundation a. «Open Data Index: Tracking the state of government open data». http://index.okfn.org/about/. Open Knowledge Foundation b. «The Open Definition». http://opendefinition.org/od/.

6. PARIS21 a. «NSDS Guidelines». http://nsdsguidelines.paris21.org/

7. PARIS21 b. Statistics for Transparency, Accountability, and Results: A Busan Action Plan for Statistics. http://www.paris21.org/sites/default/files/Busanactionplan_nov2011.pdf

8. **Stiglitz, Joseph E., A. Sen, J.-P. Fitoussi.** 2009. Report by the Commission on the Measurement of Economic Performance and Social Progress.http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf.

9. **Tauberer, Joshua.** «The Annotated Eight Principles of Open Government Data». http://opengovdata.org/.

10. The Web Foundation and the Open Data Institute. 2013. The Open Data Barometer 2013 Report.http://www.opendataresearch.org/barometer.

11. United Kingdom Cabinet Office. 2013. «The G8 Open Data Charter and Technical Annex». Accessed at (https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex

12. United Nations Foundation. «What We Do: Data2X: Closing Gender Data Gaps». http://www.unfoundation.org/what-we-do/issues/women-and-population/data2x.html.

13. United Nations a. «Fundamental Principles of Official Statistics (revised preamble 2013)». http://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx.

14. United Nations b. «Millennium Development Goals and Beyond 2015». http://www.un.org/millenniumgoals/.

15. United Nations. 2013. «A New Global Partnership: Eradicate Poverty and Transform Economies through Sustainable Development». The Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda. http://www.un.org/sg/management/pdf/HLP_P2015_Report.pdf .

16. United Nations c. Department of Economic and Social Affairs. «Open Working Group proposal for Sustainable Development Goals». http://sustainabledevelopment.un.org/focussdgs.html.

17. United Nations d. Open Working Group Proposal for Sustainable Development Goals.https://sustainabledevelopment.un.org/content/documents/1579SDGs%20Proposal.pdf .

18. World Bank. «Data on Statistical Capacity». http://datatopics.worldbank.org/statisticalcapacity/

19. World Bank. «Marrakech Action Plan for Statistics». http://www.worldbank.org/en/data/statistical-capacity-building/marrakech-action-plan-for-statistics.

[23] See the report of the Open Working Group of the General Assembly on Sustainable Development Goals. Available at: http://undocs.org/A/68/970.

КАЧЕСТВО, ОХВАТ И ОТКРЫТОСТЬ: ЗАДАЧИ ОФИЦИАЛЬНОЙ СТАТИСТИКИ

Миша В. Белкиндас

Аффилиация: Open Data Watch (Вашингтон, округ Колумбия, США). E-mail: mishabelkindas@opendatawatch.com.

Эрик В. Суонсон

Аффилиация: Open Data Watch (Вашингтон, округ Колумбия, США). E-mail: ericswanson@opendatawatch.com.

Национальные статистические службы находятся под давлением, чтобы обеспечить свободный доступ к своим официальным базам данных. В то время как многие из них заявляют о своем намерении предоставить открытые данные, прогресс идет медленными темпами и во многих странах сохраняются пробелы в качестве и охвате предоставляемой информации. Несколько индексов открытости официальных данных теперь стали доступны, но они не в полном объеме учитывают выходной продукт, производимый национальными статистическими системами, и оставляют без внимания большинство развивающихся стран. Страны должны осуществлять строгую оценку качества данных и их открытости, заложенной в планах развития их статистических систем. Увеличение донорского финансирования на основе приоритетных национальных планов необходимо для того, чтобы страны были способны производить данные, необходимые для отслеживание прогресса на пути к достижению заявленных целей устойчивого развития.

Ключевые слова: открытые данные, национальные статистические системы, международное развитие.

JEL: O19, C81, C82.