

TRANSLATIONAL ARTICLE

Microdata collection and openness in the Middle East and North Africa

Uche E. Ekhator-Mobayode*  and Johannes Hoogeveen

World Bank, Washington, District of Columbia, USA

*Corresponding author. E-mail: uekhator@worldbank.org

Received: 08 January 2022; **Revised:** 24 June 2022; **Accepted:** 22 August 2022

Key words: MENA; Microdata; National Statistical Offices; Data openness

Abbreviation: MENA, Middle East and North Africa

Abstract

This article uses a “mystery client” approach and visits the websites of National Statistical Offices and international microdata libraries to assess whether foundational microdata sets for countries in the Middle East and North Africa region are collected, up to date, and made available to researchers. The focus is on population and economic censuses, price data and consumption, labor, health, and establishment surveys. The results show that about half of the expected core data sets are being collected and that only a fraction is made available publicly. As a consequence, many summary statistics, including national accounts and welfare estimates, are outdated and of limited relevance to decision-makers. Additional investments in microdata collection and publication of the data once collected are strongly advised.

Policy Significance Statement

The methodology to examine microdata accessibility developed in the article can be applied to all countries and covers various data categories that are important for tracking progress toward the sustainable development goals. The exercise is designed to foster conversations surrounding best practices for microdata presentation online drawn from various sources including National Statistical Offices websites and established microdata libraries. This can reduce the cost that countries incur while making improvements to close data accessibility gaps where it exists as they can learn from counterparts who have made more progress and avoid costs associated with designing new systems.

1. Introduction

Timely and consistent statistics are essential to inform and monitor economic, environmental, and social development. Yet to be used in decision-making, statistics need to be more than of good quality. They need to be timely and trusted. Trust in official statistics comes, broadly speaking, from two sources (Brackfield, 2011). The statistics themselves must be trustworthy and credible. Next, the institution producing the statistics needs to be trusted. Openness and transparency affect trust in official statistics through both pathways. Transparency allows the public to assess the methods and data used and increases trust in the

organization itself. In addition to being important for trust in official statistics, statistical transparency also yields an attractive return. Research in middle-income contexts demonstrates that the availability of quality, transparent, and timely disseminated macroeconomic and financial data reduces sovereign borrowing costs on international capital markets. Adherence to the Special Data Dissemination Standards (SDDS), for instance, lowers borrowing costs by 50 basis points as it reassures international investors on the reliability and serviceability of a country's economic and financial data (Cady, 2005). Also, open access to publicly funded data maximizes its research potential and provides greater returns from the public investment in research. When this is microdata, the burden on researchers of collecting quality microdata is reduced. Furthermore, inconsistencies arising from researcher bias in microdata collection are also minimized thereby increasing the value of insights such data provides.

In this article, we examine two aspects of statistical quality, microdata collection, and access. We focus on microdata for three reasons. They are an important source of data, especially for researchers, who without it often would not have the ability to carry out their work on nationally representative samples. The demand for readily available microdata can be illustrated with the 2017 Djibouti Household Survey. Its data have been downloaded 2,078 times even though the data was only uploaded on the World Bank microdata library in June 2019. After 20 months since the data have been publicly released, Google scholar already gives 290 hits of academic articles that have been prepared using this data set (checked on February 23, 2021). The inflow of research with new data strengthens the analytical capacity of the national statistical system and has huge marginal gains especially for lower-income countries that are less likely to conduct household surveys.¹

The second reason to focus on publicly releasing microdata is that by not doing so, the public use value of the data in research is foregone. This value can be significant. The cost of collecting the data is sunk (taxpayers have already paid for it) and the marginal cost of creating another copy of the database is negligible. The benefits on the other hand can be substantial. Increased accessibility to data has been related to the MENA region's chronic low-growth syndrome and Arezki et al. (2020) estimate that the region's lack of data transparency has resulted in losses of income per person ranging between 7 and 14%. MENA often produces a considerable number of reports while allowing little-to-no microdata access. Furthermore, MENA is the only region that underperforms its GDP in terms of economics research (Das et al., 2013). This may be the result of insufficient microdata availability.

The third reason to focus on the availability of microdata is because it demonstrates a credible commitment to transparency. Between 2005 and 2008 MENA was the only region globally to experience an absolute decline in the "statistical capacity index"—an index of data transparency (Arezki et al., 2020). More data transparency may improve political trust and create more social cohesion. Releasing microdata to the public requires balancing two fundamental principles of statistics: confidentiality and access. An agency not committed to data transparency could argue (erroneously) that privacy considerations—captured in every Statistics Act, prevent it from releasing anonymized microdata.

To assess access to microdata we take the perspective of an everyday data user and visit the public-facing websites of all National Statistical Offices (NSOs) in the MENA region as well as microdata libraries maintained by the World Bank, International Household Survey Network (IHSN), IPUMS, Eurostat and the Economic Research Forum (ERF). We also visit the web portals of the MICS and DHS surveys. Though, as World Bank staff, we often already have access as part of our official duties, we opt to follow a "mystery client approach" and explore which data can be accessed through public channels. We verify if up to date microdata across several data categories is available for download either immediately or after a request is made by the user. Informed by this exercise, we make suggestions aimed at improving NSOs' ability to provide up to date microdata online. The findings from the exercise show that many microdata sets are out of date or not collected at all. Since one cannot publish what is not collected, we strongly advocate for additional investments in microdata collection as well as publication of the data.

¹ Dang et al. (2019) provide evidence that countries with higher incomes more frequently implement household surveys.

The rest of the article is structured as follows: the next section explores in greater depth the intersection between public trust in official statistics and data transparency. [Section 3](#) discusses the data categories examined in the article, describes the exercise of visiting MENA NSOs' websites, and presents the results from the exercise. [Section 4](#) offers some suggestions for progress based on observations made by the research team from visiting NSOs' websites. [Section 5](#) discusses existing indicators measuring data accessibility in MENA and makes a case for a complementary indicator building on the methodology presented in the article. [Section 6](#) concludes.

2. Transparency and Trust in Statistics

Public trust in official statistics is anchored in professional independence and impartiality of statisticians, their use of scientific methods, and equal access for all to official statistical information. To operationalize these ideas, the international statistics community has adopted a professional code comprising of 10 principles, the *Fundamental Principle of Official Statistics*, and a set of “*Good Practices*.” Together they emphasize accessibility, impartiality, transparency, accuracy, relevance, cost-effectiveness, confidentiality, professionalism, coordination, and cooperation. At times, the Principles and Practices have conflicting requirements. Confidentiality, for instance, captured in Principle no six necessitates measures to prevent the direct or indirect disclosure of data on persons, households, businesses, and other individual respondents. As this could be interpreted as a prohibition to release source data, statisticians also commit themselves to “*a framework describing methods and procedures to provide sets of anonymous micro-data for further analysis by bona fide researchers, maintaining the requirements of confidentiality*.”² In this way, the Good Practices forge a compromise between confidentiality on the one hand and transparency and access on the other.

Access to microdata is typically offered in two ways. Some agencies make anonymized microdata directly available to the public. India's statistical agency, for instance, the Ministry of Statistics and Programme Implementation (MOSPI), has a long history of running national sample surveys dating back to the 1950s when they were initiated by Professor Mahalanobis, the father of Indian statistics, and of publicly releasing the anonymized microdata. On MOSPI's website, microdata sets are available for download dating back to as far as 1975. Other known sources of downloadable microdata sets are the World Bank's (WBs) microdata library,³ the DHS,⁴ and MICS⁵ websites, the labor force surveys curated by the ILO,⁶ and IPUMS⁷ which publishes (samples of) population censuses.

Others, like EUROSTAT, make microdata available in two formats: Public and Scientific Use Files. The Public Use Files (PUFs) can be downloaded immediately. They are subsamples of the Scientific Use Files (SUFs) which allow researchers to explore data sets and build their code. These PUFs cannot be used for publications. For this, the SUF files are needed. SUF files are also made available but require a stricter two-step application process in which the organization of a researcher first has to be recognized as a research entity—a university, research institution, or research department in a public administration, bank, statistical institute, and so forth, after which a researcher can submit an application to receive the full microdata set.

In the MENA region, there is less of a tradition of making microdata available and few countries seem to provide public access to (anonymized) microdata. For example, Atamov et al. (2020) report that in 2019 only seven of the 20 countries in the region provided public or licensed access to household budget surveys which provide the source data on the basis of which the World Bank calculates its estimates of poverty (Table 1). To help with the advocacy of accessibility to microdata in MENA, it is important to

² See <https://unstats.un.org/unsd/dnss/gp/FP-Rev2013-E.pdf>.

³ <https://microdata.worldbank.org/>.

⁴ <https://www.dhsprogram.com/>.

⁵ <https://mics.unicef.org/surveys>.

⁶ <https://www.ilo.org/surveyLib/index.php/catalog/LFS/about>.

⁷ <https://usa.ipums.org/usa/>.

Table 1. Status of public and WB access to household budget surveys in MENA as of August 2019

 Public or licensed access

Djibouti
 Egypt, Arab Rep.
 Iran, Islamic Rep.
 Iraq
 Tunisia
 West Bank and Gaza
 Yemen, Rep.

Source: Atamanov et al. (2020).

have a more complete understanding of the state of microdata access, beyond the availability of household budget surveys. We do so in the remainder of this article.

3. Examining Microdata Openness in MENA

3.1. Microdata categories

The Sustainable Development Goals (SDG) provide a global agenda for disaggregated data needed to track global development progress.⁸ To facilitate reporting on the SDGs, a broad range of data is needed. The 2015 Data for Development Report recommends that countries derive their data from a total of eight sources: (a) census data; (b) household surveys; (c) agricultural surveys; (d) administrative data; (e) civil registration and vital statistics; (f) economic statistics, including labor force and establishment surveys and trade statistics; (g) geospatial data, and (h) other environmental data. In this article, we focus on microdata and examine access on NSOs' websites across four data categories (a) establishment data, (b) price data, (c) individual/household data, and (d) census data.⁹ In each data category, the degree of data accessibility provided to data users is examined by aiming to access the relevant data sets. (See Figure 1 for a snapshot of the data categories and subcategories.) We turn to discussing the representative data sets in each of the data categories in the paragraphs that follow.

In the *establishment data category*, we consider two types of surveys: enterprise surveys and annual surveys of industry—these surveys are the underlying source data for GDP estimation and are used to estimate labor market demand. In the *price data category*, we consider surveys of consumer prices (used to calculate the consumer price index, CPI) and surveys of producer prices (used to calculate the producer price index, PPI). In this category we do not look for the availability of each data point, though such information would be informative, but for the availability of price data for product categories or at the item level.

We divide the *individual/household data category* into three subcategories as follows: *consumption (welfare) data*; *labor force data and health data* and consider various possibilities under each subcategory. For the *consumption data sub-category* we look for household budget surveys, household income surveys, and/or living standard measurement survey—these surveys are typically used to measure household spending and income and are the underlying source data used to estimate poverty statistics. For the *labor force data subcategory*, we consider labor force surveys which are the underlying source data used to monitor labor supply and estimate various labor market statistics including labor force participation rate and employment rate. For the *health data subcategory*, we consider two possibilities, Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS) or any equivalent which provides the source data to estimate key health statistics including fertility, mortality, nutritional status, and various disease incidences.

⁸ See <https://sdgs.un.org/2030agenda> (accessed 4 March 2021).

⁹ Given the relatively small size of the agricultural sector in many MENA countries, we refrain from assessing the availability of agricultural censuses.

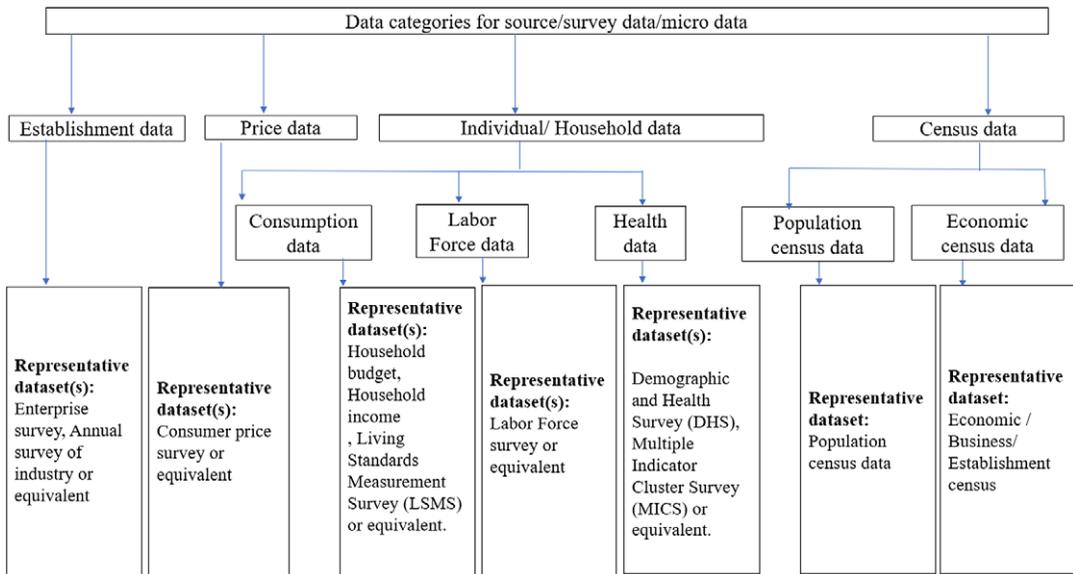


Figure 1. Data categories.
Source: Authors' illustration

Finally, we divide the *census data* category into two subcategories namely: *population and economic censuses*. Census data help define the structure and key characteristics of the population and economy and provide the framework needed for sampling different surveys. Censuses are rarely published in their entirety but many NSOs, including in the United States, Canada, and Britain, publish randomized 5–10% samples from their censuses.

3.2. Definition of recent microdata and classification of microdata accessibility

To allow for the possibility that microdata is not released because they have not been collected, we first establish the availability of recently collected data in each category, whereby recently is defined based on the data at hand. For establishment, consumption, labor force, and health surveys, we expect data to be collected at least once every 5 years. This is lenient: the 2016 State of Development Data Funding (SDDF) report published by the Global Partnership for Sustainable Development Data proposes a frequency of 2–3 years for health surveys, 5 years for consumption surveys, and annually for labor force and establishment surveys.¹⁰ The World Bank expects welfare surveys to be updated every 3 years. We expect price survey data to be collected multiple times annually—typically monthly—but examine NSOs for data within the past year. Census data is expected to be collected at least once every decade. Although the exercise of examining NSO websites for recent microdata was carried out between February and April 2021, we use year 2019 as the reference year. This is because of COVID-19-related disruption in data collection which often prevented face-to-face interviews from being conducted. Hence recent establishment, consumption, labor force, and health surveys are those carried out between 2014 and 2019 or later; recent price data are collected between 2018 and 2019 or later; recent census data are collected between 2009 and 2019 or later.

Once we have established that data has been collected recently, we assess whether the data is publicly accessible. For each data category, we classify microdata accessibility into four groups as follows:

1. *No coverage*: if no representative microdata was recently collected.
2. *No openness*: if representative microdata was recently collected but the data or a link to the data is not available on the website.

¹⁰ See <http://opendatawatch.com/knowledge-partnership/state-of-development-data-funding-2016/>.

3. *Satisfactory openness*: if representative microdata was recently collected and the data (or link) is available on website but is restricted, that is, users need to submit a request and/or register to be granted access to the data.
4. *Excellent openness*: if representative microdata was recently collected and the data (or link) is publicly available on website in machine-readable format for immediate download.

We differentiate between “satisfactory” and “excellent” openness because microdata openness is examined from the perspective of the data user. From this perspective “excellent openness” is ideal because there is no wait time for data users to access available data. However, “satisfactory openness” is acceptable because it is okay for data guardians to require registration, authorization, and clearance before releasing data to prevent unauthorized access. The best scenario being where following satisfactory registration, access to the data is granted automatically.

3.3. *Implementation exercise of microdata classification in MENA*

The exercise of visiting the websites of the NSOs¹¹ and international organizations¹² to examine microdata accessibility was designed to be cost-effective and easy to apply to countries in MENA and beyond. To prevent bias and ensure accuracy and replicability, the exercise is implemented in a 3-step process by a team comprised of three core researchers with language competencies in English, Arabic, and French—major languages in the MENA region.

1. *Step 1*: Each of the three researchers in the research team independently visits the websites to classify microdata accessibility for all data categories into one of the four groups discussed in Section 3.2.
2. *Step 2*: Researchers meet to discuss their independent findings from step 1 and reconcile any differences. When a researcher finds representative microdata for the categories covered on a public-facing website that other researchers do not, the reconciliation process involves providing a link to the portion of the website where the data was found. The research team visits the link as a group to verify the data and update the result.
3. *Step 3*: The updated result from the research team in step 2 is sent for peer review. The peer review is done by World Bank colleagues who work as country/poverty economists and are familiar with the coverage of microdata in the MENA region. Like in step 2, when country economists are aware about representative microdata for the data categories covered on a public channel not captured by the research team, they provide the link to the data. The research team then verifies the data and updates the result.

Although the methodology described here has only been implemented for MENA countries, it can be scaled globally. To minimize cost, the implementation exercise for a global scale-up may be modified. Since step 3 of the implementation exercise involves a review by credible peers to validate the results from steps 1 and 2, only one researcher may implement step 1. In this case, step 2 will be eliminated. If this modification occurs, it is preferable that the researcher chosen to implement the classification exercise for a given region is multilingual in the major languages in the region.

3.4. *Microdata coverage in MENA*

Before data can be made available, it must be collected. Hence, we first determine the collection of recent data for each data category. On NSO websites, we do this by searching explicitly in the “survey/ data section” and/or microdata dashboard/library or implicitly like checking for any mention or reference to the

¹¹ See Supplementary Table A1 for list of NSOs in MENA and their websites.

¹² See Supplementary Table A2 for list of the websites of international microdata repositories.

data in a report, summary table, survey calendar/event schedule, and/or announcement page. We also check international microdata libraries to determine recent collection of representative data for each data category. In [Table 2](#), we summarize the results from the exercise. At the start, we expected to be able to identify a total of 140 microdata sets—seven data subcategories across 20 countries: eventually, we could verify that around half (83) of these microdata sets had been collected.

All MENA NSOs except the Republic of Yemen collect price data for their CPI and or PPI and about half are up to date with respect to their labor force, consumption, health, and census data. Twelve NSOs report recent surveys in the Labor Force microdata category and 14 recent surveys are found in the consumption data category. For establishment data, only a quarter of NSOs (5) collected such data recently: the 2018 Kuwait's Annual Survey of Establishments, 2016 Malta's Labor Cost Survey, 2019 Morocco's National Business Survey, 2019 Saudi Arabia's Economic Indicator Survey, and the 2018 Palestinian Economic Survey Series.

The NSOs of Saudi Arabia and West Bank and Gaza are up to date with their microdata collection across all data categories—seven out of seven recent microdata sets expected. They are closely followed by the Arab Republic of Egypt, Jordan, and Morocco which collected data for six out of the seven recent microdata sets expected. By contrast Lebanon, the Syrian Arab Republic, and the United Arab Emirates, only report two recent microdata sets.

3.5. *Accessibility of microdata nationally*

Having collected data does not necessarily imply that the (anonymized) microdata is publicly accessible. For all the data categories, we examine NSO websites¹³ for accessibility of the microdata indicated to have been collected. This is reported in [Table 3](#), where entries are only provided whereas in [Table 2](#) it was indicated that a recent microdata set has been collected. Of the 83 microdata sets, only 23 are accessible to a user visiting NSO websites. Of these, only 12 can be downloaded immediately, seven of which are price data for product categories. The remaining five are the 2018–2019 Lebanon Labor Force and Household Conditions Survey (LFHLCs), the 2014 Morocco National survey on Household Consumption and Expenditure, the 2015 Tunisia National survey on budget, consumption, and household living standard,¹⁴ the 2017 Tunisia National Population and Employment Survey and a subset of the 2014 population census microdata for Morocco. All others require prior registration.

We conclude that NSOs in the MENA region face two major challenges with respect to microdata. Except for price data which are up to date across the board, in all other data categories only about half the countries have up to date microdata sets on which they can draw. Note that this is a very lenient interpretation as microdata sets collected as far back as 2014 are counted toward being up to date. If a stricter definition of up to date were used, the number of countries with recent data would fall lower.

With respect to making the data that has been collected publicly available, NSOs in the region face even more challenges. Only 16 microdata sets, out of a potential 140 that ideally would have been collected, and 83 that have been collected, are downloadable from NSO websites. Consequently, and depending on the definition used, only 10–20% of the expected microdata are available to the public on NSO websites. Within the health data category, none of the NSOs makes microdata publicly available.

3.6. *Accessibility of microdata internationally*

We have not (yet) considered non-NSO websites and/or repositories from which a country's data could be available. We excluded these on purpose in [Table 3](#) as data users—most of whom would be nationals, should be able to access data for their country from their national NSO (or other national agencies: health surveys, for instance, are at times collected and published by Ministries of Health). Yet there are instances

¹³ Microdata available nationally may also be on the platforms of other national agencies besides the NSO. If this is the case, we examine the website of the national agency as well.

¹⁴ For the 2015 Tunisia Budget survey, it is important to note that not all variables are included in the microdata set available for immediate download.

Table 2. Status of survey data (with year collected) in MENA on NSOs website and other public channels

Economy/data category	Establishment survey	Price survey	Consumption survey	Labor force survey	Health survey	Population census	Economic census	Total data categories with recently collected data per country
Algeria	No	Yes (2021)	No	Yes (2015)	No	No	Yes (2011)	3/7
Bahrain	No	Yes (2021)	Yes (2014/2015)	No	Yes (2018)	Yes (2020)	No	4/7
Djibouti	No	Yes (2021)	Yes (2017/2018)	No	No	Yes (2011)	No	3/7
Egypt, Arab Rep.	No	Yes (2021)	Yes (2017/2018)	Yes (2020)	Yes (2014)*	Yes (2017)	Yes (2017)	6/7
Iran, Islamic Rep.	No	Yes (2020)	Yes (2019/2020)	Yes (2018)	No	Yes (2016)	No	4/7
Iraq	No	Yes (2019)	Yes (2017/2018)*	No	Yes (2018)*	No	No	3/7
Jordan	No	Yes (2021)	Yes (2017)	Yes (2020)	Yes (2017/2018)	Yes (2015)	Yes (2018)	6/7
Kuwait	Yes (2018)	Yes (2020)	Yes (2019/2021)	Yes (2015)	No	No	No	4/7
Lebanon	No	Yes (2020)	No	Yes (2018/2019)	No	No	No	2/7
Libya	No	Yes (2020)	No	No	Yes (2014)	Yes (2012)	No	3/7
Malta	Yes (2016)	Yes (2021)	Yes (2015)	Yes (2020)	No	Yes (2011)	No	5/7
Morocco	Yes (2019)	Yes (2021)	Yes (2014)	Yes (2019)	Yes (2018)	Yes (2014)	No	6/7
Oman	No	Yes (2020)	No	No	Yes (2014)*	Yes (2020)	Yes (2020)	4/7
Qatar	No	Yes (2020)	No	Yes (2019)	No	Yes (2015)	Yes (2015)	4/7
Saudi Arabia	Yes (2019)	Yes (2021)	Yes (2018)	Yes (2020)	Yes (2017/2018)	Yes (2010)	Yes (2010)	7/7
Syrian Arab Republic	No	Yes (2019)	No	No	No	No	Yes (2019)	2/7
Tunisia	No	Yes (2021)	Yes (2015)	Yes (2017)	Yes (2018)*	Yes (2014)	No	5/7
United Arab Emirates	No	Yes (2020)	Yes (2019)	No	No	No	No	2/7
West Bank and Gaza	Yes (2018)	Yes (2021)	Yes (2017)	Yes (2019)	Yes (2019/2020)*	Yes (2017)	Yes (2017)	7/7
Yemen, Rep.	No	No	Yes (2014)	No	No	Yes (2014)	Yes (2014)	3/7
Total economies	5/20	19/20	14/20	12/20	10/20	14/20	9/20	
indicates collection of recent data for each data category Total recent data indicated to have been collected across all economies								

Note. Evidence that a survey was collected can be explicit like in a “survey section” of the website or “implicit” like in a report, summary table, and/or any mention or reference to the data on the website.

*Indicates instances where collection of recent microdata was not indicated on NSOs website, but the research team discovered it on an external website. These include Iraq: Rapid welfare monitoring survey SWIFT 2017/2018 downloadable from <https://microdata.worldbank.org/>, Egypt (2014) downloadable from <http://www.dhsprogram.com/> and Iraq MICS 2018, Oman MICS 2014, Tunisia: MICS 2018, West Bank and Gaza (Palestine) MICS 2019/20 downloadable from <https://mics.unicef.org/surveys>.

where microdata sets are available in international repositories, even while they are unavailable locally. For example, the National Statistics Office of Malta makes microdata from some surveys available to Eurostat which then makes it available to data users upon successful registration and application for the data—these data may not be available on the website of Malta’s National Statistical Office.¹⁵ Also international data repositories such as the ERF Open Access Micro Data Initiative (OAMDI) launched in 2013 serve as the only archive for various microdata for countries of the ERF region.¹⁶ Hence, to complete the picture of microdata accessibility for each country, we explore what is available in international microdata libraries. We do so by visiting the WB microdata library, the web-portals of the MICS and DHS surveys as well as the microdata libraries maintained by the International Household Survey Network (IHSN), IPUMS, Eurostat, and the ERF.¹⁷ The results from this exercise are summarized in Table 4. Overall microdata accessibility improves by around 50% when we consider international accessibility in addition to national accessibility—from 23/140 to 34/140. Some countries like Iraq and Malta which had *no microdata openness* for all data categories when we examined only NSOs website now have *satisfactory data openness* for some data categories. However, despite these improvements, microdata accessibility in MENA remains poor.

4. Opportunities for NSOs to Improve Microdata Accessibility

Collecting microdata is costly, which may be one reason why relatively few microdata sets are collected in the MENA region. While the frequency with which microdata are collected may not change overnight, our search for microdata revealed opportunities for NSOs to improve their data accessibility at almost no additional cost. Some MENA NSO makes price indices for product categories available in PDF format even though such information would be relevant to a host of users in machine-readable format. Almost all MENA NSOs possess recent population census data, but few make them publicly available. The exceptions are Morocco and the Islamic Republic of Iran where a sample of anonymized individual and household level data is available for download. Additional suggested practices that can improve accessibility of microdata on NSO websites are outlined below.

4.1. Suggested practice 1: Provide an English version of the website

While the primary audience for NSO statistics is nationals, many potential data users live abroad. Since English is understood by majority of people in almost every region of the world, it is best practice for NSOs to make available an adequate English version/translation of their website. At present, not all MENA NSOs have an adequate English version of their website. For instance, an English version of the website of the Islamic Republic of Iran NSO exists, but several datasets available on the Persian version of the website are not available on the English version. This includes the consumption (welfare), labor force survey as well as the population census data reported to have satisfactory openness in Table 4. Consequently, non-Persian speakers would have difficulty identifying the wealth of data that is available, particularly as the Islamic Republic of Iran is exemplary in providing data access. All recent, available microdata sets are downloadable from the website, some like the household budget surveys at an annual frequency.

4.2. Suggested practice 2: Provide a microdata catalog, data tab, and a search button on website landing page

Given the multiplicity of information that is typically available on an NSO’s website, ensuring a good routing through the website is critical. For primary microdata users, a data tab and/or microdata catalog

¹⁵ For example, Malta National Statistics Office sends microdata from its European Statistics on Income and Living Conditions Survey (EU-SILC), Household Budgetary Survey as well as Labor Cost Survey—Enterprise survey—to Eurostat where it can be requested by data users.

¹⁶ See <https://erf.org.eg/oamdi/>.

¹⁷ See Supplementary Table A2 for the links to these microdata libraries.

Table 3. Publicly accessible microdata sets on website of MENA NSOs

Economy/data category	Establishment survey	Price survey	Consumption survey	Labor force survey	Health survey	Population census	Economic census	
Algeria	—	No openness*	—	No openness	—	—	No openness	
Bahrain	—	Excellent	No openness	—	No openness	No openness	—	
Djibouti	—	Excellent	Satisfactory	—	—	No openness	—	
Egypt, Arab Rep.	—	Excellent	Satisfactory	No openness	No openness	No openness	Satisfactory	
Iran, Islamic Rep.	—	No openness	Satisfactory	Satisfactory	—	Satisfactory	—	
Iraq	—	No openness*	No openness	—	No openness	No openness	—	
Jordan	—	Excellent	No openness	No openness	No openness	No openness	No openness	
Kuwait	No openness	No openness*	No openness	No openness	—	—	—	
Lebanon	—	Excellent	—	Excellent	—	—	—	
Libya	—	No openness*	—	—	No openness	No openness	—	
Malta	No openness	No openness*	No openness	No openness	—	No openness	—	
Morocco	No openness	No openness	Excellent	No openness	No openness	Excellent	—	
Oman	—	Excellent	—	No openness	No openness	No openness	—	
Qatar	—	No openness*	—	—	—	No openness	No openness	
Saudi Arabia	No openness	Excellent	No openness	No openness	No openness	No openness	No openness	
Syrian Arab Republic	—	No openness*	—	—	—	—	No openness	
Tunisia	—	No openness	Excellent	Excellent	No openness	No openness	No openness	
United Arab Emirates	—	No openness	No openness	—	—	—	—	
West Bank and Gaza	Satisfactory	No openness	Satisfactory	Satisfactory	No openness	Satisfactory	Satisfactory	
Yemen, Rep.	—	—	No openness	—	—	No openness	No openness	
Total economies with some degree of accessibility of microdata for data category	1/20	7/20	6/20	4/20	0/20	3/20	2/20	
Total surveys/ census with some degree of accessibility of microdata on NSOs website								23/140

Note. “—” indicates up to date microdata have not been collected.

*Price data available for product categories but not in machine-readable format.

Table 4. Openness of recent source/survey/micro data on public-facing websites of MENA NSOs and international microdata libraries

Economy/data category	Establishment survey	Price survey	Consumption survey	Labor force survey	Health survey	Population census	Economic census
Algeria	—	No openness	—	No openness	—	—	No openness
Bahrain	—	Excellent (NSO)	No openness	—	No openness	No openness	—
Djibouti	—	Excellent (NSO)	Satisfactory (NSO, WB)	—	—	No openness	—
Egypt, Arab Rep.	—	Excellent (NSO)	Satisfactory (NSO, ERF)	Satisfactory (ERF*)	Satisfactory (WB, DHS)	No openness	Satisfactory (NSO)
Iran, Islamic Rep.	—	No openness	Satisfactory (NSO)	Satisfactory (NSO)	—	Satisfactory (NSO)	—
Iraq	—	No openness	Satisfactory (WB)	—	Satisfactory (WB, MICS)	—	—
Jordan	—	Excellent (NSO)	No openness	Satisfactory (ERF*)	Satisfactory (WB, DHS)	No openness	No openness
Kuwait	No openness	No openness	No openness	No openness	—	—	—
Lebanon	—	Excellent (NSO)	—	Excellent (NSO)	—	—	—
Libya	—	No openness	—	—	No openness	No openness	—
Malta	No openness	No openness	Satisfactory (Eurostat)	Satisfactory (Eurostat)	—	No openness	—
Morocco	No openness	No openness	Excellent (NSO)	No openness	No openness	Excellent (NSO, IPUMS)	—
Oman	—	Excellent (NSO)	—	No openness	Satisfactory (WB, MICS)	No openness	—
Qatar	—	No openness	—	—	—	No openness	No openness
Saudi Arabia	No openness	Excellent (NSO)	No openness	No openness	No openness	No openness	No openness
Syrian Arab Republic	—	—	—	—	—	—	No openness

(Continued)

Table 4. Continued

Economy/data category	Establishment survey	Price survey	Consumption survey	Labor force survey	Health survey	Population census	Economic census
Tunisia	—	No openness	Excellent (NSO)	Excellent (NSO)	Satisfactory (MICS)	No openness	No openness
United Arab Emirates	—	No openness	No openness	—	—	—	—
West Bank and Gaza	Satisfactory (NSO)	No openness	Satisfactory (NSO)	Satisfactory (NSO, ERF)	Satisfactory (MICS)	Satisfactory (NSO, IPUMS)	Satisfactory (NSO)
Yemen, Rep.	—	—	No openness	—	—	No openness	No openness
Total economies with some degree of microdata accessibility for each data category	1/20	7/20	8/20	7/20	6/20	3/20	2/20
Total surveys/ census with some degree of microdata accessibility on NSO, WB, IHSN, IPUMS, Eurostat, and/or DHS and MICS websites							

34/140

Note. “—”, microdata not collected; green, microdata available nationally only; red, microdata available internationally only; blue, microdata available both nationally and internationally. In Table 2, we report that Egypt NSO’s website indicates collection of LFS for 2020, it is important to note that the available most recent LFS microdata for Egypt on ERF portal at the time of access is for 2019. WB, IHSN microdata library as well as IPUMS, Eurostat, DHS, and MICS data was accessed on August 8, 2021 while ERF data portal was accessed on January 7, 2022.

that present all data available on the website is a useful tool. This will make microdata on the website easy to find and download. Egypt for instance has a tab “MetaData” on its landing page that leads visitors to a central data catalog. This is very helpful for website visitors interested in the country’s data. Some countries go even further. UAE’s open data portal allows users to search for data by the organization within UAE that owns the data. More generally, a search button is important to facilitate finding relevant information on the website and ultimately ensuring a favorable user experience. To date, not all MENA NSOs include a data tab, search button, and/or microdata catalog on the landing page of their website. There is a freely available, World Bank-approved microdata cataloging tool available at <http://nada.ihsn.org/> that can serve as a guideline for NSOs.

4.3. Suggested practice 3: Provide links to other websites with country’s data

Earlier we reported that not all microdata sets are hosted on the NSOs websites and about half of the publicly available microdata sets are accessible through international repositories. Where this is the case, providing a link to the websites with the relevant country data is best practice. Microdata available in other Microdata Library of the WB, IHSN, Eurostat, MICS, and IPUMS data can be easily linked on the NSO’s website whether the NSO owns all the data available on these websites. Djibouti sets a good example for providing external links to its country’s data. At the time of the study, on the landing page of the website of the National Institute of Statistics of Djibouti, there is a tab named “database” with three dropdown tabs as follows (a) survey data. (b) Open data (c) key indicators. The survey data tab links to the World Bank’s microdata library.

4.4. Suggested practice 4: Provide clarity for requesting restricted data

In the classification of microdata accessibility in Section 3.2, we differentiate between two classes of microdata accessibility—“satisfactory openness” and “excellent openness” where the former involves a situation where authorization and/or registration is required before a data user can access available data and the later a situation where microdata is available for immediate download on the website. As discussed earlier, “excellent openness” is ideal from the perspective of a data user, however requiring registration, authorization, and clearance before data is released by data guardians is acceptable. When microdata has “satisfactory openness,” it is important that NSO’s provide clarity regarding the steps that need to be followed to gain access, that access is granted within a reasonable period of time and that granting permission is “rule based” and not dependent on ad hoc criteria. However, for some MENA NSOs for which satisfactory openness is reported in Table 3, the website indicates that the data is available upon request without clear instructions about the steps needed to obtain the data. The best scenario for “satisfactory openness” is where following satisfactory registration, access to the data is granted automatically. This is standard practice for international organizations such as the WB, MICS, DHS, IPUMS, and ERF.

Apart from these best practices that could be implemented by any NSO at a negligible expense, we also strongly advocate to close the microdata gap by investing in regular microdata collection.

5. Open Data Inventory and Data Accessibility in MENA

5.1. The open data inventory

The evidence in Section 3 shows that the availability and accessibility of microdata in the MENA region are very constrained. Yet MENA ranks highly on the Open Data Inventory (ODIN)¹⁸ published by Open Data Watch.

The ODIN covers 178 countries in its 2018/2019 version including 17 MENA countries and 187 countries in its 2020/2021 version including all 20 MENA countries. Additionally, ODIN also has

¹⁸ See Open Data Watch—Open Data Inventory, <http://www.opendatawatch.com>.

a substantial proportion of its elements assessing data accessibility or openness—it assesses the coverage and openness of data available on National Statistics Offices (NSOs) websites based on ten elements across two dimensions—coverage and openness. Five of the 10 elements measure data coverage, that is, the degree to which data is available, and while the others measure access/openness, that is, the degree to which available data is accessible. Each of the five elements in the coverage dimension is assessed as follows: representative indicators are available and are disaggregated appropriately; data are available for the preceding 5 years; data are available for the preceding 10 years; data are disaggregated at the first administrative level and data are disaggregated at the second administrative level. Each of the five elements assessed in the ODIN data accessibility/openness dimension is assessed as follows: machine readability; nonproprietary; download options; metadata available and terms of use. All the elements in the ODIN coverage and openness dimensions are assessed across several data categories and data dimensions. Given the foregoing, ODIN is clearly comprehensive and covers a substantial amount of MENA countries. However, ODIN’s methodological guide mention that the terms “data,” “statistics,” and “indicators” are used interchangeably.¹⁹ These three terms are not synonyms, as data in the ODIN context does not include microdata. ODIN’s measure only captures access to generated statistics and indicators. Hence it is possible that ODIN suggests data accessibility where in fact there is no access to microdata. In the next section, we discuss recent ODIN scores for MENA vis-à-vis the results of microdata accessibility presented in Section 3.

5.2. Performance of MENA countries on the ODIN

Countries in the MENA region perform rather well on the ODIN. As shown in Figure 2, in the 2018/2019 ODIN, MENA generally does better than Sub-Saharan Africa, and is on par with South Asia, Latin America and East Asia, and the Pacific.

The ODIN is well established and recognized and when the World Bank (WB) launched, beginning in 2021 its own Statistical Performance Indicator (SPI), it relies on data provided by ODIN to complete its subperformance indicators on data access (Dang et al. 2021). How can our assessment of very limited data access in the MENA region and ODIN’s assessment differ so much? There are two possible explanations. Microdata availability and access in MENA are on par with that in other regions. This is a possibility. Our intuition is, however, that this is less likely as, for example, many countries in Latin America have very well-developed microdata programs that pride themselves in the public accessibility that they provide. Instead, we are convinced this has more to do with the fact that the ODIN measures data access based on the ability of NSO’s to make available summary statistics, data that represent a summary measure derived from survey/source/microdata but does not capture the public release of (anonymized) microdata.

This can be illustrated by the availability of “poverty statistics” which ODIN assesses through the availability of two indicators (a) the poverty rate and (b) the distribution of income by deciles or Gini coefficient. An NSO that publishes these statistics, without making available the underlying household consumption/expenditure/income survey, gets a full score on the ODIN indicator, irrespective of when the microdata on which these statistics have been based are collected and irrespective of whether these microdata are publicly accessible. Thus, Oman which provides no public access to its household budget surveys, receives a perfect ODIN score on “poverty statistics.” Lebanon does not obtain a perfect score but scores an average (45 out of 100 points). Yet not only is the microdata on which this score is based inaccessible, the last Household Budget Survey on which the official poverty estimates are based dates from 2011. Clearly, any poverty statistics that are officially released are outdated and of limited relevance today, particularly considering the economic decline the country is experiencing.

ODIN’s measures are of value because of the meticulous and transparent way in which it documents its scores. As discussed earlier, it is based on 10 elements across two dimensions—coverage and openness assessed across several data categories and data dimensions. But the data under these categories are not

¹⁹ See https://docs.google.com/document/d/1MBK0hN6MoQrii7_E1bmRXmsUcE8Fbb-Q32nxm8d8qTw/edit (accessed 18 October 2021).

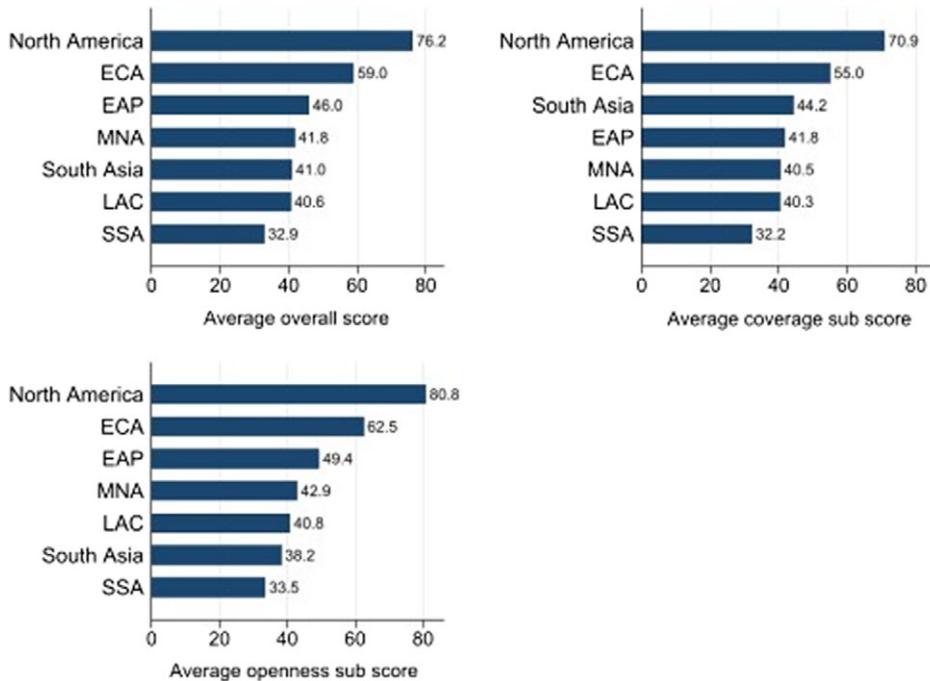


Figure 2. Regional comparison of ODIN scores, coverage sub scores, and openness subscore. EAP, East Asia and Pacific; ECA, Europe and Central Asia; LAC, Latin America and Caribbean; MNA, Middle East and North Africa; SSA, Sub-Saharan Africa. Source: Author's compilation using 2018/2019 ODIN data from Open Data Watch—Open Data Inventory <http://www.opendatawatch.com>.

required to be at a high level of disaggregation, that is, they are not required to be individual/household level data—regional and subregional level data satisfies ODINs scoring guidelines. ODIN's scores present an excellent basis for data users interested in summary statistics, scoring countries on data availability, degree of disaggregation, and the ability to download data in machine-readable format. However, given the evidence presented in Section 3, the usefulness of the ODIN by NSOs and development partners as part of a measuring rod for the development of the statistical system, improving data access, and encouraging dialogue with data users is limited for MENA without a complementary indicator measuring microdata access. Thus, an indicator focusing on microdata openness building on the methodology discussed in Section 3 combined with the ODIN will give a more balanced view of data openness in MENA.

6. Conclusion

Evidence driven decision-making requires trusted statistics. For statistical offices, this straightforward statement means that core microdata is regularly collected, and that the data are made publicly available. For this article, we assessed the availability of anonymized microdata sets for the MENA region across seven categories: population and economic censuses, price statistics and consumption, labor, establishment, and health surveys. We visited the websites of each NSO in the region as well as international data libraries and checked whether these core microdata sets had been collected recently and whether they are available for download (either immediately or after registration). We used a lenient definition of “recent” and required census data be not older than 10 years, survey data no more than 5 years old and price data to have been collected at least once a year. Because our website visits took place during the COVID-19

epidemic, during which face-to-face data collection came to a standstill, we used 2019 as benchmark year, implying that any censuses done after 2009 and surveys done after 2014 were considered up to date.

Our findings are threefold. First price data are typically collected (often at a monthly basis), but census and survey data are often out of date. Only 14 out of 20 countries are current on their population census; nine out of 20 are up to date on their economic census. Only five out of the 20 countries carried out an establishment survey recently and about half the countries are up to date with respect to their health, labor force, and consumption surveys (having been completed in 10, 12, and 14 countries, respectively). The implication is that in almost half the cases, no or outdated microdata is used to produce core statistics including National Accounts and SDG reporting.

Our second finding is that only in few instances where microdata has been collected, they are made publicly accessible. Of the 140 potential microdata sets, we looked for (seven data categories in 20 countries) 83 had been collected and as few as 34 were accessible. Remarkably, of these 34 about a third are not accessible through the website of the NSO; they can only be downloaded from international microdata repositories. Our third finding is that recent microdata is scarce in MENA. Summary statistics are generally available—as evidenced by the ODIN. However, many of these statistics are necessarily based on outdated microdata and decision-makers relying on such information would need to consider them with care.

Our findings show that an indicator focusing on microdata accessibility is needed in MENA. This indicator together with the ODIN will give a robust picture of data openness in MENA. Although, ODIN's measures are of value because of the meticulous and transparent way in which it documents its scores, it focuses on summary statistics and not on the underlying microdata from which the statistics are calculated. Focusing on the availability of recent microdata would make avoidable the situation where decision-makers are informed by summary statistics that no longer reflect their economic and social realities. If such microdata were also made publicly available, it would further improve statistical transparency while also soliciting researchers to contribute their knowledge to help answer the pressing development questions of our time.

Acknowledgments. The authors would like to thank Henry Gannat and Federica Alfani for excellent research assistance. The authors would also like to thank Umar Serajuddin, Hai-Anh Dang, Brian Stacy, and Daniel Mahler for providing suggestions on earlier drafts of this article. Special thanks to several colleagues from the Poverty and Equity Global Practice at the World Bank who participated in the peer review process to validate results from the data collection exercise. Finally, the authors acknowledge reviewers and participants of Data for Policy conference 2021 for their useful comments to an earlier version of the article.

Funding Statement. This work is a product of the staff of the Poverty and Equity Global Practice of The World Bank.

Competing Interests. The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of the Executive Directors of The World Bank or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

Author Contributions. Conceptualization: U.E.E.-M., J.H.; Data curation: U.E.E.-M.; Data visualization: U.E.E.-M., J.H.; Methodology: U.E.E.-M., J.H.; Writing—original draft: U.E.E.-M., J.H. All authors approved the final submitted draft.

Data Availability Statement. Data availability is not applicable as all the new data collected are presented within the article.

Supplementary Materials. To view supplementary material for this article, please visit <http://doi.org/10.1017/dap.2022.24>.

References

- Arezki R, Lederman D, Abou Harb A, El-Mallakh N, Fan RY, Islam A and Zouaidi M (2020) *How Transparency Can Help the Middle East and North Africa*. Washington, DC: World Bank. <https://doi.org/10.1596/978-1-4648-1561-4>**
- Atamanov A, Tandon S, Lopez-Acevedo G, Vergara B and Mexico A (2020) *Measuring Monetary Poverty in the Middle East and North Africa (MENA) Region. Data Gaps and Different Options to Address Them*. World Bank Policy Research Paper # 9259, May 2020.**

- Brackfield D** (2011) *OECD Work on Measuring Trust in Official Statistics*. Int Statistical Inst.: Proceedings of the 58th World Statistical Congress, Dublin (session STS070).
- Cady J** (2005) Does SDDS subscription reduce borrowing costs for emerging market economies? *IMF Staff Papers* 52(3), 503–517.
- Dang H, Jolliffe D and Carletto C** (2019) Data gaps, data incomparability, and data imputation: A review of poverty measurement methods for data scarce environments. *Journal of Economic Surveys* 33(3), 757–797. <https://doi.org/10.1111/joes.12307>
- Dang HH, Pullinger J, Serajuddin U and Stacy B** (2021) *Statistical Performance Indicators and Index: A New Tool to Measure Country Statistical Capacity*. Policy Research Working Paper; No. 9570. World Bank, Washington, DC. Available at <https://openknowledge.worldbank.org/handle/10986/35301>. Accessed February 1, 2022.
- Das J, Do Q-T, Shaines K and Srikant S** (2013) U.S. and them: The geography of academic research. *Journal of Development Economics* 105(1), 112–130. <https://doi.org/10.1016/j.jdeveco.2013.07.010>