

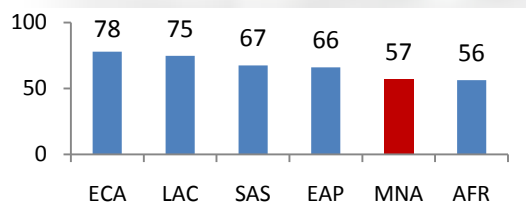


ENHANCING ACCESS TO MICRO-DATA IN THE MIDDLE EAST AND NORTH AFRICA

Diego F. Angel-Urdinola, Anne Hilger, and Ingrid B. Ivins¹

Introduction: Greater transparency and accountability has been one of the core issues behind the current upheavals in the region. Similarly, lack of or inability to gain access to data has been a long standing complaint by local and international agencies and organizations. This state of affairs makes it difficult to design good policies and to put together indicators to hold governments and others accountable. This Arab World Brief looks at the current state-of-play in terms of access to data as well as measures to improve the situation.

Figure 1: World Bank Statistical Capacity Index 2010



Source: World Bank Statistical Capacity Index (2010)

Background: In spite of recent achievements concerning the quality and availability of data in the MENA region, there is still much room for improvement. The World Bank Statistical Capacity Index (see Figure 1) provides an overview of the statistical capacity of developing countries². The Index consists of three dimensions, each capturing a different aspect of statistical capacity such as: (i) the ability of a country to adhere to internationally recommended standards and methods, (ii) whether a country conducts periodical data

collection and whether data are available and reliable, and (iii) the extent to which data are made accessible to users through the transformation of source data into timely statistical outputs. A higher index indicates a higher statistical capacity.³

The index reveals important differences in statistical capacity between low performing regions, mainly the MENA and Sub-Saharan Africa (AFR) regions, and high performing regions, i.e. Eastern Europe and Central Asia (ECA), and Latin America and Caribbean (LAC). A low statistical capacity index may result when countries do not produce micro-data, or despite producing it they do not make it available to users - whether government or private. In the particular case of MENA, all independent sub-indices are below world average, but the gap is largest when it comes to the methods used for data collection (i.e. sub-index 1).⁴

Lack of access to micro-data and Policy Design

Failures: In many MENA countries open access to data/micro-data remains restricted, even among government officials and public institutions. As a consequence, many recent and important social reforms (such as the introduction of family allowances and universal healthcare in Morocco and the introduction of the food-nutrition support program in Algeria, among others) were not well studied/assessed before their implementation (Bibi and Nabli, 2010). Research is indispensable for informed policy making. Indeed, in key areas such as labor markets and social protection, the number of existing publications is very limited, often due to the lack of access to micro-data. As illustrated in Figure 2, while MENA is among the regions with

¹ This Arab World Brief was cleared by Jonathan Walters, Director, Regional Strategy and Program, The Middle East and North Africa Region, The World Bank.

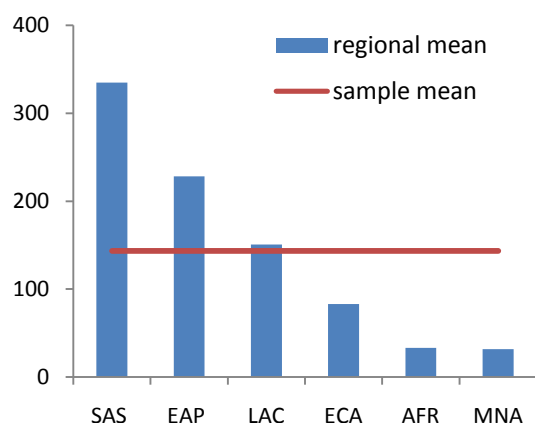
² A full description of the Index Methodology can be found here: <http://go.worldbank.org/3J9X57XKY0>.

³ While the index has been constructed for low and middle income countries, only middle income countries are considered in the following analyses.

⁴ Note that the correlation between statistical capacity index and the level of GDP per-capita is not significant for countries in the MENA region.

the lowest statistical capacity index, it is also the region which has the least number of publications in the area of labor markets.

Figure 2: Number of Publications on Labor Markets in EconLit by Region



Source: EconLit (2010)

Benefits of Enhanced Access to Micro-Data: The main benefits of open access to micro-data are the formulation of informed policies, and also the capacity to better monitor existing ones. Informed public policy design requires detailed information about target groups and their characteristics in order to conduct social and impact assessments of planned policies and reforms. Moreover, data access increases transparency which in turn leads to a higher public credibility. Through open access to data, results from researchers can be verified, fostering open research communities and better analysis. This in turn would generate new research further adding to the value of existing data and spurring the demand for additional analysis and data – in effect creating a virtuous cycle.

Availability of Micro-Data and Better Governance: There is a positive association between a country’s statistical capacity index and commonly used governance indicators, although the causality is probably bi-directional, i.e. better production and dissemination of micro-data is likely to contribute to public transparency and better governance in the same way that better governance is likely to foster data production and dissemination (Table 1).

Table 1: Statistical Capacity and Governance Indicators

	Correlation with Statistical Capacity Index
Voice & Accountability	0.3742*
Political Stability	0.2415*
Governance Effectiveness	0.5349*
Regulatory Quality	0.5285*
Rule of Law	0.4381*
Control of Corruption	0.4347*
Democratization index ⁵	0.3376*

Source: Author’s own calculations using Statistical Capacity Index (2009), WGI (2009), and Polity IV (2008) for EAP, ECA, LAC and MENA

Regional Initiatives to Enhance the Production of and Access to Micro-data: One way of improving production/access to micro-data is to engage in regional coordination and cooperation programs. An exemplary program is the MECOVI program⁶, established in Latin America in the 1990s. Before the program was introduced, access to and availability of micro-data in Latin America was very limited. While data was produced frequently by most statistical agencies in the region (similar to what occurs today in MENA), the data was largely inaccessible to public and private users. At the time, data produced by most national statistical offices were not comparable regionally due to differences in sampling frameworks, definitions, and data collected. To address these shortcomings, the MECOVI program was launched in 1996 and lasted a total of 8 years. The program’s main purpose was to improve surveys and the measurement of living standards in Latin America. The program was developed in a partnership between the World Bank, the Inter-American Development Bank and the Economic Commission for Latin America and the Caribbean (ECLAC).

⁵ (1): institutionalized autocratic regime, (2) weak autocratic regime, (3) weak democracy, (4) institutionalized democracy. Available at www.systemicpeace.org/polity/polity4.htm

⁶ MECOVI stands for “Programa para el Mejoramiento de las Encuestas y la Medición de Condiciones de Vida” - Program for the Improvement of Surveys and the Measurement of Living Conditions.

The MECOVI Program: MECOVI was launched in Paraguay but expanded quickly to other countries in the region, namely Argentina, Bolivia, Colombia, Ecuador, El Salvador, Guatemala, Honduras, Nicaragua, Paraguay and the Dominican Republic. Three components were at the core of the program: (i) improving the quality of regional data production, (ii) ensuring international comparability, and (iii) promoting data access and utilization. MECOVI turned out to be an overwhelmingly successful program, leading to better harmonization between regional statistical offices and to promoting public access to micro-data. The program contributed to the standardization of questionnaires in the region for most household budget surveys, the production of several studies and publications, the development of regional GPS, and to the modernization/standardization of software and equipment in the region. Today, several household budget surveys and labor force surveys in Latin America are available through the internet to public institutions, academia, NGOs and researchers. Better data dissemination and the introduction/promotion of panel surveys within the MECOVI framework also contributed to the improvement of the credibility of national statistical offices. No confidentiality issues arose and no negative experiences of the program were reported.

SEDLAC⁷: As a continuation of MECOVI, a follow-up program, SEDLAC, was introduced. SEDLAC is a data base for Latin America and the Caribbean, consisting of (i) regional socio-economic statistics computed from household surveys accessible from the web, (ii) labor statistics computed from labor surveys (LABLAC), (iii) reports, briefs, maps and other material describing the main patterns and trends in poverty, inequality, and other indicators in the region, and (iv) research on poverty and other social and labor issues in LAC. MECOVI and SEDLAC collectively led to a substantial improvement in data quality and comparability in Latin America. The achievements of the two programs can be recognized in the empirical evidence presented above, as the LAC region

⁷ Socio-Economic Data Base for Latin America and the Caribbean

currently performs above average concerning the World Bank Statistical Capacity Index, with an above average level of publications in the area of labor markets.

Piloting a MECOVI-type Program in MENA: Given that the regional context in MENA has some similarities with Latin America in the early 90s, a regional program to enhance access and production of micro-data for public use could be considered. Such a program could be piloted first in a sub-region, such as the Maghreb or the Mashreq. A regional MECOVI-like program would help the region in making its already existing data sets available for research and public policy analyses, leading to more knowledge on the countries' socio-economic situations and ultimately to more informed policy making. A regional program would also allow for better comparison across countries and for better harmonization and coordination of national statistical offices.

Confidentiality Issues: Regardless of the advantages of more access to micro-data, it is imperative to assure confidentiality. In considering whether or not to produce and disseminate micro-data, an important trade-off emerges between the benefits of data access and the necessity to guarantee confidentiality. Ethical and legal considerations require statistical agencies to strive for the confidentiality of the information supplied by survey respondents. From a pragmatic point of view, confidentiality is a necessary condition for participants to continue sharing confidential information. There are ways to balance data access and confidentiality by minimizing potential risks. This can be done by setting a threshold risk (to the individual record, or to the entire file), by removing unique/rare sample data, or by ensuring the anonymity of the data. Assuring confidentiality was one of the pillars of the MECOVI program and should be considered as a key component for the development of any regional or country-specific program aimed at improving statistical capacity. As a first step, MENA countries could also consider establishing a regional data set similar to Europe's LIS (Cross National Data Center in Luxemburg see www.Lisdatacenter.org). LIS

created a data base containing harmonized micro-data from middle and high income countries (40 countries, mostly Europe and North America). LIS allows for remote-execution, so that researchers can write programs (e.g. in STATA), send them via email to the LIS server, and then receive the results without having actual access to the micro-data. Of course, a system like LIS needs to sustain technical capacity to respond to client's demands, which in turn requires a sustainable mechanism of financing.

Contact Arab World Briefs:

Director, Regional Strategy and Programs
Jonathan Walters

Arab World Briefs Team: *Omer Karasapan, Roby Fields, and Hafed Al-Ghwel* – Tel #: (202) 473 8177

The Arab World Briefs are intended to summarize lessons learned from the Knowledge and Learning activities of the World Bank Group and our partners. The Briefs do not necessarily reflect the views of the World Bank, its board or its member countries.