Policy Research Working Paper 9892

Microdata Collection and Openness in the Middle East and North Africa (MENA)

Introducing the MENA Microdata Access Indicator

Uche Eseosa Ekhator-Mobayode
Johannes Hoogeveen

WORLD BANK GROUP
Poverty and Equity Global Practice
December 2021

Abstract

This paper 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. Following the exercise, a new microdata access indicator that measures the degree of opennes of microdata and the ease with which microdata users can understand and navigate the websites of national statistical offices is presented. 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 As a consequence, many summary statistics, including national accounts and welfare estimates, are outdated and of limited relevance to decision makers.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. National statistical offices in the region should make considerable improvements to the outlook of their websites to make them more user friendly. Specifically, microdata libraries and updated survey calendars should be a standard feature of the websites to ensure easy access to available microdata.

This paper is a product of the Poverty and Equity Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at uekhator@worldbank.org and jhoogeveen@worldbank.org.

Microdata Collection and Openness in the Middle East and North Africa (MENA): Introducing the MENA Microdata Access Indicator

Uche Eseosa Ekhator-Mobayode[*1] and Johannes Hoogeveen[*]

Key words: National Statistical Offices, Statistical Indicators, Data Openness, Microdata, MENA

JEL: C18, H00, I00, O1, O53

---

* MENA Poverty and Equity Global Practice. 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 paper. 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.
[1] Corresponding author. Email: uekhator@worldbank.org

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).

In this paper, 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 2078 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 23 Feb 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.[2]

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 data base 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%.

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

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

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 micro data 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 micro data.

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 and Eurostat. 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 do two things. First, we make suggestions aimed at improving NSOs' ability to provide up to date microdata online. Second, we present a new Microdata Access Indicator (MAI) that measures the degree of openness of available microdata across MENA countries and the degree to which microdata users can access, understand, and navigate the websites of MENA NSOs. The new MENA MAI provides insights on microdata accessibility in MENA and complements the Open Data Access Indicator (ODIN)[3] published by Open Data Watch that focuses on NSOs' ability to provide access to produced statistics and indicators (as opposed to anonymized raw data). Together, both the MENA MAI and the ODIN provide a robust picture of data accessibility in MENA and can over time serve as a useful tool for NSOs and development partners to measure the development of the statistical system and advocate for greater data transparency. They can also be used to encourage dialogue between NSOs, development partners and data users. The findings from the MENA MAI demonstrate 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. The findings also show that there is room for improvement with regards to the outlook of the websites of many MENA NSOs – NSOs should make microdata libraries and updated survey calendars a standard feature of their websites so that microdata users do not spend unnecessary time searching for available microdata.

The rest of the paper is structured as follows: the next section explores in greater depth the intersection between public trust in official statistics and data transparency. Section 3 describes the exercise of visiting MENA NSOs' websites, discusses the results from the exercise and offers

[3] Open Data Watch - Open Data Inventory http://www.opendatawatch.com

some suggestions for progress based on observations made by the research team. Section 4 examines existing indicators measuring data accessibility and discusses results from section 3 vis-à-vis the Open Data Access Indicator (ODIN). Section 5 presents the new microdata access indicator for MENA while 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 ten 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 6 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*." [4] 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,[5] the DHS[6] and MICS[7] websites, the labor force surveys curated by the ILO,[8] and IPUMS[9] 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

[4] See https://unstats.un.org/unsd/dnss/gp/FP-Rev2013-E.pdf
[5] https://microdata.worldbank.org/
[6] https://www.dhsprogram.com/
[7] https://mics.unicef.org/surveys
[8] https://www.ilo.org/surveyLib/index.php/catalog/LFS/about
[9] https://usa.ipums.org/usa/

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 etc., 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 7 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 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 paper.

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

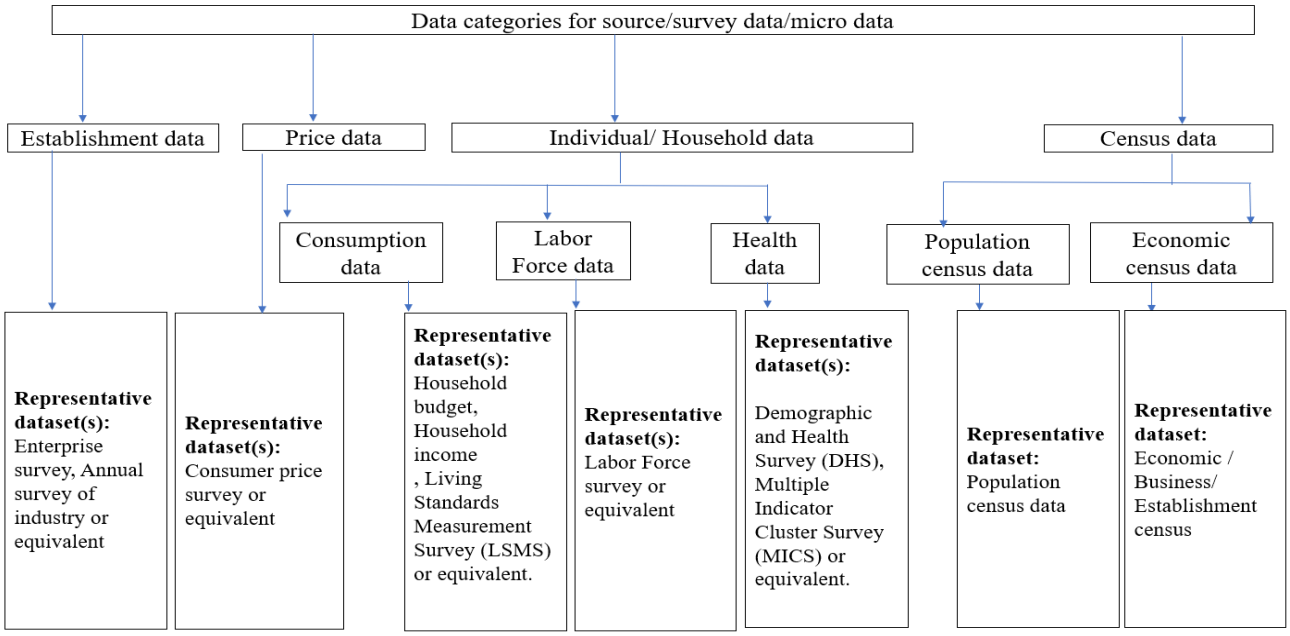
Public or licensed access
Djibouti
Egypt, Arab Rep.
Iran, Islamic Rep.
Iraq
Tunisia
West Bank and Gaza
Yemen, Rep.
Source: Atamanov et al. 2020

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.[10] 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: (i) census data; (ii) household surveys; (iii) agricultural surveys; (iv) administrative data; (v) civil registration and vital statistics; (vi) economic statistics, including labor force and establishment surveys and trade statistics; (vii) geospatial data and (viii) other environmental data. In this paper, we focus on microdata and examine access on NSOs' websites across four data categories (1) establishment data, (2) price data, (3) individual/household data, and (4) census data.[11] 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.

Figure 1: Data categories



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

[10] See https://sdgs.un.org/2030agenda, accessed 4 March 2021.

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

(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 item level price series.

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 sub-category*, 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 sub-category*, we consider two possibilities, demographic and health surveys and multiple indicator cluster survey or any equivalent which provides the source data to estimate key health statistics including fertility, mortality, nutritional status, and various disease incidences.

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 are not released because they have not been collected recently, 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.[12] The World Bank expects welfare surveys to be updated every three 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 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 considered to be those

[12] See http://opendatawatch.com/knowledge-partnership/state-of-development-data-funding-2016/

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 4 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.
3. **Satisfactory openness:** if representative microdata was recently collected and the data (or link) is available on website but is restricted i.e., 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 NSOs[13] and international organizations[14] 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 3 core researchers with language competencies in English, Arabic and French – major languages in the MENA region.

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 4 groups discussed in section 3.2.

Step 2: Researchers meet to discuss their independent findings from step 1 and reconcile any differences. When a researcher finds a representative microdata for the categories covered on a public facing website that other researchers do not, the reconciliation process involves providing

[13] See Annex table A1 for list of NSOs in MENA and their websites.

[14] See Annex table A2 for list of the websites of international microdata repositories.

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 updates their result.

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 1 researcher may implement step 1. In this case, step 2 will be eliminated. If this modification occurs, it will be 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 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 results from the exercise. At the start, we expected to be able to identify a total of 140 micro data sets – 7 data subcategories across 20 countries: eventually we could verify that around half (81) of these microdata sets had actually been collected.

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	No	No	No	Yes (2011)	2/7
Bahrain	No	Yes (2021)	Yes (2014/15)	No	Yes (2018)	Yes (2020)	No	4/7
Djibouti	No	Yes (2021)	Yes (2017/18)	No	No	Yes (2011)	No	3/7
Egypt, Arab Rep.	No	Yes (2021)	Yes (2017/18)	Yes (2020)	Yes (2014) *	Yes (2017)	Yes (2017)	6/7
Iran, Islamic Rep.	No	Yes (2020)	Yes (2019/20)	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 (2017/18)	Yes (2015)	Yes (2018)	6/7
Kuwait	Yes (2018)	Yes (2020)	Yes (2019/21)	Yes (2015)	No	No	No	4/7
Lebanon	No	Yes (2020)	No	Yes (2018/19)	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)	No	Yes (2014)	No	5/7
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 (2017/18)	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 (2019/20) *	Yes (2017)	Yes (2017)	7/7
Yemen, Rep.	No	No	Yes (2014)	No	No	Yes (2014)	Yes (2014)	3/7
Total economies indicating collection of recent data for each data category	5/20	19/20	14/20	11/20	9/20	14/20	9/20	
Total recent data indicated to have been collected across all economies				81/140				

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.

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, and census data. Eleven 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 micro data collection across all data categories- 7 out of 7 recent microdata sets expected. They are closely followed by the Arab Republic of Egypt and Jordan which collected data for 6 out of the 7 recent microdata

sets expected. By contrast Algeria, Lebanon, the Syrian Arab Republic and the United Arab Emirates, only report 1 or 2 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 websites[15] for accessibility of the microdata indicated to have been collected. This is reported in Table 3, where entries are only provided where in Table 2 it was indicated that a recent microdata set has been collected. Of the 81 microdata sets, only 16 are accessible to a user visiting NSO websites. Of these, only 5 can be downloaded immediately: the 2018-19 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[16], 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 towards 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 81 that have been collected, are downloadable from NSO websites. Consequently, and depending of the definition used, only 10%-20% of the expected microdata are available to the public on NSO websites. Within the price and health data categories none of the NSOs makes microdata publicly available.

[15] 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.

[16] 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 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
Bahrain	-	No openness	No openness	-	No openness	No openness	-
Djibouti	-	No openness	Satisfactory	-	-	No openness	-
Egypt, Arab Rep.	-	No openness	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	-	No openness	No openness	No openness	No openness	No openness	No openness
Kuwait	No openness	No openness	No openness	No openness	-	-	-
Lebanon	-	No openness	-	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	-	Excellent	-
Oman	-	No openness	-	No openness	No openness	No openness	-
Qatar	-	No openness	-	-	-	No openness	No openness
Saudi Arabia	No openness	No openness	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	0/20	6/20	4/20	0/20	3/20	2/20
Total surveys/ census with some degree of accessibility of microdata on NSOs website.				16/140			

Note: - indicates up to date microdata have not been collected.

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 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 who 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.[17] 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 and Eurostat.[18] The results from this exercise are summarized in table 4.

[17] 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.

[18] See annex table A2 for the links to these microdata libraries.

Overall microdata accessibility improves by around 50% when we consider international accessibility in addition to national accessibility – from 16/140 to 25/140. Some countries like Iraq and Jordan who 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.

Table 4: Openness of recent source/survey/micro data on public facing website of MENA NSOs and international microdata libraries.[19]

Economy/data category	Establishment survey	Price survey	Consumption survey	Labor force survey	Health survey	Population census	Economic census
Algeria	-	No openness	-	-	-	-	No openness
Bahrain	-	No openness	No openness	-	No openness	No openness	-
Djibouti	-	No openness	Satisfactory (NSO, WB)	-	-	No openness	-
Egypt, Arab Rep.	-	No openness	Satisfactory (NSO)	No openness	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	-	No openness	No openness	No openness	Satisfactory (WB, DHS)	No openness	No openness
Kuwait	No openness	No openness	No openness	No openness	-	-	-
Lebanon	-	No openness	-	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	-	Excellent (NSO, IPUMS)	-
Oman	-	No openness	-	No openness	Satisfactory (WB, MICS)	No openness	-
Qatar	-	No openness	-	-	-	No openness	No openness
Saudi Arabia	No openness	No openness	No openness	No openness	No openness	No openness	No openness
Syrian Arab Republic	-	No openness	-	-	-		No openness
Tunisia	-	-	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)	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	0/20	8/20	5/20	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.				25/140			

Legend

- Microdata not collected
- Microdata available nationally only
- Microdata available internationally only
- Microdata available both nationally and internationally

[19] WB, IHSN microdata library as well as IPUMS, Eurostat, DHS and MICS data was accessed on 8th August 2021.

3.7 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. No MENA NSO makes price surveys (or at least item level price indices) publicly available even though such information would be relevant to a host of users. 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.

Suggested practice 1: Provide an English version of the website

While the primary audience for NSO statistics are 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 data sets 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.

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 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.

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 3 dropdown tabs as follows (1) survey data. (2) Open data (3) key indicators. The survey data tab links to the World Bank's microdata library.

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" where following satisfactory registration, access to the data is granted automatically is standard practice for international organizations such as the WB, MICS, DHS, IPUMS reported in table 4.

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.

4. Existing indicators measuring data accessibility in MENA

4.1 Existing indicators

The evidence in section 3 shows that the availability and accessibility of microdata in the MENA region is very constrained. Yet the existing indicators measuring data openness do not capture this reality. These indicators include the Open Data Inventory (ODIN)[20] published by Open Data Watch; Open Data Barometer (ODB)[21] by World Wide Web Foundation and Global Open Data Index (GODI)[22] by Open Knowledge Foundation[23] (see table 5 for a summary).[24] Of these indicators, the ODIN ranks highest with regards to country coverage.

The ODIN covers 178 countries in its 2018/19 version including 17 MENA countries and 187 countries in its 2020/21 version including all 20 MENA countries. Additionally, ODIN also has 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 ten elements measure data coverage i.e., the degree to which data is available and while the others measure access/openness i.e., 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 five years; data are available for the preceding ten 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 are assessed as follows: machine readability; non-proprietary; 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 (see annex table A3 and A4 for the full list of data categories, representative indicators in each category and scoring option for the 2018 ODIN). The default ODIN overall score gives equal weights to the three dimensions. ODIN scores are calculated as a percentage of the maximum score obtainable.

[20] Open Data Watch - Open Data Inventory http://www.opendatawatch.com
[21] World Wide Web Foundation- https://opendatabarometer.org/barometer/
[22] Global Open Data Index - https://index.okfn.org/
[23] The Statistical Access Indicator (SAI) by Almeida and Hoogeveen (2015) with 72% (36 out of 50) of its elements assessing data access and covering 49 African countries would have made the fifth indicator. However, it is unpublished, hence excluded from the review.
[24] The Statistical Access Indicator (SAI) by Almeida and Hoogeveen (2015) with 72% (36 out of 50) of its elements assessing data access and covering 49 African countries would have made the fifth indicator. However, it is unpublished, hence excluded from the review.

Table 5: Indicators/ data sets measuring data openness

Indicators/Data sets	Author	% of elements focusing on data access	Number of countries covered (year)	MENA countries coverage
1. Open Data Inventory (ODIN) -2018/2019 and 2020	Open Data Watch	50% (5 out of 10)	178 (2018/2019) 187 (2020)	17 of 20 in 2018/2019 - Algeria, Djibouti, Egypt, Arab Rep., Iraq, Iran, Islamic Rep., Jordan, Kuwait, Lebanon, Libya, Malta, Morocco, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, Tunisia, United Arab Emirates All 20 MENA countries in 2020
2. Open Data Barometer (ODB) – The leadership edition	World Wide Web Foundation	Around 33.3 %	30 (2017) Note: Scope of study was reduced to 30 countries in 2017 to include only countries that publicly committed to adopt the International Open Data Charter Principles or the equivalent G20 Anti-Corruption Open Data Principles.	1 of 20 – Saudi Arabia
3. Global Open Data Index (GODI)	Open Knowledge Foundation	About 50%	94 (2015)	3 of 20- Iran, Islamic Rep., Tunisia, Oman

Source: Author's compilation

Unlike the ODIN, the ODB assesses not just NSOs but overall government data, i.e., all government data regardless of whether the NSO contributes to making the data available. The most recent edition (the fifth edition) known as the leaders' edition of the ODB covers far fewer countries than the ODIN. Although previous editions of the ODB covered over 100 governments, the *Leaders Edition* covers only 30 governments that have publicly committed to adopt the International Open Data Charter Principles or the equivalent G20 Anti-Corruption Open Data Principles – only one of the 30 countries – Saudi Arabia- is in the MENA region. The ODB scores countries on three sub-indexes as follows: readiness; implementation and; impact. Each of the sub-indexes are weighted equally and further subdivided into various components weighted equally within each sub index. The readiness sub index measures the ability of governments to secure positive outcomes from open government data initiative. This is measured through three components focusing on: (1) Government; (2) Citizens and Civil Society; and (3) Entrepreneurs and Business. These components include measures relating to the existence of open data, and a range of interventions that support engagement with and re-use of open data. For the implementation sub index, the ODB asks researchers to complete a checklist of 10 questions[25] assessing data openness for various data categories (see annex table A5 for the various data categories).[26] Finally, for the impacts sub index, the ODB takes the approach of treating online, mainstream media and academic publications about open data impacts as a proxy for existence of impacts. Researchers were asked to score the extent of impact on a 0 – 10 scale. The scoring guidance directed assignment of the highest scores for peer-reviewed studies showing impact and emphasized the importance of sources making a direct connection between open data and observed impacts. For scores over 5, researchers were asked to cite at least two separate examples in the given category. To calculate each score for each component an average of the variables in that component is taken and the average of components is used to generate each sub-index. The weighted average of the sub-indexes is used to generate the overall ODB score.

Like the ODB, the GODI also assesses government data based on the open definition and the open data charter. However, unlike the ODB, GODI limits its inquiry to the publication of national government data and ignores other aspects of the common open data assessment framework such as context, use or impact. It covers 94 countries including 3 MENA countries- the Islamic Republic of Iran, Tunisia, and Oman- in its most recent version. GODI data categories are refined

[25] These 10 questions are as follows: Does the data exist? Is it available online from government in any form? Is the data set provided in machine-readable formats?; Is the machine-readable data available in bulk?; Is the data set available free of charge?; Is the data openly licensed?; Is the data set up to date?; Is the publication of the data set sustainable?; Was it easy to find information about this data set and; ; Are (linked) data URIs provided for key elements of the data?

[26] Excerpt from ODB methodology report- '"By putting forward categories of data, rather than specific named data sets, we allowed researchers to exercise judgement as to the extent to which countries were making data of this kind available, whilst also sourcing specific examples of data sets that fit into these categories in different countries, and generating a rich collection of qualitative information about the reasons that certain data may or may not be available in different countries, and the extent to which certain data sets tend to exist at national or federal levels."

each year to reflect key data that is relevant for civil society at large and are developed in partnership with domain experts, including organizations championing open data in their respective fields (see annex table A6 for most recent data categories). Each data set is evaluated using a set of questions in a survey that examines the openness of the data sets. Each survey question measures a crucial aspect of either the legal, technical, or practical 'openness' of data (see annex table A7 for detailed questions and scoring guidelines). As shown in annex table A7, GODI survey questions check different aspects of data access and usability. The scoring guidelines for GODI's survey questions are such that fairly high scores may not always mean open data, but access-controlled data, or public data in poorly structured, or not machine-readable formats. Specifically, GODI does not add many filters, such as exclusively considering data that is machine-readable - even though it might give a more realistic image of open data. For example, budget data can be assessed in PDF form which may be in public domain, available online for free, but in a format making it practically unusable. This data is presented as 80% open. The score suggests a fairly high degree of openness, but in fact, the data is not open. Only 100% means that the data is open. GODI claims that this approach, seeks to demonstrate which data is already available and how it can be further improved.

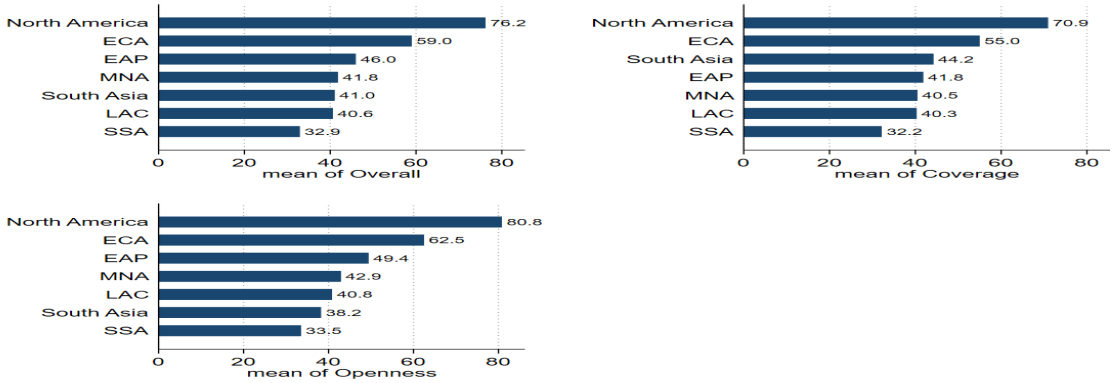
In comparison to the ODB and GODI, the ODIN has comprehensive country coverage and covers the most MENA countries. It is also easier to understand because its measure of data accessibility credits data when it is machine readable as opposed to the GODI that assigns scores to data in formats like PDF which is not really open at least from the perspective of a researcher seeking machine readable data for analytical work. However, ODIN's methodological guide mention that the terms "data," "statistics," and "indicators" are used interchangeably.[27] These three terms are clearly 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.

4.2 Performance of MENA countries on the ODIN

Countries in the MENA region perform rather well on the Open Data Access Indicator (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.

[27] 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 sub score



Source: Author's compilation using 2018/2019 ODIN data from Open Data Watch -- Open Data Inventory http://www.opendatawatch.com

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 sub-performance 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 is on par with that in other regions. This is a possibility. Our intuition is, however, that this is less likely as e.g., 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/micro data but does not capture the public release of (anonymized) microdata.

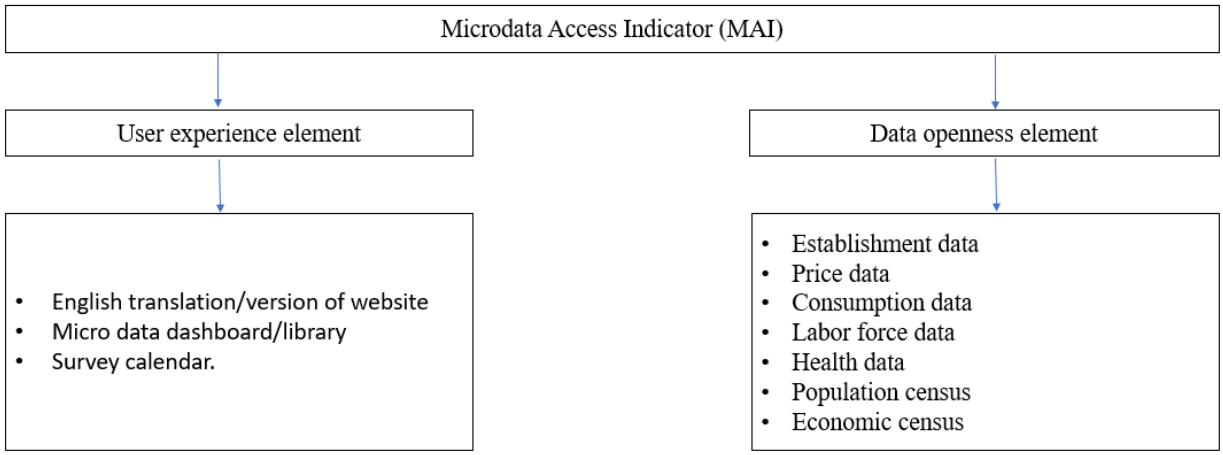
This can be illustrated by the availability of *"poverty statistics"* which ODIN assesses through the availability of two indicators (i) the poverty rate and (ii) the distribution of income by deciles or Gini coefficient. An NSO that publishes these statistics, without making available the underlying 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 micro data 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 ten elements across two dimensions - coverage and openness assessed across several data categories and data dimensions. But the data under these categories are not required to be at a high level of disaggregation i.e., they are not required to be individual/household level data- regional and sub regional 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 3.1 and 3.2, 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. Given the foregoing, we build on the methodology discussed in section 3 and present an indicator focusing on microdata openness in the next section. This indicator together with the ODIN will give a more balanced view of data openness in MENA.

5. The MENA Microdata Access Indicator (MENA MAI)

In this section, we present the MENA Microdata Access Indicator (MAI). It comprises of 2 elements - a user experience element and a data openness element. The data openness element focuses on measuring the degree of openness of available microdata while the user experience element measures the ease with which data users can access, understand, and navigate the website of MENA NSOs. Each element comprises of various features (See figure 3) discussed in the following section.

Figure 3: The MENA Micro Data Access Indicator



5.1 User experience element

As discussed in section 3, NSOs website typically contain a multiplicity of information, hence it is important that they provide microdata users with navigation tools to locate available microdata. We measure the ability of NSOs to do this with the user experience element following the framework given by the Standardized User Experience Percentile Rank Questionnaire (SUPR-Q) developed by Sauro (2015). The SUPR-Q covers 8 items measuring 4 factors of quality of website user experience - usability, trust, appearance, and loyalty. The 8 items and their corresponding factors are as follows: (1) The website is easy to use (usability); (2) It is easy to navigate within the website (usability); (3) I feel comfortable purchasing from the website (trust); (4) I feel confident conducting business on the website (trust); (5) How likely are you to recommend this website to a friend or colleague? (loyalty); (6) I will likely return to the website in the future (loyalty); (7) I find the website to be attractive (appearance); (8) The website has a clean and simple presentation (appearance).

To measure user experience for microdata users visiting NSOs website, we focus on the two items in the SUPR-Q that corresponds to usability. These items relate to the ease of use and navigation of websites. With regards to microdata accessibility, this can be measured with three key features that are important for both local and international microdata users to easily understand and navigate the website of NSOs efficiently. These features are as follows: (1) an English

translation/version of the landing page of NSOs website; (2) a micro data dashboard/library and (3) a survey calendar. These features are chosen for several reasons. First, English is the most widely spoken language globally and has a dominant presence in the media. Hence a larger group of international audience including content creators and researchers will be able to better use an English version of NSOs' websites. When an English version is unavailable, it is possible for international users to leverage other translation resources such as Google translate. However, it is still vital for NSOs to maintain an outstanding English translation of its website so that the information disseminated in the local language is translated appropriately from the original material and retain its intended meaning. Similarly, organizing available microdata in a microdata dashboard/library/tab on NSOs website makes it easy to find. A survey calendar also helps microdata users know and track available and forthcoming surveys by the NSO or other agencies. Combined, a microdata library and survey calendar on an NSO's website helps microdata users navigate NSOs website in search of microdata without wasting unnecessary time.

Following the implementation procedure outlined in section 3.3, countries are assigned scores for each of the 3 key features using the scoring criteria outlined in table 6.

Table 6: Scoring criteria for the user experience element

Element	Representative indicators	Scoring criteria
User experience	English translation	0 if no English translation/version is available on NSO's website landing page. 50 if an English translation/version for NSO's website is available but does not reasonably reflect the local language(s)and/ or not available for other pages besides the landing page. 100 if English translation/version for NSO's website is available, reasonably reflects the local language and available for other pages besides the landing page.
	Micro data tab/dashboard/library	0 if no micro data tab/dashboard/library is available on NSO's website 100 if micro data tab/dashboard/library is available on NSO's website
	Survey calendar	0 if no survey calendar is available on NSO's website 50 if survey calendar is available on NSO's website but not up to date. 100 if survey calendar is available on NSO's website and up to date.

5.2 Data openness element

In section 3, we describe and present results of an exercise that examines both the website of NSOs and international repositories such as web-portals of the MICS and DHS surveys and microdata libraries maintained by the World Bank (WB), International Household Survey Network (IHSN), IPUMS and Eurostat to determine the degree of accessibility of microdata. For each data

category, the data openness element assigns scores to countries based on the classification of microdata data accessibility summarized in table 4 using the scoring criteria outlined in table 7.

Table 7: Scoring criteria for the data openness element

Data category	**Data sub-category**	**Representative indicator/ data sets**	**Scoring** criteria
Firm data	N/A	Enterprise survey or Annual survey of industry or equivalent.	0 if up to date representative microdata accessibility is classified as "No coverage" or "No openness". 50 if up to date representative microdata accessibility is classified as "Satisfactory openness". 100 if up to date representative microdata accessibility is classified as "Excellent openness".
Price data	N/A	Commodity price survey or equivalent.	
Individual/ Household data	**Consumption data**	Household budget survey, Household income and expenditure survey, Living Standards Measurement Survey (LSMS) or equivalent.	
	Labor Force data	Labor Force survey or equivalent.	
	Health data	Demographic and Health Survey (DHS), Multiple Indicator Cluster Survey (MICS) or equivalent.	
Census data	**Population**	Population census	
	Economic	Economic / Business/ Establishment census	0 if up to date representative microdata accessibility is classified as "No coverage" or "No openness". **<u>If available for one or more sectors</u>** 25 if up to date representative microdata accessibility is classified as "Satisfactory openness". 50 if up to date representative microdata accessibility is classified as "Excellent openness". **<u>If available for all sectors</u>** 50 if up to date representative microdata accessibility is classified as "Satisfactory openness". 100 if up to date representative microdata accessibility is classified as "Excellent openness".

5.3 Constructing scores for the MENA MAI

The MENA MAI uses equal nested weights with its 2 elements. This weighting structure is symmetric, monotonic, and decomposable by subgroup. It is based on Atkinson's (2003) counting method and has been used to construct several indices including the Multi-dimensional Poverty Index, Social Exclusion Index and Statistical Performance Index (see Alkire and Foster, 2011; Cameron et al., 2021; Chakravarty & D'Ambrosio, 2006; Dang et al. 2021). The equal nested weighting structure is such that the indicator is sequentially aggregated at each level so that all features within an element receive an equal share of the total weight of that element (see annex table A8 for weights of features within the MENA MAI). Since, the MENA MAI has a two-level structure – the first level comprises of the two elements and the second comprises of the features within each element-, this implies the following.

1. The sub-score for each of the two elements in each country is derived by taking the average of the scores for features within the element.
2. The overall score in each country is derived by taking the average of the sub-scores for the 2 elements.

Hence the scores for each country are given as follows:

User experience sub-score= $\frac{1}{n}\sum_{i=1}^{n} Feature_i$

Where:
n=3 (total number of features in the user experience element) and
i represents each feature.

Data openness sub-score= $\frac{1}{m}\sum_{k=1}^{m} DataCategories_k$

Where:
m=7 (total number of data categories in the data openness element) and
k represents each data category

Overall MENA SAI score =$\frac{1}{2}$ (user experience sub score) + $\frac{1}{2}$ (data openness sub score).

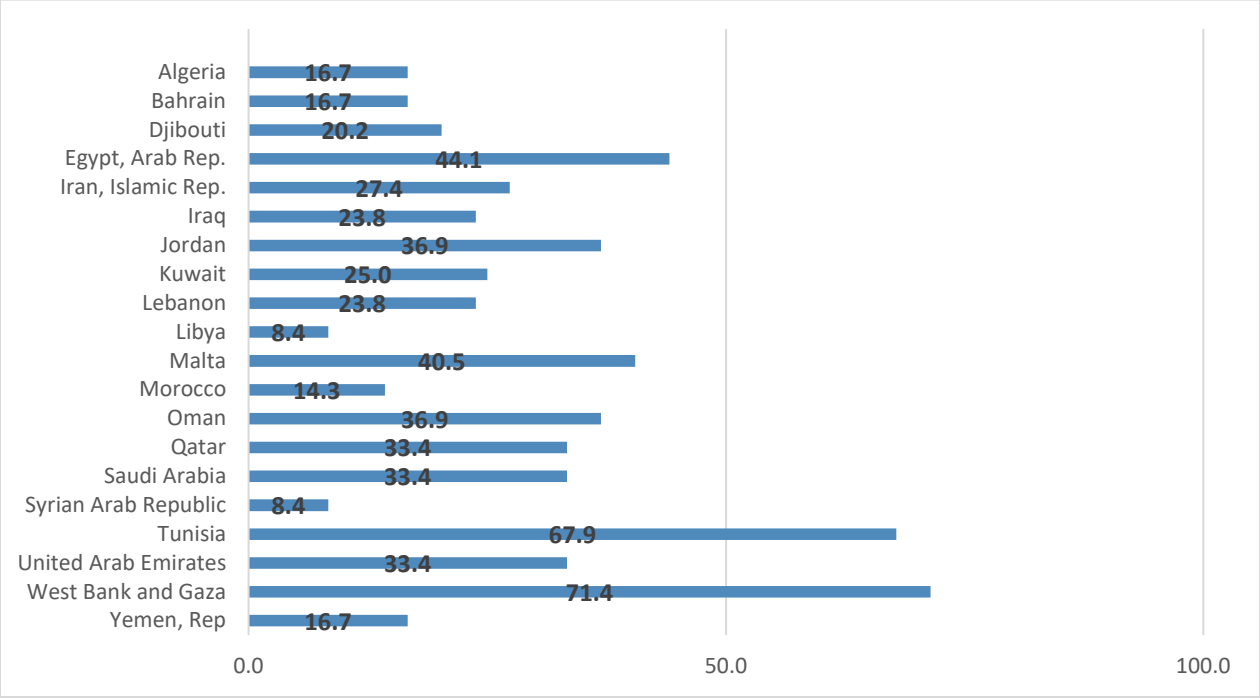
The data openness cross-element cut-off score is fixed at 50. This is because when countries receive a score of at least 50 in the data openness element, it implies that on average they have at least satisfactory openness in all data categories examined. Similarly, receiving at least a score of 50 on the user experience element is equivalent to receiving the midpoint score across all 3 features of the element. This suggests that microdata users can easily understand and navigate NSOs website at an acceptable level. The scores for each country are summarized in the next section.

5.4 Results

The raw scores of the MENA MAI based on the scoring criteria in table 6 and 7 for MENA countries are reported in Annex table A9. Only Tunisia and West Bank and Gaza obtain the cut off score or

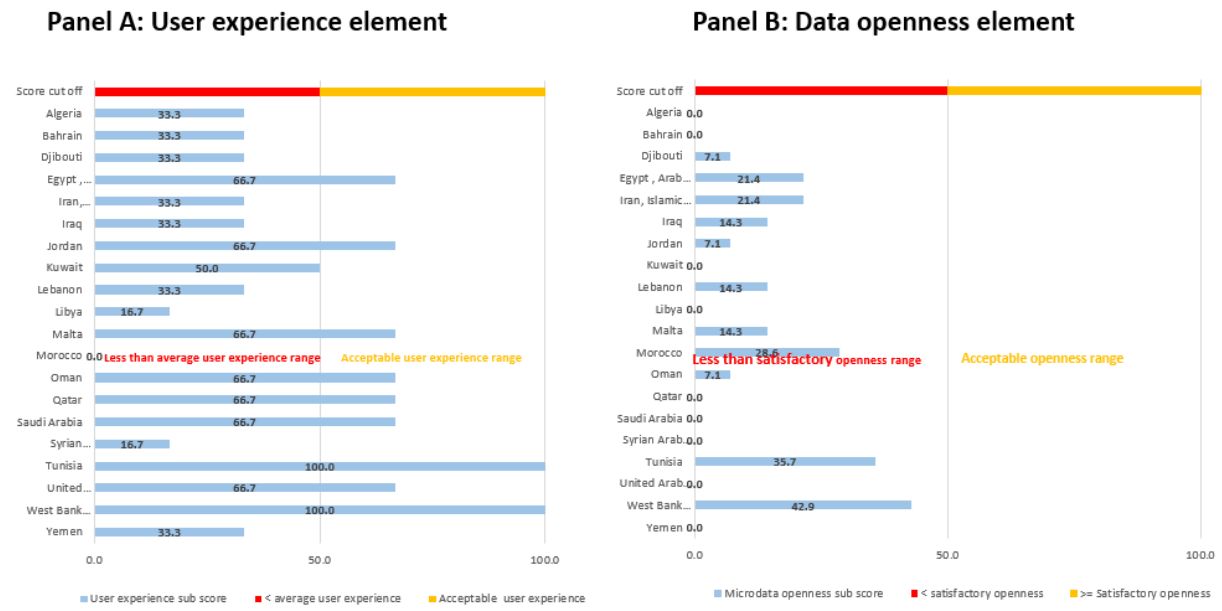
better (see figure 4). NSOs perform lower than average on the user experience element. Although most NSOs obtain a perfect score for providing an English translation of their websites landing page, they fall considerably short with regards to providing updated survey calendars and micro data libraries on their websites. Majority of NSOs falling below the cut off score for the user experience element implies that they can make considerable improvements to the experience of microdata users who visit their websites (see panel A of figure 5). In general, all NSOs perform poorly on the overall score of the MENA MAI. The data openness element contributes significantly to this poor performance (see panel B of figure 5) since a lot of countries obtain no score for data openness for the various data categories assessed because recent data is not available on their NSOs websites and/or internationally.

Figure 4: MENA micro data access indicator – overall score



Source: Author's compilation using own data

Figure 5: MENA micro data access indicator – user experience element and data openness element scores

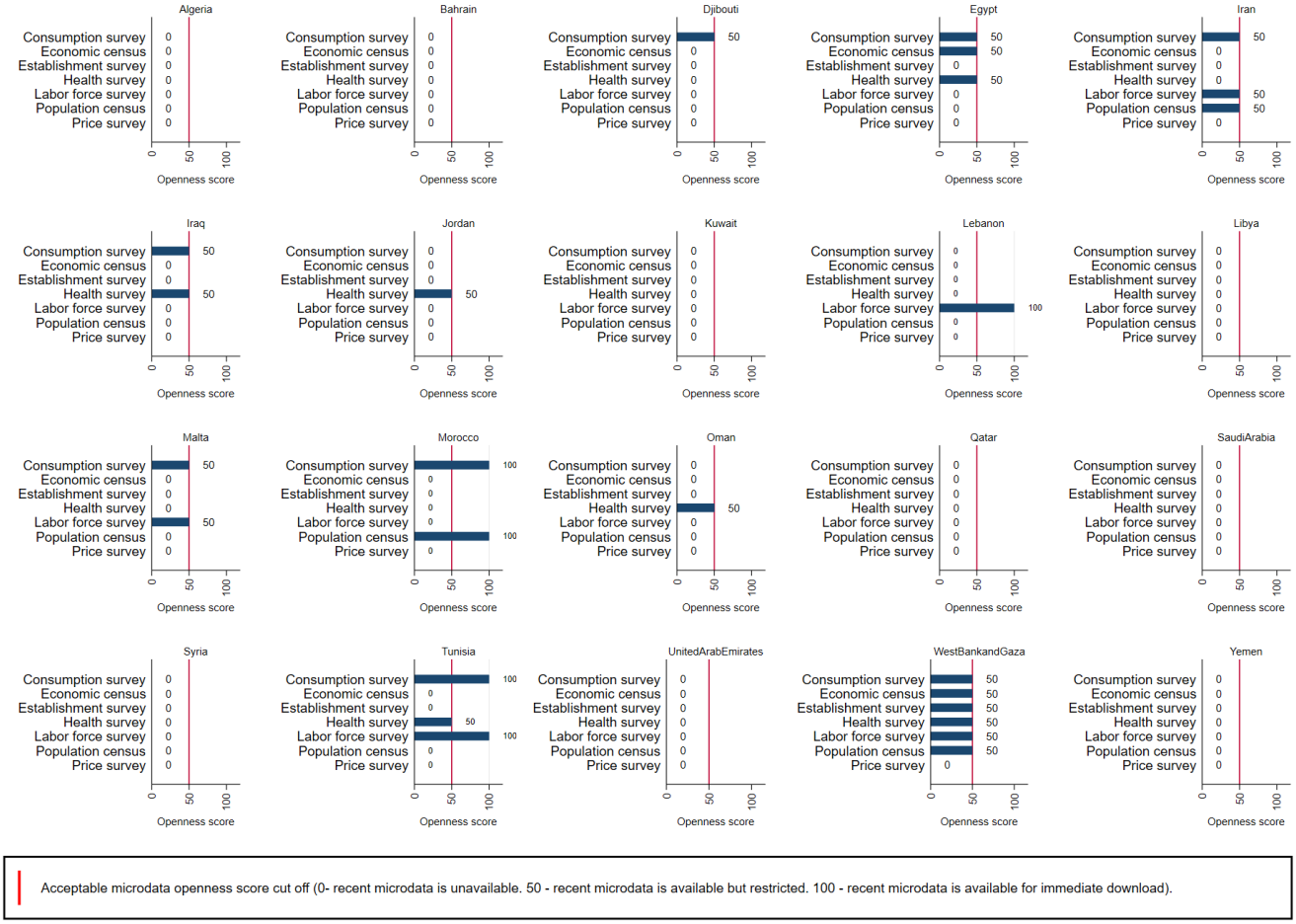


Source: Author's compilation using own data

For countries that obtain a positive score on the data openness element, as expected, the scores on data openness for all data categories assessed is poor. Across all MENA countries, only five (5) - the 2018-19 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[28], the 2017 Tunisia National Population and Employment Survey and a subset of the 2014 population census microdata for Morocco- of the 140 data sets assessed receive a perfect score for data openness (see figure 6). Yet as discussed in section 4.2, the ODIN suggests that MENA countries do well on data openness. Our results show that this does not consider openness of micro data. All countries do very poorly on the data openness element of the MENA MAI compared to their recent ODIN openness sub score (see figure 7). In figure 8, we show scores of the "poverty and statistics" category of the ODIN for two consecutive rounds alongside scores of the consumption data category - the underlying survey for calculating poverty statistics - of the data openness element in the MENA MAI. As expected, several MENA countries that receive positive openness scores for "poverty and statistics" on the ODIN receive a score of 0 on the consumption data category of the MENA MAI.

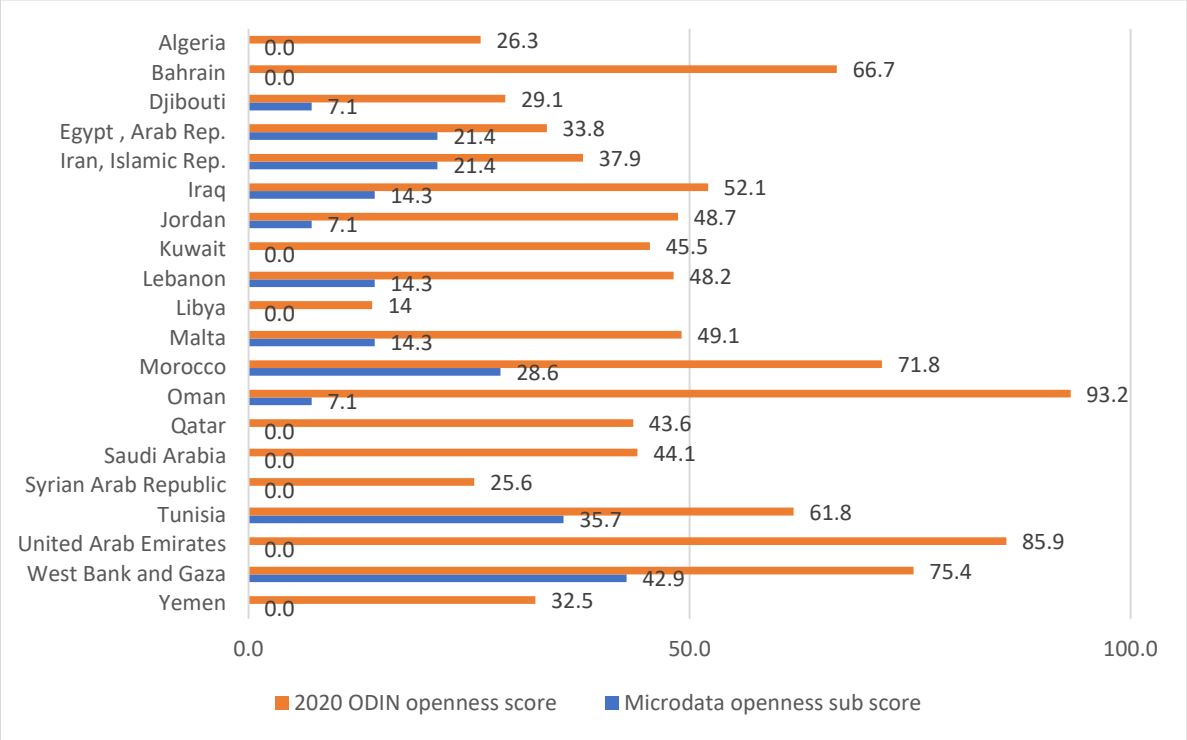
[28] For the 2015 Tunisia Budget survey, it is important to note that not all variables are included in the microdata set available for immediate download.

Figure 6: Openness scores for source/survey/micro data categories.



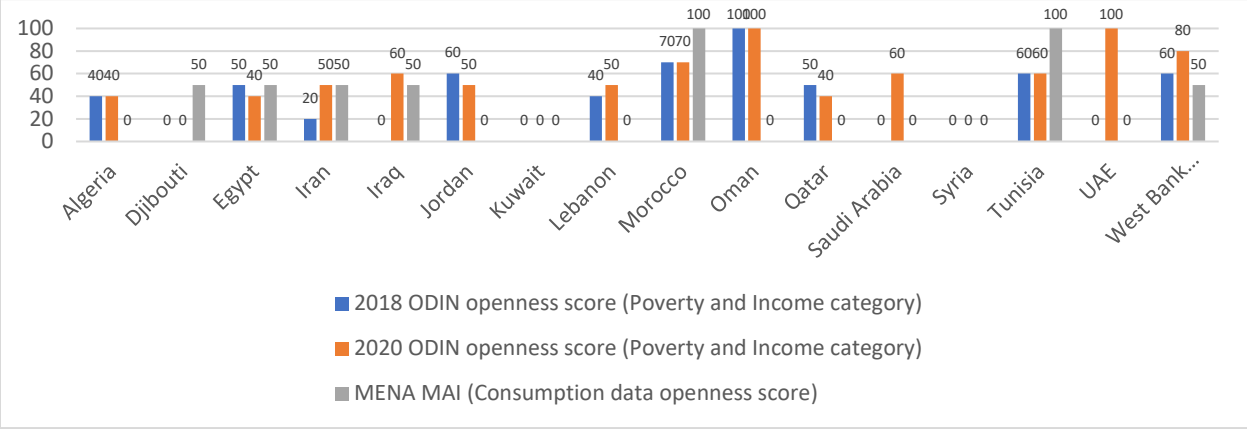
Source: Author's compilation using own data

Figure 7: MENA SAI micro data openness sub score vs. ODIN 2020 openness sub score.



Source: Author's compilation using own data and 2020 ODIN data from Open Data Watch -- Open Data Inventory http://www.opendatawatch.com

Figure 8: ODINs poverty and income openness scores vs MENA MAI consumption/welfare data openness score

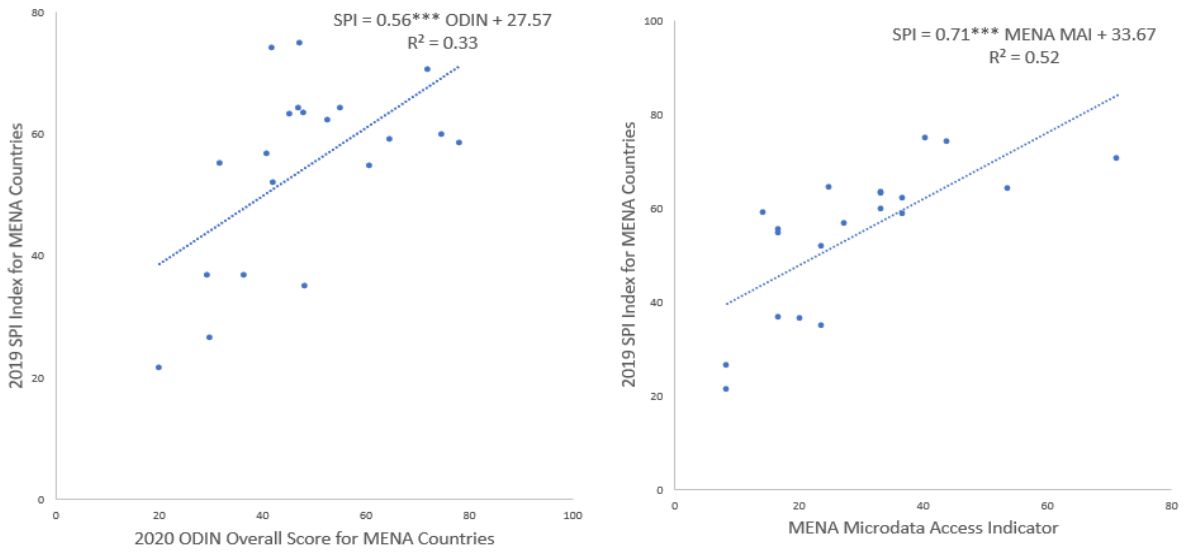


Source: Author's compilation using own data, 2018/19 and 2020 ODIN data from Open Data Watch -- Open Data Inventory http://www.opendatawatch.com

Note: Countries without poverty and income openness score for the 2018/19 and 2020 ODIN are not shown.

Next, we graph the fitted line of the regression of the recently released Statistical Performance Index (SPI) which measures the performance of entire national statistical systems, on the MENA MAI and the ODIN (see figure 9). The estimated slope of both regression lines is positive and strongly statistically significant, suggesting that by improving access to source/survey/micro data in addition to summary data, countries can improve the overall performance of their statistical systems.

Figure 9: The MENA micro data access indicator and the WB SPI



Source: Author's calculation using data from the MENA MAI in the current paper and the 2019 WB SPI Index downloaded from https://github.com/worldbank/SPI/
Note: ***Statistical significance at the 1% level.

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 paper we assessed the availability of anonymized microdata sets for the MENA region across 7 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; 9 out of 20 are up to date on their economic census. Only 5 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 9, 11, and 14 countries respectively). The implication is that in almost half the cases, no or outdated microdata are 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 (7 data categories in 20 countries) 81 had been collected and as few as 25 were accessible. Remarkably, of these 25 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 Open Data Inventory (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.

Hence, we present a new indicator to complement the ODIN. The new indicator assesses the degree of openness of recent microdata as well as the ease with which microdata users can access this data on NSOs' websites. Together with the ODIN, it gives a robust picture of data openness in MENA. It is cost effective, takes less time to collect and can be updated periodically as it only requires visits by the research team to the websites of all NSOs and international microdata libraries to determine the scores of its various elements with no input from the NSOs. The scores of MENA countries on the new microdata access indicator show that there is a strong case to be made to invest in the collection and release of microdata sets in the MENA region. 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.

References

Alkire, S., Foster J. (2011). *Counting and Multidimensional Poverty Measurement*. Journal of Public Economics, 95(7): 476-487.

Arezki, R., Lederman, D., Abou Harb, A., El-Mallakh, N., Fan, R. Y., Islam, A., Zouaidi, M. (2020). *How Transparency Can Help the Middle East and North Africa.* Washington, DC: World Bank. doi:10.1596/978-1-4648-1561-4.

Atamanov, A., Tandon,S., Lopez-Acevedo, G., Vergara B., 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.

Atkinson, A. B. (2003). *Multidimensional Deprivation: Contrasting Social Welfare and Counting Approaches*. Journal of Economic Inequality, 1(1): 51-65

Brackfield, D. (2011). *OECD Work on Measuring Trust in Official Statistics*. Int Statistical Inst.: Proc 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), 503-17.

Cameron, G.J., Dang, H.H, Dinc, M., Foster, J., Lokshin, M.M. (2021). *Measuring the Statistical Capacity of Nations*. Oxford Bulletin of Economics and Statistics. Doi: https://doi.org/10.1111/obes.12421

Chakravarty, S.R., D'Ambrosio. C. (2006). *The Measurement of Social Exclusion*. Review of Income and Wealth. 52(3): 377-398.

Dang, H.H., Pullinger, J., Serajuddin, U., 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. © World Bank. https://openknowledge.worldbank.org/handle/10986/35301 License: CC BY 3.0 IGO.

Dang H., Jolliffe D., & 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). Pg 757-797 https://doi.org/10.1111/joes.12307

Sauro, J. (2016). SUPR-Q: A Comprehensive Measure of the Quality of the Website User Experience. Journal of Usability Studies. Vol. 10, Issue 2, February 2015, pp. 68-86. https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Sauro_Feb2015.pdf

Annex

Table A1: Websites of National Statistical Agencies in MENA

S/N	Economy	National Statistical Agencies	Website
1	Algeria	The National Office of Statistics	https://www.ons.dz/
2	Bahrain	National statistical office of Bahrain Information and eGovernment Authority	http://www.data.gov.bh; http://www.iga.gov.bh
3	Djibouti	National Institute of Statistics of Djibouti	www.instad.dj
4	Egypt, Arab Rep.	Central Agency for Public Mobilization and Statistics (CAPMAS)	https://www.capmas.gov.eg/
5	Iraq	Central Statistical Organization	http://cosit.gov.iq/en/
6	Iran, Islamic Rep.	Statistical Centre of Iran	https://www.amar.org.ir/
7	Jordan	Department of Statistics	http://dosweb.dos.gov.jo/ar
8	Kuwait	Central Statistical Bureau	https://www.csb.gov.kw/
9	Lebanon	Central Administration for Statistics	http://www.cas.gov.lb/
10	Libya	Bureau of Statistics and Census Libya	http://www.bsc.ly/
11	Malta	National Statistics Office	https://nso.gov.mt/en/
12	Morocco	Direction de la Statistique	https://www.hcp.ma/Direction-de-la-statistique_a716.html
13	Oman	National Centre for Statistics and Information	https://www.ncsi.gov.om/Pages/NCSI.aspx
14	Qatar	Planning and Statistics Authority	https://www.psa.gov.qa/en/Pages/default.aspx
15	Saudi Arabia	General Authority for Statistics	https://www.stats.gov.sa/en
16	Syrian Arab Republic	Central Bureau of Statistics	http://cbssyr.sy/index-EN.htm
17	Tunisia	National Institute of Statistics (INS)	http://www.ins.tn/en/statistics-tunisia-national-institute-statistics
18	United Arab Emirates	Federal competitiveness and Statistical Center	https://fcsc.gov.ae/en-us
19	West Bank and Gaza	Palestinian Central Bureau of Statistics	http://www.pcbs.gov.ps/default.aspx
20	Yemen, Rep.	Central Statistical Organization	http://www.cso-yemen.com/

Table A2: International organizations and their microdata library websites

S/N	Organization	Online microdata library
1	World Bank	https://microdata.worldbank.org/
2	IPUMS	https://ipums.org/
3	International Household Survey Network (IHSN)	https://www.ihsn.org/
4	Multiple Indicator Cluster Survey (MICS)	https://mics.unicef.org/surveys
5	Eurostat	https://ec.europa.eu/eurostat/web/microdata
6	Demographic and Health Surveys (DHS)	https://dhsprogram.com/data/available-datasets.cfm

Table A3: ODIN data dimension, category and indicators

S/N	Data dimension	Data category	Representative indicators	Categorical disaggregation	Coverage element 1 scoring guidelines*
1	Social Statistics	Population and Vital Statistics	(1.1) Population by 5-year age groups (1.2) Birth rate (1.3) Death rate	(1.1) Sex; marital status (1.2) Sex; marital status (1.3) Sex	To receive full point: Must have all indicators disaggregated by sex. To receive half point: Must have at (1.1) with one disaggregation, or (1.2) and (1.3) with one disaggregation each.
2		Education Facilities	(2.1) Number of schools or classrooms; (2.2) Number of teaching staff; (2.3) Education budget data.	(2.1) School stage; school type. (2.2) School stage; school type; (2.3) School stage; functional categories.	To receive full point: Must have all three indicators disaggregated as follows: (2.1) and (2.2) by school stage (see notes) and one other disaggregation and (2.3) with one disaggregation. To receive half point: Must have one indicator with two disaggregation or two indicators with one disaggregation each.
3		Education Outcomes	(3.1) Enrollment rate; (3.2) Completion or graduation rate; (3.3) Competency exam results.	(3.1, 3.2) Sex; school stage; age; school type. (3.3) Sex.	To receive full point: Must have all indicators disaggregated as follows: (3.1) and (3.2) by sex and one other disaggregation, and (3.3) by sex. To receive half point: Must have (3.1) or (3.2) with two disaggregations, or (3.3) disaggregated by sex. Two indicators with only one disaggregation each is not enough.
4		Health Facilities	(4.1) Number of health facilities; (4.2) Number of beds or data on health care staff; (4.3) Health budget data.	(4.1) Facility type; (4.2) Facility type; department type; staff type (4.3) Functional categories	To receive full point: Must have two indicators with one disaggregation each. To receive half point: Must have one indicator with one disaggregation.

Table A3 continued

Table A3 continued

S/N	Data dimension	Data category	Representative indicators	Categorical disaggregation	Coverage element 1 scoring guidelines*
5	Social Statistics	Health Outcomes	(5.1) Immunization rate. (5.2) Diseases prevalence or incidence. (5.3) Stunting, wasting, or obesity rate.	(5.1) Age; sex. (5.2) Age; sex; disease type. (5.3) Age; sex.	To receive full point: Must have (5.1), (5.2) and (5.3) by sex, and (5.2) by disease type. To receive half point: Must have at least one indicator with one disaggregation.
6		Reproductive Health	(6.1) Maternal mortality ratio or rate. (6.2) Infant mortality rate or neonatal mortality rate; (6.3) Under-5 mortality rate; (6.4) Fertility rate; (6.5) Contraceptive prevalence rate; (6.6) Adolescent birth rate.	(6.1) None; (6.2, 6.3) Sex; (6.4-6.6) None.	To receive full point: Must have five indicators with one disaggregation each. To receive half point: Must have two indicators with one disaggregation each.. One must be a mortality rate.
7		Gender Statistics	(7.1) Proportion or number of women who are victims of physical, sexual, or psychological violence; (7.2) Proportion of women in government, management or senior positions; (7.3) Data on child marriages.	(7.1) Age; disability status; relationship to perpetrator; (7.2) None; (7.3) None.	To receive full point: Must have all indicators with one disaggregation. To receive half point: Must have at least one indicator with one disaggregation.

Table A3 continued

Table A3 continued

S/N	Data dimension	Data category	Representative indicators	Categorical disaggregation	Coverage element 1 scoring guidelines*
8	Social Statistics	Crime and Justice Statistics	(8.1) Homicide rate or count; (8.2) Crime rate or count; (8.3) Persons in prison or incarceration rate.	(8.1) Sex of victim; age of victim; sex of perpetrator; age of perpetrator; victim/perpetrator relationship. (8.2) Crime type; age of victim; sex of victim; sex of perpetrator; age of perpetrator; victim/perpetrator relationship. (8.3) Sentencing status; age; sex.	To receive a full point: Must have all three indicators with one disaggregation each or two indicators with two disaggregations each. (8.2) must be disaggregated by crime type. To receive half point: Must have one indicator with two disaggregations or two indicators with one disaggregation each.
9		Poverty Statistics	(9.1) Poverty rate; (9.2) Distribution of income by deciles or Gini coefficient.	(9.1, 9.2) None	To receive full point: Must have two indicators. To receive half point: Must have one indicator.
10	Economic statistics	National Accounts	(10.1) GDP (production approach) or gross value added; (10.2) GDP (expenditure approach).	(10.1) Industrial classification; (10.2) Major expenditure categories.	To receive full point: Must have all indicators disaggregated as follows: (10.1) by industrial classification and (10.2) by major expenditure categories. Any data in the most recent 5 years (2013 onward) must be presented on at least quarterly basis to receive a full point. To receive half point: Must have at least one indicator with one disaggregation. Data can be presented on a monthly, quarterly, or annual basis.
11		Labor Statistics	(11.1) Employment rate; (11.2) Employment distribution; (11.3) Unemployment rate.	(11.1) Sex; age; (11.2) Industry; Occupation type; Sex; (11.3) Sex; age.	To receive full point: Must have: (11.1) by sex, or (11.2) by sex and one other disaggregation, and (11.3) by sex. To receive half point: Must have at one indicator with one disaggregation.

Table A3 continued

Table A3 continued

S/N	Data dimension	Data category	Representative indicators	Categorical disaggregation	Coverage element 1 scoring guidelines*
12	Economic statistics	Price Indexes	(12.1) Consumer price index (CPI); (12.2) Producers price index (PPI)	(12.1, 12.2) None	To receive full point: Must have all indicators. Any data in the most recent 5 years (2013 onward) must be presented on at least quarterly basis to receive a full point. To receive half point: Must have at least one indicator. Data can be presented on a monthly, quarterly, or annual basis.
13		Government Finance	(13.1) Actual revenues; (13.2) Actual expenditures.	(13.1) Revenue source; (13.2) Administrative classification; economic classification; functional classification.	To receive full point: Must have all indicators with one disaggregation each. To receive half point: Must have one indicator with one disaggregation.
14		Money and Banking	(14.1) Money supply; (14.2) Interest rates.	(14.1) M1; M2; M3; (14.2) Rate type.	To receive full point: Must have all indicators with one disaggregation. Must have at least three rates. To receive half point: Must have one disaggregated indicator. If indicator is (14.2), only one rate is needed.
15		International Trade	(15.1) Merchandise exports; (15.2) Merchandise imports.	(15.1, 15.2) Major product categories (agricultural products, fuels, mining, manufactures, etc.).	To receive full point: Must have (15.1) and (15.2) imports by major product categories. Any data in the most recent 5 years (2013 onward) must be presented on at least quarterly basis to receive a full point. To receive half point: Must have one indicator with one disaggregation. Data can be presented on a monthly, quarterly, or annual basis.

Table A3 continued

Table A3 continued

S/N	Data dimension	Data category	Representative indicators	Categorical disaggregation	Coverage element 1 scoring guidelines*
16	Economic statistics	Balance of Payments	(16.1) Current account; (16.2) Capital and Financial account.	(16.1) Goods and services, income, and current transfers (or secondary income); (16.2) Direct investment or international investment position.	To receive full point: Must have (16.1) and (16.2) with one disaggregation each. To receive half point: Must have (16.1) or (16.2) with one disaggregation.
17	Environmental statistics	Land Use	(17.1) Data on land use or land cover. (17.2) Data on protected lands.	(17.1) Urban and rural; agricultural use (crop type); environmental zones. (17.2) None	To receive full point: Must have (17.1) with two disaggregations, as well as (17.2). Land use data with one disaggregation and land cover data with one disaggregation is accepted. To receive half point: Must have one indicator with one disaggregation.
18		Resource Use	(18.1) Data on fishery harvests; (18.2) Data on timber harvests or deforestation; (18.3) Data on major mining or extractive activities; (18.4) Water supply and/ or consumption.	(18.1, 18.2) None (18.3) Type of mining activity; (18.4) None	To receive full point: Must have three indicators, disaggregated To receive half point: Must have two indicators, disaggregated
19		Energy Use	(19.1) Energy consumption	(19.1) Energy type; end-use sector; industrial sector.	To receive full point: Must have (19.1) by energy type and one other disaggregation. Three energy types must be present. To receive half point: Must have (19.1) by energy type and one other disaggregation. Two energy types must be present.

Table A3 continued

Table A3 continued

S/N	Data dimension	Data category	Representative indicators	Categorical disaggregation	Coverage element 1 scoring guidelines*
20	Environmental statistics	Pollution	(20.1) CO2 or other greenhouse gas (GHG) emissions; (20.2) Emissions of air or water pollutants.	(20.1,20.2) None	To receive full point: Must have all indicators. CO2 must be specified. To receive half point: Must have at least one indicator.
21		Built Environment	(21.1) Proportion of households with access to water; (21.2) Proportion of households with access to sanitation; (21.3) Housing quality indicators.	(21.1) Water supply type; (21.2) Sanitation facility type; (21.3) Number of houses by type; number of rooms; Houses by construction material; Houses by piping type; other.	To receive full point: Must have (21.1) and (21.2); as well as (21.3) with one disaggregation. To receive half point: Must have either (21.1) and (21.2), or (21.3) with at least one disaggregation.

Source: Authors own compilation from ODIN 2018/19 Methodology Report. https://docs.google.com/document/d/1ubPL1I_3im9bjlCVZ6W2ICAy6UAiXl1hGeA1aXImkxI/edit
*Scoring options for elements 2 to 10 are presented in Table A4.
For the elements of data openness, scoring is calculated independent of the data coverage.

Table A4: ODIN 2018/2019 scoring options

Dimension	Element	Scoring options	Notes
Coverage*	2. Data are available for the preceding five years	1 point if all published data are available for 3 of the last 5 years. 0.5 points if some published data are available for 1-2 of the last 5 years.	Scores for this element cannot be greater than the score for coverage element 1. Scores are given by data category, not indicator.
	3. Data are available for the preceding ten years	1 point if all published data are available for 6 of the last 10 years. 0.5 points if some published data are available for 3-5 of the last 5 years. 0 points if all published data are unavailable for 2 or fewer of the last 10 years.	Scores for this element cannot be greater than the score for coverage element 1. Scores are given by data category, not indicator.
	4. Data are disaggregated at the first administrative level	1 point if all published data in a data category are available at first administrative level. 0.5 points if some published data in a data category are available at first administrative level. 0 points if no data are available at this level	Scores for this element cannot be greater than the score for coverage element 1. Scores are given by data category, not indicator. Additionally, data disaggregated at the first administrative level is only scored if national level data also exists for that indicator.
	5. Data are disaggregated at the second administrative level	1 point if all data in a data category are available at second administrative level. 0.5 points if some published data are available at second administrative level. 0 points if no data are available at this level.	Scores for this element cannot be greater than the score for coverage element 1. Scores are given by data category, not indicator. Additionally, data disaggregated at the second administrative level are only scored if national level data also exists for that indicator.
Openness	1. Machine readability	1 point if all published data are available in a machine-readable format (such as XLS, XLSX, CSV, Stata, SAS, SPSS, JSON and so forth). 0.5 points if some published data are available in machine-readable format. 0 points if all published data are not available in machine-readable format.	Scores are not penalized for having identical data sets in both machine readable and non-readable formats. Compression formats do not affect machine readability scores, only non-proprietary scores (see next page). Scores are given by data category, not indicator.
	2. Non-proprietary	1 point if all published data are available in non-proprietary format (such as XLSX, DOCX, CSV, XML, HTML, and JSON). 0.5 points if some published data are available in non-proprietary format. 0 points if no data are available in a non-proprietary format.	If data files are compressed in RAR format (which is proprietary), data for that indicator should be considered proprietary even if the enclosing files are in a non-proprietary format. Files compressed in ZIP format are not affected.

Table A4 continued

Table A4 continued

Dimension	Element	Scoring options	Notes
Openness	3. Download options	1 point if all published data has a bulk download option and an API or user-selectable download option. 0.5 points if some published data has an API, bulk download, or user-selectable download options. 0 points if no published data have any download options	A bulk download is defined at the indicator level as: The ability to download all data recorded in ODIN for a particular indicator (all years, disaggregations, and subnational data) in one file, or multiple files that can be downloaded simultaneously. Bulk downloads are a key component of the Open Definition, which requires data to be "provided as a whole . . . and downloadable via the internet." User-selectable download options are defined as: Users must be able to select an indicator and at least one other dimension to create a download or table. These dimensions could include time periods, geographic disaggregations, or other recommended disaggregations. An option to choose the file export format is not enough. API stands for Application Programming Interface. Ideally, APIs should be clearly displayed on the website. ODIN assumes APIs are available for the NSOs entire data collection used in ODIN, unless clearly stated. ODIN assessors do not register for use or test API functionality. Scores are given by data category, not indicator.
	4. Metadata available	1 point if all published data have complete metadata. 0.5 points if some published data have complete or incomplete metadata, or all published data have incomplete metadata. 0 points if all published data have no metadata.	ODIN classifies metadata into three categories: (1) Not Available, (2) Incomplete, and (3) Complete. The following must be available to classify metadata as complete: • Definition of the indicator, or definition of key terms used in the indicator description (as applicable), or how the indicator was calculated. • Publication (date of upload), compilation date (date on front of report is not sufficient), or date data set was last updated. • Name of data source (what agency collected the data). If the metadata only have one or two of the above elements, they are scored as incomplete

Table A4 continued

Table A4 continued

Element category	Element	Scoring options	Notes
Openness	5. Terms of use	1 point if all published data have a terms of use classified as open. 0.5 points if some published data have a terms of use classified as open, or if some published data has a terms of use classified as semi-restrictive or if all published data has a terms of use classified as semi-restrictive. 0 points If no terms of use are found or if all published data have a terms of use classified as restrictive.	Generally, terms of use (TOU) will apply to an entire website or data portal (unless otherwise specified). In these cases, all data found on the same website and/or portal will receive the same score. If a portal is located on the same domain as the NSO website, the terms of use on the NSO site will apply. If the data are located on a portal or website on a different domain, another terms of use will need to be present. For a policy/ license to be accepted as a terms of use, it must clearly refer to the data found on the website. Terms of use that refer to nondata content (such as pictures, logos, etc.) of the website are not considered. A copyright symbol at the bottom of the page is not sufficient. A sentence indicating a recommended citation format is not sufficient. Terms of use are classified the following ways: (1) Not Available, (2) Restrictive, (3) Semi-Restrictive, and (4) Open.

Source: Authors own compilation from ODIN 2018/19 Methodology Report. https://docs.google.com/document/d/1ubPL1l_3im9bjlCVZ6W2ICAy6UAiXl1hGeA1aXImkxI/edit

*Scoring options for coverage element 1- Representative indicators are available and are disaggregated appropriately - varies by each indicator. Hence it is presented in Annex Table A2.

For the elements of data openness, scoring is calculated independent of the data coverage.

Table A5: ODB data categories covered in technical survey

Variable	Short Name	Long Name	Description
ODB.2013.D1	Map	Mapping data	A detailed digital map of the country provided by a national mapping agency and kept updated with key features such as official administrative borders, roads and other important infrastructure. Please look for maps of at least a scale of 1:250,000 or better (1cm = 2.5km).
ODB.2013.D2	Land	Land ownership data	A data set that provides national level information on land ownership. This will usually be held by a land registration agency, and usually relies on the existence of a national land registration database.
ODB.2013.D4	Stats	National statistics	Key national statistics such as demographic and economic indicators (GDP, unemployment, population, etc), often provided by a National Statistics Agency. Aggregate data (e.g. GDP for whole country at a quarterly level, or population at an annual level) is considered acceptable for this category.
ODB.2013.D5	Budget	Detailed budget data	National government budget at a high level (e.g. spending by sector, department etc). Budgets are government plans for expenditure, (not details of actual expenditure in the past which is covered in the spend category).
ODB.2013.D6	Spend	Government spend data	Records of actual (past) national government spending at a detailed transactional level; at the level of month to month government expenditure on specific items (usually this means individual records of spending amounts under $1m or even under $100k). Note: A database of contracts awarded or similar is not sufficient for this category, which refers to detailed ongoing data on actual expenditure.
ODB.2013.D7	Company	Company registration data	A list of registered (limited liability) companies in the country including name, unique identifier and additional information such as address, registered activities. The data in this category does not need to include detailed financial data such as balance sheet etc.
ODB.2013.D8	Legislation	Legislation data	The constitution and laws of a country, including national laws and statutes but excluding case-law and administrative regulations.
ODB.2013.D9	Transport	Public transport timetable data	Details of when and where public transport services such as buses and rail services are expected to run. Please provide details for both bus and rail services if applicable. If no national data is available, please check and provide details related to the capital city.
ODB.2013.D10	Trade	International trade data	Details of the import and export of specific commodities and/or balance of trade data against other countries.
ODB.2013.D11	Health	Health sector performance data	Statistics generated from administrative data that could be used to indicate performance of specific services, or the healthcare system as a whole. The performance of health services in a country has a significant impact on the welfare of citizens. Look for ongoing statistics generated from administrative data that could be used to indicate performance of specific services, or the healthcare system as a whole. Health performance data might include: Levels of vaccination; Levels of access to health care; Health care outcomes for particular groups; Patient satisfaction with health services.
ODB.2013.D12	Education	Primary and secondary education performance data	The performance of education services in a country has a significant impact on the welfare of citizens. Look for ongoing statistics generated from administrative data that could be used to indicate performance of specific services, or the education system as a whole. Performance data might include: Test scores for pupils in national examinations; School attendance rates; Teacher attendance rates. Simple lists of schools do not qualify as education performance data.

Table A5 continued

Table A5 continued.

Variable	Short Name	Long Name	Description
ODB.2013.D13	Crime	Crime statistics data	Annual returns on levels of crime and/or detailed crime reports.Crime statistics can be provided at a variety of levels of granularity, from annual returns on levels of crime, to detailed real-time crime-by-crime reports published online and geolocated, allowing the creation of crime maps.
ODB.2013.D14	Environment	National environmental statistics data	Data on one or more of: carbon emissions, emission of pollutants (e.g. carbon monoxides, nitrogen oxides, particulate matter etc.), and deforestation. Please provide links to sources for each if available.
ODB.2013.D15	Elections	National election results data	Results by constituency / district for the most all national electoral contests over the last ten years.
ODB.2013.D16	Contracting	Public contracting data	Details of the contracts issued by the national government.

Table A6: GODI data categories covered in technical survey

Category	What we look at?	Why we look at it?	Characteristics
Budget	National government budget at a high level. This is planned government expenditure for the upcoming year, and not the actual expenditure.To develop this category the Index drew on work from Open Spending.	Open budget data allows for well-informed publics. It showing what money is spent on, how public funds develop over time, and why certain activities are funded. See here a list of cases how budget data has been used in the past.	Following data must be online to qualify for assessment: • Budget for each national government department, ministry, or agency • Descriptions for budget sections • Level of granularityBudget separated into sub-department, political program, or expenditure type
Spending	Records of actual (past) national government spending at a detailed transactional level. Data must display ongoing expenditure, including transactions. A database of contracts awarded or similar will not be considered sufficient. Also, a database only showing subsidies will not be sufficient. To develop this category the Index drew on work from Open Spending.	Open spending data shows whether public money is efficiently and effectively used. It helps to understand spending patterns and to display corruption, misuse, and waste.	Following data must be online to qualify for assessment: • Government office which had the transaction • Date of transaction • Name of vendor • Nominal amount of individual transactionLevel of granularity • Individual record of each transaction
Procurement	All tenders and awards of the national/federal government aggregated by an office. It does not look into procurement planning or other procurement phases such as implementation (i.e. actual money transfers, which are part of our spending category). To develop this category the Index drew on work from the Open Contracting Partnership.	Open procurement data may enable fairer competition among companies, allow to detect fraud, as well as deliver better services for governments and citizens. Monitoring tenders helps new groups to participate in tenders and to increase government compliance.	Following data must be online to qualify for assessment: **Tender phase** • Tenders per government office • Tender name • Tender description • Tender status **Award phase** • Awards per government office • Award title • Award description • Value of the award • Supplier's name
Election results	This data category looks at results for the latest national electoral contest. Election data informs about voting outcomes and voting process. What are electoral majorities and minorities? How many votes are registered, invalid, or spoilt? The Index consulted the National Democratic Institute (NDI) to develop this data category, but did not take their latest recommendation which will be considered for the next edition. For more information, see the NDI's Open Elections Data Initiative.	To enable the highest level of transparency, the Index assesses polling station-level data. Polling stations are the locations at which voters cast their vote. Having this data allows for independent scrutiny of each stage of the voting and counting process. It also helps electoral stakeholders better target their voter education and mobilization efforts for the next elections.	Following data must be online to qualify for assessment: • Results for major national electoral contests (such as general elections) • Number of registered votes • Number of invalid votes • Number of spoiled votes (not required, if a digital voting system is assessed, that does not recognize spoiled votes) Level of granularity • Data available at polling station level

Table A6 continued.

Table A6 continued

Category	What we look at?	Why we look at it?	Characteristics
Company register	Lists of registered (limited liability) companies. The submissions in this data category do not need to include detailed financial data such as balance sheets.This category draws on the work of OpenCorporates.	Open data from company registers may be used to many ends: enabling customers and businesses to see with whom they deal, or to see where a company has registered offices.	Following data must be online to qualify for assessment: • Name of company • Company address • Unique identifier of the company • Register available for entire country (usually assessed through sample: it is answered with „Yes“ if a register indicates companies in different regions)
Land ownership	Maps of lands with parcel layer that displays boundaries. Also a land registry with information on registered parcels of land.The assessment criteria were developed in collaboration with Cadasta Foundation. For more information on land ownership data sets, see Cadasta Foundation's Data Overview.	The Index focuses on assessing open land tenure data (describing the rules and processes of land property). Responsible use may enable tenure security and increase the transparency of land transactions.	The following characteristics must be included in cadastral and registry information submitted. • Parcel boundaries • Parcel ID • Property Value (price paid for transaction or tax value) • Tenure Type (public, private, customary, etc.)
National maps	A geographical map of the country including national traffic routes, stretches of water, and markings of heights. The map must at least be provided at a scale of 1:250,000 (1 cm = 2.5km), a scale feasible for most countries. The Index developed this category based on a landmark report of the United Nations Committee of Experts on Global Geospatial Information Management (UNGGIM).	Geographic information is instrumental for many use cases, including journey planning, the mapping of topography, as well as demographic indicators.	Following data must be online to qualify for assessment: • Markings of national traffic routes • Markings of relief/heights • Markings of water stretches • National borders Coordinates - Note: To qualify, data must contain geographic projections that enable to interpret coordinates
Administrative Boundaries	Data on administrative units or areas defined for the purpose of administration by a (local) government.The development of this category draws on work of FAO Global Administrative Unit Layers (GAUL)project, as well as the UNGIWG.	Open data about administrative zones has many use cases: Who are the candidates in my region? Which government bodies administer my region? How is wealth distributed across regions? The Index assesses two administrative boundary levels (e.g. federal states = level 1, and municipalities = level 2).	Following data must be online to qualify for assessment: • Boundary level 1 • Boundary level 2 (not required, if country has only one level) • Coordinates of administrative zone (latitude, longitude) • Name of polygon • Borders of polygon - Note: To qualify, data must contain geographic projections that enable to interpret coordinates

Table A6 continued.

Table A6 continued

Category	What we look at?	Why we look at it?	Characteristics
Locations	A database of postcodes/zipcodes and the corresponding spatial locations regarding latitude and longitude (or similar coordinates in an openly published coordinate system). The data has to be available for the entire country. The Index drew on work of the Universal Postal Union to develop this category.	Open location data shows the addresses of public and private buildings. While mainly used to route postal services, this data has many use cases: to calculate the number of persons in a city district, to provide homes with services, or for direct mailing and marketing.	Following data must be online to qualify for assessment: • Zipcodes Addresses (required, if zip code does not include the address) • Coordinates (latitude, longitude) • Data available for entire country - Note: To qualify, data must contain geographic projections that enable to interpret coordinates
National statistics	Key national statistics on demographic and economic indicators such as Gross Domestic Product (GDP), or unemployment and population statistics. These statistics can be published as aggregates for the entire country.	As Open Data Watch states "Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy and the public with data about the economic, demographic, social and environmental situation."	Following data must be online to qualify for assessment: • Country Population (Required: census data, updated every year, Optional: vital statistics of birth and death) • Gross Domestic Product (measured in current or constant prices, updated quarterly, last update must not be more than 3 months ago) • National unemployment (absolute numbers, or expressed as percentage of entire population, updated quarterly, last update must not be more than 3 months ago)
Draft legislation	Data about the bills discussed within national parliament as well as votes on bills (not to be confused with passed national law). Data on bills must be available for the current legislation period. This data category draws on work by the National Democratic Institute (NDI) and the Declaration of Parliamentary Openness.	Open data on the law-making process is crucial for parliamentary transparency: What does a bill text say and how does it change over time? Who introduces a bill? Who votes for and against it? Where is a bill discussed next so that the public can participate in debates?	Following data is required. It must be online for the data to qualify for assessment: • Content of bill • Author of bill • Status of bill • Available for current legislation period Following data is assessed optionally (only if available): • Votes on bill per member of parliament • Transcripts of debates on bill • Note on optional data: This category is newly added in 2016. Not all data needs to be available online to qualify. The Index team used minimum requirements to explores how much data is currently available online. In future editions, the category may require more data elements.

Table A6 continued.

Table A6 continued

Category	What we look at?	Why we look at it?	Characteristics
National law	This data category requires all national laws and statutes to be available online, although it is not a requirement that information on legislative behaviour e.g. voting records is available.This data category draws on work by the National Democratic Institute (NDI) and the Declaration of Parliamentary Openness.	Access to open data on a country's legal code (i.e. national law) supports compliance with law, enables to keep track of legal changes, and also enables public deliberation around a law.	Following data must be online to qualify for assessment: • Content of the law / status • Date of last amendment • Amendments to the law (if applicable)
Air quality	Data about the daily mean concentration of air pollutants, especially those potentially harmful to human health. Data should be available for all air monitoring stations or zones in a country, including at least 3 major cities. The Index evaluates the openness of key pollutants as defined by the World Health Organisation (WHO).	Air quality is a key factor for human health and environment.	Following data must be online to qualify for assessment: • Particulate matter (PM) • Sulphur oxides (SOx) • Nitrogen oxides (NOx) • Carbon monoxide (CO) • Ozone (O3) • Available per air monitoring station (at least for 3 major cities Following data is assessed optionally (if available):Volatile organic compounds (VOCs)
Water quality	Water quality data by water source. The data category regards the quality of designated drinking water sources. If data on designated drinking water sources is not available, it refers to environmental water sources (lakes, rivers, groundwater). Data per each water source is desirable. But for this year the Index also accepted if a country only published country-wide aggregated reports. As the review shows, we either find local and granular data or aggregated national reports.	This information is essential for both the delivery of services and the prevention of diseases.	In order to satisfy the minimum requirements for this category, data should be available on level of the following chemicals: • Fecal coliform • Arsenic • Fluoride levels • Nitrates • Total Dissolved Solids • Data per water source • Available for the entire country

Table A7: GODI survey question and scoring

Question	Description	Score	Rationale
Is the data collected by government (or a third-party related or linked to government)?	Answer "Yes" if the chosen data is collected by the government, or a third party is officially representing the government. This is the case for state-owned-enterprises or contractors delivering public services for government. Answer "No" if one of the following cases apply: i) The data is collected by organisations that do not represent government; ii) The data is collected but not for the relevant government level; iii) The data is not collected at all	Not scored	Data collection by itself is not a characteristic of 'open' data. Our knowledge of edge cases or exceptions from the rule (such as legal arrangements of data publication in cases of public-private partnerships) is too limited to develop valid statements about a reasonable scoring.
Is the data available online without the need to register or request access to the data?	Answer "Yes", if the data is made available by the government on a public website. Answer "No" if the data are NOT available online or are available online only after registering, requesting the data from a civil servant via email, completing a contact form or another similar administrative process.	15 points	Online availability is a requirement for openness: everyone has to have online access to specific data. Furthermore, it is a condition for all following questions and mandatory registration can deter people from using data (focus on user perspective). We put emphasis on the additional requirement that data must also be available without mandatory registration
Is the data available online at all?	Tell us if the data is available online at all (after registering, after getting authentication.	Not scored	We currently do not aim to reward mandatory registration. Administrative processes may entail terms of use that contradict open data: such as agreeing to terms of use. A zero score is a indicates to governments that their way of online publication is not ideal for all user groups.
Is the data available free of charge?	The data is free if you don't have to pay for it.	15 Points	Data has to be for free in order to be accessible to everyone. We cannot expect users to pay for data sets in order to evaluate them for us. Some data (especially when provided in machine-readable file formats) have to be paid for.
Where did you find the data?	Indicate a URL and a description of the URL. Example: If you find data on a financial department website, please fill in: "Website of National Department of Finances". Sometimes you can find data in a lot of places in the web. To limit your search, tell us the first 5 URLs you can easily find for each source type. Make sure the URLs are from an official government source.	Not scored	This is a subjective assessment. The results may be affected by a submitter's topical expertise or familiarity with government websites.
How much do you agree with the following statement: "It was easy for me to find the data."	Submitters answer with a Likert scale.	Not scored	This is a subjective assessment. The results may be affected by a submitter's topical expertise or familiarity with government websites. We experiment with the results to develop a better findabillity assessment.

Table A7 continued.

Table A7 continued

Question	Description	Score	Rationale
Is the data downloadable at once?	Answer “Yes”, if you can download all data at once from the URL at which you found them. In case that downloadable data files are very large, their downloads may also be organised by month or year or broken down into sub-files. Answer “No” if you have to do many manual steps to download the data, or if you can only retrieve very few parts of a large data set at a time (for instance through a search interface).	15 Points	We score if a data set can be downloaded at once. This question therefore rewards the technical possibility to retrieve all data from the internet without having to download dozens of small pieces of information, getting access to data through a search interface only, sending requests, having captchas or other limits to download. Important note: data may be split into smaller sub-sets. This applies for instance for long time series, or large geospatial data. It is important that these sub-sets are logically linked, and that it is possible to retrieve data automatically from one or several URLs.
Data should be updated every [Time Interval]: Is the data up-to-date?	Please base your answer on the date at which you answer this question. Answer “No” if you cannot determine a date, or if the data are outdated.	15 Points	Some of the data we assess are most valuable right after their releases such as short-term weather forecasts, election results or budget data. Timely provision of these data is crucial. - Some data is not as time sensitive as others. Our scoring wants to strike a balance between both cases and therefore amounts to 15 points, in order not to avoid an over-emphasis of this category.
Is the data openly licensed/in public domain?	This question measures if anyone is legally allowed to use, modify and redistribute data for any purpose. Only then data is considered truly "open" (see Open Definition). • Answer ”Yes” if the data are openly licensed. The Open Definition provides a list of conformant licenses. Also, consult the terms of use which often indicate whether data can be freely reused • Answer “Yes” if there is no open license, but a statement that the data set is in “public domain”. To count as public domain, the data set must not be protected by copyright, patents or similar restrictions. • If you are not sure whether an open licence or public domain notice is compliant with the Open Definition 2.1, seek feedback on the Open Data Index discussion forum. Answer “No” whenever it is not fully evident that the license or terms of use are compliant with the Open Definition.	20 Points	Legal usability of data is a core requirement of the open definition. It is a prerequisite for unrestricted usability for everyone. Our old scoring was fairly high, emphasizing the legal usability of data. The current scoring is lowered to give us some space to stress other aspects of openness. This question will not lose its significance for openness (still scored higher than in the Open Data Barometer)

Table A7 continued.

Table A7 continued

Question	Description	Score	Rationale
Is the data in open and machine-readable file formats?	We automatically compare them against a list of file formats that are considered machine-readable and open. A file format is called machine-readable if your computer can process, access, and modify single elements in a data file.The Index considers formats to be "open" if they can be fully processed with at least one free and open-source software tool. Potentially these formats allow more people to use the data because people do not need to buy specific software to open it. The source code of these format does not have to be open.	20 points	Both features (machine-readable and open format) are key aspects of the open definition. Machine-readability is a major enhancement of technical usability. However, if a file is only usable with proprietary software (such as ArcGIS) 'normal' users are exempt from using them. Open formats put no copyright, monetary restrictions or other restrictions on their use (important for people who cannot / do not want to afford proprietary software).
How much human effort is required to use the data. (1 = little to no effort is required, 3 = extensive effort is required)	The submitters tell us their use case and the steps they took to make the data usable (example: "I have to reformat the data").	Not scored	The question is a subjective assessment. Furthermore, usability depends on context and the purposes for which a person wants to use the data.

Table A8: Weights of Features of the MENA Micro Access Indicator (MENA MAI)

Element	Feature	Weight in overall score	
User experience	English translation	1/6	1/2
	Micro data tab/dashboard/library	1/6	
	Survey calendar	1/6	
Data openness	Firm data	1/14	1/2
	Price data	1/14	
	Consumption data	1/14	
	Labor Force data	1/14	
	Health data	1/14	
	Population Census	1/14	
	Economic Census	1/14	

Table A9: 2021 MENA Microdata Access Indicator Scores

		User experience element				Data access element								Overall score
S/N	**Economy**	English translation	survey calendar	microdata library	**user experience sub score**	Establishment survey	Price survey	Consumption survey	Labor force survey	Health survey	population census	Economic census	**Data openness sub score**	**Overall score**
1	Algeria	0	100	0	33.3	0	0	0	0	0	0	0	0.0	16.7
2	Bahrain	100	0	0	33.3	0	0	0	0	0	0	0	0.0	16.7
3	Djibouti	0	0	100	33.3	0	0	50	0	0	0	0	7.1	20.2
4	Egypt, Arab Rep.	100	0	100	66.7	0	0	50	0	50	0	50	21.4	44.1
5	Iran, Islamic Rep.	100	0	0	33.3	0	0	50	50	0	50	0	21.4	27.4
6	Iraq	100	0	0	33.3	0	0	50	0	50	0	0	14.3	23.8
7	Jordan	100	0	100	66.7	0	0	0	0	50	0	0	7.1	36.9
8	Kuwait	100	50	0	50	0	0	0	0	0	0	0	0.0	25.0
9	Lebanon	100	0	0	33.3	0	0	0	100	0	0	0	14.3	23.8
10	Libya	50	0	0	16.7	0	0	0	0	0	0	0	0.0	8.4
11	Malta	100	100	0	66.7	0	0	50	50	0	0	0	14.3	40.5
12	Morocco	0	0	0	0	0	0	100	0	0	100	0	28.6	14.3
13	Oman	100	100	0	66.7	0	0	0	0	50	0	0	7.1	36.9
14	Qatar	100	100	0	66.7	0	0	0	0	0	0	0	0.0	33.4
15	Saudi Arabia	100	100	0	66.7	0	0	0	0	0	0	0	0.0	33.4
16	Syrian Arab Republic	50	0	0	16.7	0	0	0	0	0	0	0	0.0	8.4
17	Tunisia	100	100	100	100	0	0	100	100	50	0	0	35.7	67.9
18	United Arab Emirates	100	100	0	66.7	0	0	0	0	0	0	0	0.0	33.4
19	West Bank and Gaza	100	100	100	100	50	0	50	50	50	50	50	42.9	71.4
20	Yemen, Rep.	0	0	100	33.3	0	0	0	0	0	0	0	0.0	16.7