

Data Transparency in the Middle East and North Africa

Asif M. Islam



WORLD BANK GROUP

Middle East and North Africa Region

Office of the Chief Economist

March 2023

Abstract

Data transparency about critical economic issues may be key to driving growth and enhancing trust in government in the Middle East and North Africa. Several knowledge products and technical analyses on the region have been greatly constrained by the lack of availability of detailed data, and the relatively outdated nature of many available datasets. The goal of this study is to ascertain the state of data systems in the Middle East and North Africa region. Through analysis of several indicators, with their limitations in mind, the study uses descriptive analyses and uncovers six stylized facts of the region: (i) developing economies in the Middle East and North Africa have poor data ecosystems, largely due to the prevalence of conflict; (ii) developing economies in the Middle East and North Africa as a group have experienced the largest deterioration in data systems over time; (iii) data systems in richer economies in the

Middle East and North Africa region underperform relative to their income peers; (iv) Gulf Cooperation Council economies underperform in data openness, especially online access, despite having the resources for online features; (v) the regulatory framework for data (data infrastructure) is poor throughout the region, especially in Gulf Cooperation Council economies; and (vi) the dispersion of source data scores—a measure of availability and timeliness of micro data—in the region suggests that national statistical offices in the region could learn from each other. Furthermore, the study summarizes data availability and timeliness for specific macro, micro, and public health indicators for countries across the region. The need for forging a social contract for data is discussed, as well as the role international institutions can play through a statistics compact for the region.

This paper is a product of the Office of the Chief Economist, Middle East and North Africa Region. 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 author may be contacted at aislam@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Data Transparency in the Middle East and North Africa

Asif M. Islam¹

JEL Codes: C81, C82, D83, EO1

Keywords: Statistical capacity, data openness, data access, transparency, MENA

¹ Asif M. Islam is a Senior Economist in the Chief Economist Office of the Middle East and North Africa at the World Bank. The analysis benefited greatly from feedback and conversations with Johannes Hoogeveen, Minh C. Nguyen, Umar Serajuddin, and Brian William Stacy. The opinions expressed in this paper do not represent the views of the World Bank Group, its Board of Directors, or the Governments they represent. All errors and omissions are the author's responsibility.

Data Transparency in the Middle East and North Africa

Introduction

Modern and well-coordinated traditional data collection systems are a feature of advanced economies. Granting access to the research community has been characteristic of these data systems. However, many developing economies have lagged in their statistical capacity to generate data. Furthermore, when data is available, many economies have shied away from sharing it with the wider research community. This is particularly the case for many economies in the Middle East and North Africa (MENA) region. While the fears of data privacy are not unfounded, little attention has been afforded to the costs of inadequate data systems that persist in silos and are unavailable to the best minds to advance the boundaries of knowledge.

The objective of this study is to ascertain the state of data systems in the MENA region, while making the case that the long run benefits of good quality data are considerable. Through analysis of several indicators, the study uncovers six stylized facts of the region: (i) developing MENA economies have poor data ecosystems, largely due to the prevalence of conflict; (ii) MENA developing economies as a group have experienced the largest deterioration in data systems over time; (iii) data systems in richer economies in the MENA region underperform relative to their income peers; (iv) Gulf Cooperation Council (GCC) economies underperform in data openness, especially online access, despite having the resources for online features; (v) the regulatory framework for data (data infrastructure) is poor throughout the region, especially in GCC economies; and (vi) dispersion of source data scores – a measure of the availability and timeliness of micro data – in the region suggests that National Statistical Offices (NSOs) in the region could learn from each other.

Data transparency about critical economic issues will be the key to driving growth and enhancing trust in government in the MENA region (Arezki et al., 2020 – April 2020 MEU). The need for greater transparency has been more vivid during the pandemic when the prevalence of real time information has been crucial. Furthermore, the onset of the war in Ukraine has generated shockwaves throughout the region, and lack of data makes it hard to establish the effects on the poor and vulnerable. Equally important for economic and social outcomes is the sharing of data for private sector stakeholders and citizens alike. The free and uncensored exchange of information can improve governance, citizen trust and accountability within the MENA region, while encouraging investments. Several knowledge products and technical analyses on the MENA region have been greatly constrained by the lack of availability of detailed data about the region, and the relatively outdated nature of many available datasets. MENA's transformation agenda must include data transparency as one of its main pillars.

A core component of the Sustainable Development Goals is that countries must generate their own indicators with their own capabilities. The costs of such efforts are not trivial, and the trade-offs between investing in data capacity and systems versus other pressing needs are difficult for developing country governments to ascertain. Without significant investments and a change in the mindset or a paradigm shift towards data transparency, the MENA region risks being left behind with accelerating data gaps with the increasing digitization of data. The 2021 World Development Report on Data for Better Lives makes the case that a new social contract for data is needed in the region, and globally. The study summarizes what the conditions would be for such a contract, and how the MENA Statistics Compact can advance the data agenda.

Why Data Systems Matter

The long run benefits of good data systems and data transparency are considerable (Islam and Lederman, 2020; World Bank (2021) - WDR 2021: Data for Better Lives). There are several channels through which the quality of the data ecosystem could aid development. First, data serves as the basis for policy and reforms. Policies can only be as good as the data they are based on. At a fundamental level, data is about records. Take the example of a business. A manager has the primary goal of raising profits. To achieve this, performance must be benchmarked historically and compared with competitors. Collateral must be evaluated and leveraged to access financing to pursue new ventures. Risk and reward must be balanced. Investors need to be enticed. Without any record keeping, none of these can be achieved. A similar set of challenges face governments and policy makers in the MENA region. Countries need to grow, and to expand options, data must be harnessed to provide guidance. Countries need to compare with past performance and benchmark against other countries at similar levels of development. Countries with high quality information are better able to make better decisions. Through data and evaluation, existing policies may be reformed and refined, while new policies may be experimentally evaluated.

Second, data that is accessible to a wider group of agents can generate better policies and reforms. A powerful tool confined to the tool shed has limited value. When data is accompanied by a large base of sophisticated users, there are substantial expansions in the frontier of knowledge. As researchers test hypotheses, debate and dispute findings from each other, while establishing robust facts and relationships using the same data, the best ideas and new pathways of addressing challenges can emerge. This is the process where simple data can be turned into a stock of knowledge that could generate and direct further knowledge creation and better policy making. It is not surprising that richer economies are researched more than poorer economies. Publications in top ranking journals in the economics profession are skewed towards wealthier economies, which themselves capitalize on this analysis to better understand and improve their economies. Less wealthy economies, on the other hand, cannot benefit from such knowledge, largely because of the lack of data accessibility.

Third, when data is poor or unavailable, the gap between perceptions and reality may grow. Important reforms may lead to real welfare improvements but may have little impact on public perceptions as there is limited data tracking such improvements. These perceptions may foster a narrative resulting in frustrations that manifest in the form of social protests and unrest. Similarly, if data is of dubious quality, the public may lose confidence in such information and may not alter their perceptions despite positive findings from the data. This may contribute to a lack of trust between citizens and their governments, weakening the social contract. More importantly, once a government treads down the path of unreliable or limited data, it may be challenging to trek back. The public may be less willing to trust information from the government, and thus perceptions are calcified and difficult to change resulting in economies that are more prone to social upheavals.

However, for data to inform policy, availability and accessibility of good quality data is necessary but not enough. Several institutions play an important role in facilitating debate fostered by data to ensure policy is an outcome of the process and the public takes ownership of this debate. The process typically resembles the following path. Data is first generated which is not easily digestible by the public at large. Academics generate knowledge from the data that fosters debate among themselves. Think tanks and policy makers join in. Media institutions communicate the information to the public at large who then participate and take ownership of the debate. The outcome of this process is hopefully an optimal set of policies that promote overall welfare. Each of the institutions is critical in fostering and enveloping the public into meaningful debate. Failure of the media to inform the public, or the absence of think tanks in the process would limit the value of information in guiding policy and diminish the returns of investment in data. All the parts must move together for the data ecosystem to be effective.

A Brief Primer on the Data and Transparency Literature

Centuries ago, rulers acknowledged the need for systematic data collection. Historical data collection can be traced to the first recorded census undertaken by Babylonians around 3800 BC (Grajalez et al., 2013). Early data collection served rulers with the aim of accounting for wealth and power. Information was gathered for taxation purposes, counting of men for military recruitment and workforce, and ascertaining conquered populations and territories (World Bank 2021). Data was kept secret from the public and not meant to improve their lives but to help rulers pursue their agenda. This raised general distrust of data collection activities among the populace. In contrast, the enlightenment ideals in eighteenth century Europe emphasized objective scientific inquiry. The role of data evolved to a means of examining society and became more public. In the late eighteenth century, statistical agencies were set up in Europe and North America to publish official statistics and inform the public.

Data transparency is defined as the regular publication of credible statistics by the state. This includes the frequency and availability of micro-data, socio-economic indicators, and adherence of data to international standards. The literature has acknowledged the importance of statistics and the benefits of evidence-based policy in terms of generating reforms through experimentation (Rodrik, 2010). However, theoretical arguments have been made that better statistics may not necessarily lead to reforms. Binswanger and Oechslin (2020) theoretically show that under certain political conditions better statistics can inhibit reforms. For instance, with better statistics, voters are less inclined to give politicians the benefit of the doubt when confronted with disappointing economic data, thereby weakening politicians' incentives to experiment and carry out reforms in the first place. Thus, the benefits of data transparency warrant empirical validation. However, recent research indicates that democracies are more likely to transform talk of reform during economic downturns into self-correcting reforms than less democratic societies (Arezki et al. 2020).

The role of transparency in facilitating good institutions by limiting corruption has been well debated. Press freedom and political institutions that increase accountability have been found to limit corruption across economies (Brunetti and Weder, 2003; Lederman et al., 2005; McMillan and Zoido, 2004). As noted by Djankov et al. (2003), two challenges in institutional design are the control of disorder and dictatorship, and transparent governments are better able to control disorder while incurring few social losses. Cordis and Warren (2014) find that strengthening the Freedom of Information Act (FOIA) in the United States reduces corruption and increases the likelihood that corrupt acts will be uncovered. In Mozambique, Armand et al. (2020) implemented a large-scale field experiment following the dissemination of information about a substantial natural gas discovery to test whether information can limit the political resource curse. The study finds that when information reaches citizens, it increases mobilization and reduces violence, but when information does not spread beyond local leaders, it has negative effects such as elite capture and increase in rent-seeking activities.

However, studies have also shown that transparency is not always desirable, or it is insufficient to limit corruption. Bac (2001) argues that transparency can have adverse effects by providing information to outsiders on who to bribe. Prat (2005) contends that complete transparency is not always beneficial and the type of information matters. This is part of the rationale of executive privilege - an agent may care more about appearances than engage in frank discussions in decision-making if it is understood that the internal discussions will be made public. Furthermore, information alone may not be sufficient to produce positive outcomes if there is little political engagement by citizens (Khemani et al., 2016). More importantly, information has to have a high signal-to-noise ratio to be impactful (Kosec and Wantchekon, 2020).

Studies have empirically explored elements of data transparency on governance, while not explicitly labeling it as such. For instance, Islam (2006) constructs a transparency index based on the frequency and availability of 11 indicators from four sectors (real, fiscal, financial, and external) and examines its effect

on measures of governance. Hollyer et al, (2011) create a transparency index based on the availability of country-level data – 172 variables in total categorized under Economic Policy and Debt in the World Bank’s World Development Indicators – to show that democratic countries are more transparent. Similarly, Williams (2009) constructs a transparency index based on the quantity of data released by governments through the availability of socio-economic data contained in the World Development Indicators and the International Finance Statistics databases. Choi and Hashimoto (2018) proxy for data transparency policy reforms using subscriptions to the IMF’s Data Standards Initiatives and find that such reforms lead to falls in sovereign bond spreads. The study finds that the larger quantity of data produced, as measured by the index, is positively correlated to the quality of bureaucracy, investment and financial sector development.

Finally, there is the question of whether data transparency can contribute to long term growth. The role of data transparency in facilitating economic growth is closely tied to institutions in the literature. The quality of institutions is largely dependent on their degree of accountability and corruption. Institutions operating in secrecy – lack of transparency - tend to be poor in quality as public officials can seek personal gain with little accountability. A large body of literature documented the detrimental effects of corruption on economic growth (Mauro 1995; Ehrlich and Lui, 1999). The secretive nature of corruption makes it distortionary, thereby discouraging investment and slowing growth (Shleifer and Vishny, 1993). Corruption can discourage foreign direct investment, entry of new firms, and misallocate public expenditures (De Soto, 1989; Mauro, 1998; Wei, 2000; Brada et al., 2019). Furthermore, institutions or even corruption that is “unpredictable” discourages investment and lowers growth (Campos et al., 1999). Corruption may also reduce parental investments in human capital (Varvarigos and Arsenis, 2015; Brunetti et al., 1998). Islam and Lederman (2020) find a statistically significant impact on transitional growth to a higher potential level of gross domestic product per capita. The estimates indicate an elasticity of the magnitude of 0.03 percent per year, which is much larger than the elasticity of trade openness and schooling in their estimation sample. Thus, there is some evidence that better data ecosystems can have far-reaching economic benefits.

Measuring the Performance of Data Systems

The ideal data transparency measure covers the capacity of the data system to produce timely and good quality data that is widely accessible. Additional elements include the legal framework surrounding data and the features or services data systems provide. There are three main measures that are widely used: (i) the Statistical Capacity Indicator (SCI) from the World Bank (that is to be discontinued), (ii) the Statistical Performance Indicator (SPI) also from the World Bank released as part of WDR 2021, and (iii) the Open Data Inventory (ODIN) from the Open Data Network.²

The SCI is a composite score assessing the capacity of a country’s statistical system. It is based on a diagnostic framework assessing the following areas: methodology; data sources; and periodicity and timeliness. Countries are scored against 25 criteria in these areas, using publicly available information and/or country input. The SCI score is then calculated as the simple average of all three area scores on a scale of 0-100. The SCI only covers developing economies but has the advantage of going back historically to 2004/2005, allowing for tracking the evolution of data systems over time. See Table A2 in the annex for definitions of subcomponents.

The SPI, which is the successor to the SCI, measures the capacity and maturity of national statistical systems by assessing the use of data, the quality of services, the coverage of topics, the sources of information, and the infrastructure and availability of resources. The goal is to improve development outcomes and track progress toward the Sustainable Development Goals. The Statistical Performance

² See Table A1 in the annex for a summary of all three indicators.

Indicators (SPI) is a framework of 5 pillars and 22 dimensions to assess the maturity of national statistical systems. The five pillars include variety of data users (*Data use*); the provision of services that connect users and producers such as online access, IMF SDSS subscription, and NADA data catalogue (*Data services*); production of indicators that track the SDGs (*data products*); availability and frequency of micro data (*data sources*); and *data infrastructure*, which encompasses adherence to international standards, legislation, skills, and partnerships. The SPI is only available from 2016 and is still a work in progress. For instance, the pillar on data use only covers the variety of international organizations as users of data. The pillar on data infrastructure only covers legislation. The data sources pillar is limited in coverage of administrative data. The SPI has built on the SCI in several meaningful ways. It is global in coverage, includes several additional sources of micro data, and includes additional dimensions of openness under the category of *data services* (see table A5). See Table A3 in the annex for definitions of SPI subcomponents.

The Open Data Inventory (ODIN) measures how complete a country's statistical offerings are and whether their data meet international standards of openness. Data assessed in ODIN must be official country data that are published on the national statistics office's (NSO) website or any other official country website that is linked from the NSO website. Coverage scores are based on the availability of key indicators and appropriate disaggregation over time and for geographic subdivisions. These encompass social statistics, economic and financial statistics, and environmental statistics. Openness scores are based on whether data can be downloaded in machine-readable and non-proprietary formats, are accompanied by metadata and whether download options exist such as bulk download and user-selection or APIs, and have an open terms of use or data license. Thus, this openness measure is really one of online access restricted to NSO or NSO-linked websites. Access to microdata is excluded and so is adherence to international standards. See Table A3 in the annex for definitions of SPI subcomponents.

The correlation matrix between the three indicators is presented in table 1. The SCI and SPI are highly correlated (0.86), while the SPI is also highly correlated with ODIN (0.76). The pairwise correlations are statistically significant at the 1% level.

Table 1: Correlation between SCI, SPI, and ODIN

	Statistical Capacity Indicator (SCI)	Statistical Performance Indicator (SPI)	Open Data Inventory (ODIN)
Statistical Capacity Indicator (SCI)	1		
Statistical Performance Indicator (SPI)	0.862***	1	
Open Data Inventory (ODIN)	0.672***	0.761***	1

***statistically significant at the 1% level

Six Stylized Facts about Data Systems in MENA

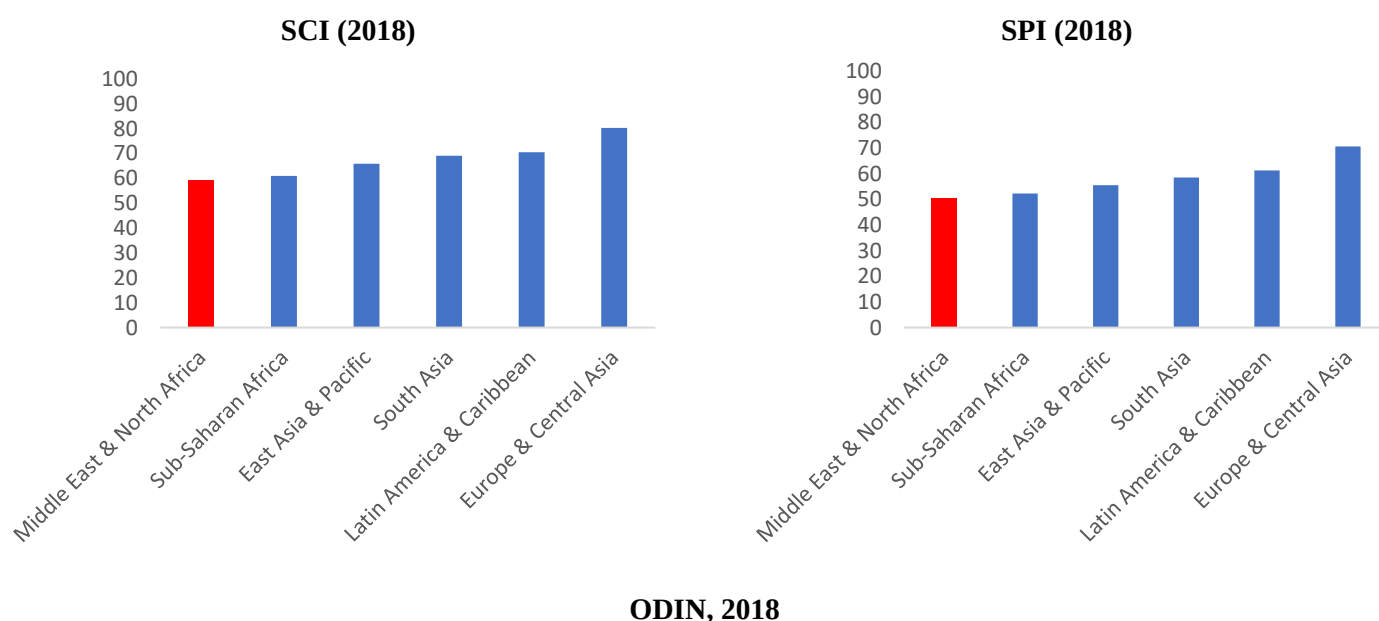
The stylized facts below describe the state of data systems across the MENA region. Data systems in MENA are characterized on the one hand by underperforming developing economies due to the prevalence of conflict, and on the other hand by rich economies that are severely underperforming relative to their income

peers. In the middle are several economies making strides for better data systems but are still far away from the frontier. Do note that the cross-sectional comparison of data systems in MENA is largely based on the year 2018, which is the year when data availability overlaps for the SPI, SCI, and ODIN.

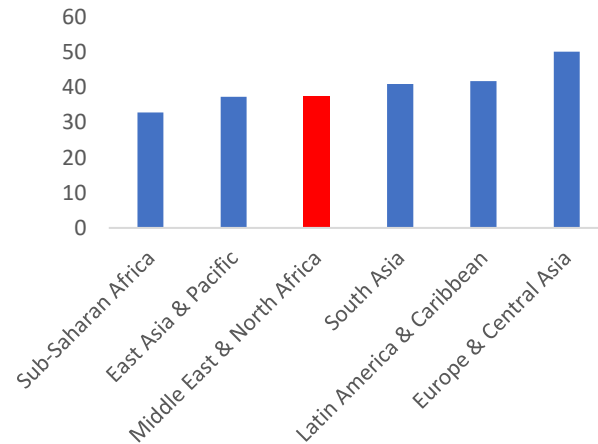
Stylized Fact 1: Developing MENA economies have poor data ecosystems, largely due to the prevalence of conflict

Data systems across the developing MENA economies are poor in comparison to other developing regions. The MENA region scores the lowest for both the SCI and SPI (figure 1). The ODIN overall score places MENA at a virtual tie with East Asia and the Pacific, both being above Sub-Saharan Africa. This is mainly due to a higher ODIN coverage score for the MENA region, indicating a higher prevalence of indicators in NSO websites available at a disaggregated level. Note that ODIN does not consider access to micro data or the adherence of data to international standards.³

Figure 1: Data Systems across Developing Economies by Region

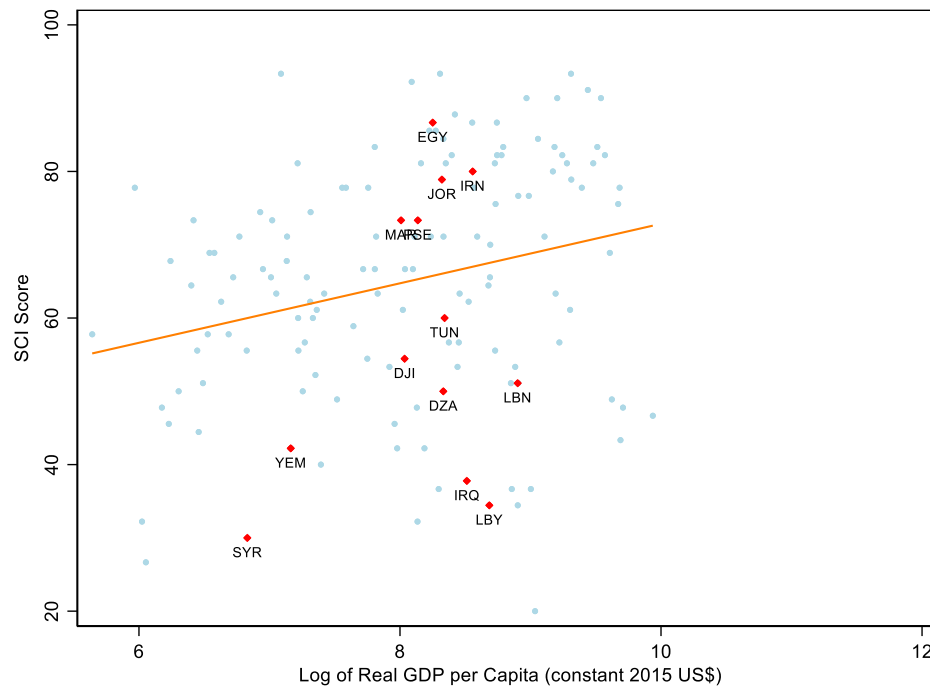


³ One element not discussed in the study is the openness and coverage of gender statistics as provided by the ODIN indicator. The MENA region has the second lowest score of all regions, but it is quite similar to the other regions except for North America and Europe and Central Asia.



The low performance in the SCI score for the developing MENA region is largely, but not completely, due to the prevalence of economies in conflict or fragile situations. The low scoring economies include the Syrian Arab Republic, the Republic of Yemen, Iraq and Libya. Figure 2 shows a positive correlation between SCI and the log of real GDP per capita for 129 developing economies. Countries such as the Arab Republic of Egypt, Jordan, and the Islamic Republic of Iran tend to perform better than their income peers. However, they remain far from the frontier, even for developing economies.

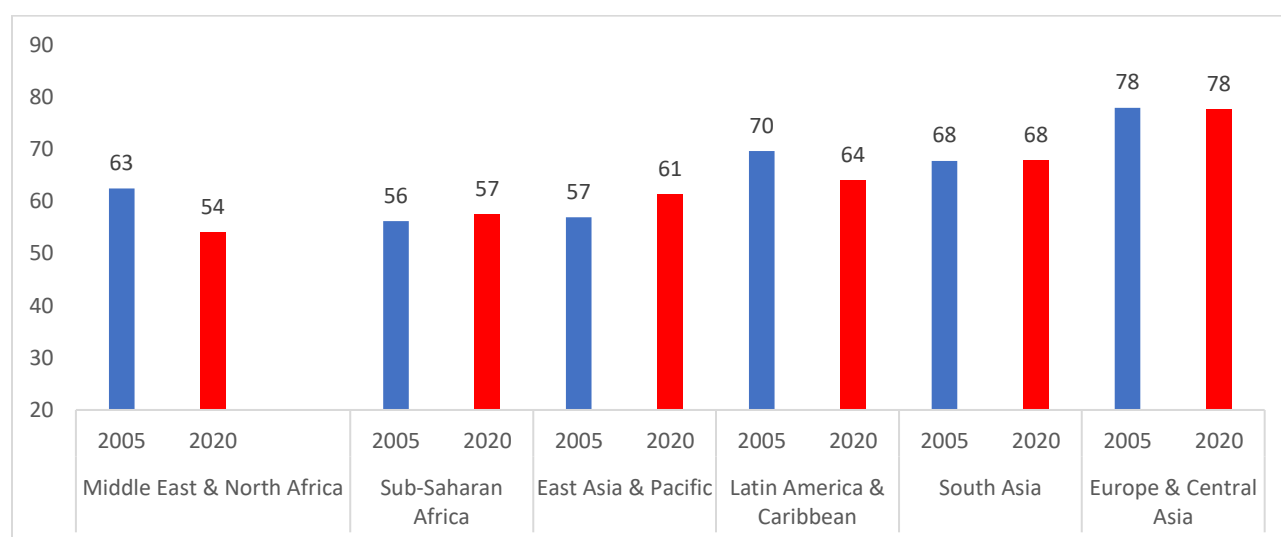
Figure 2: SCI - MENA Developing Economies and Income Peers



Stylized Fact 2: MENA developing economies as a group have experienced the largest deterioration in data systems over time

The statistical capacity indicator (SCI) is the only indicator of the three examined that can trace the evolution of data systems over a long period of time. The SCI is available only for developing economies. Although data is available from 2004, data for the bulk of countries are largely available from 2005. Therefore, figure 3 provides the SCI scores by region for 2005 and 2020. The MENA region shows the largest fall in SCI scores within this timeframe, despite starting at a low base. This drop scores it below Sub-Saharan Africa to be the lowest scoring region among all the regional groupings of developing economies. The reasons for the fall vary across countries, with some aging out of standards based on the recency of data, or the absence of micro and macro indicator data. Conflict also plays a big role with Libya and Yemen experiencing large declines in data systems over this period.

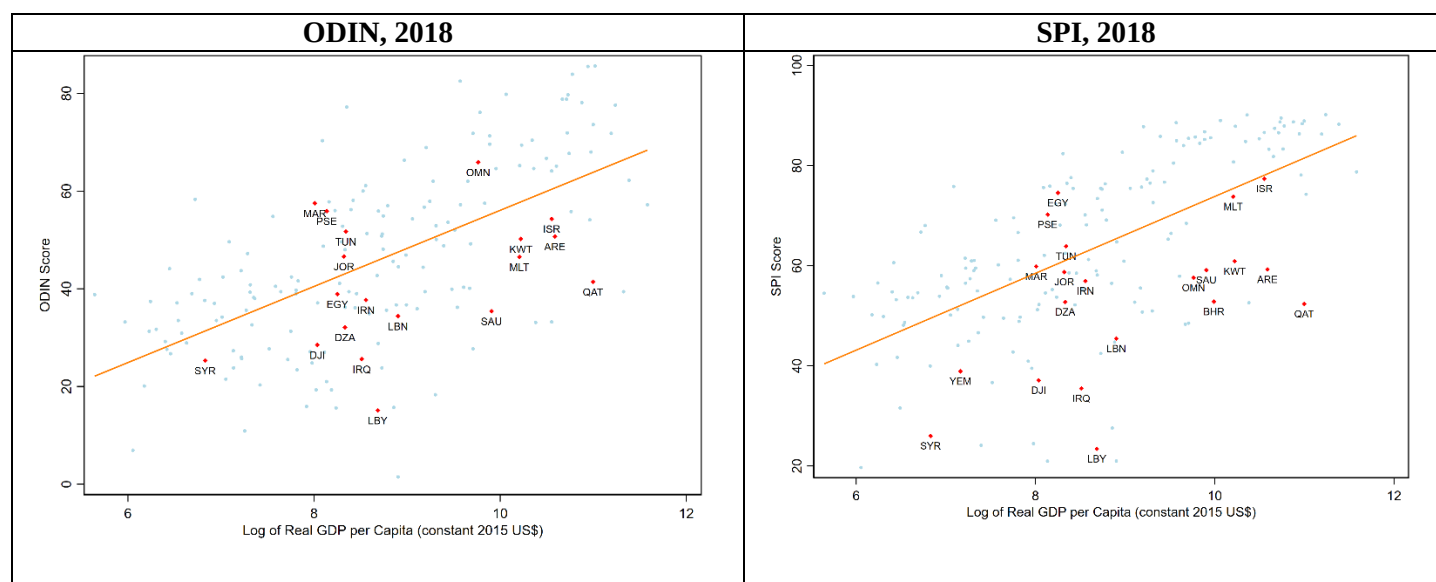
Figure 3: SCI – 2005-2020



Stylized Fact 3: Data systems in richer economies in the MENA region underperform relative to their income peers

A consistent pattern across MENA economies is the underperformance of data systems in the GCC economies relative to their income peers. Figure 4 presents the correlation between the ODIN score and log of GDP per capita, and the correlation between the SPI score and log of GDP per capita. Both indicators, unlike the SCI, cover both developing and advanced economies. Figure 4 shows that all GCC economies underperform relative to their income peers in the SPI index. This pattern is also replicated for the ODIN score, although Oman is the only exception, outperforming its income peers. The ODIN score however is more limited in scope by only focusing on NSO websites, and the online access features in these websites. A striking finding is that there are developing economies in the MENA region that overperform with respect to their income peers, in many cases even better than GCC economies, although they remain far from the frontier countries.

Figure 4: SPI and ODIN Scores



Stylized Fact 4: GCC economies underperform in data openness despite having better resources

There are largely two sub-indicators that try to capture data openness – the ODIN openness score and the SPI data services score. The ODIN openness score evaluates whether data can be downloaded in machine-readable and non-proprietary formats, are accompanied by metadata and whether download options exist such as bulk download and user-selection or APIs, and have an open terms of use or data license. This score captures online access features restricted to NSO or NSO-linked websites. The SPI data services score includes the ODIN openness score but adds the NADA micro data catalogue indicator and the IMF SDDS/e-GDDS subscription. Subscription to the IMF SDDS/e-GDDS entails (i) committing to using the e-GDDS as a framework for statistical development; (ii) designating a country coordinator; and (iii) preparing metadata that describe (a) current practices in the production and dissemination of official statistics, and (b) plans for short- and longer-term improvements in these practices.⁴ The NADA Data Catalog is an open-source software designed for researchers to browse, search, compare, apply for access and download research data.⁵ Thus the SPI data services score provides a broader measure of openness. Figure 5 provides the correlation of the openness scores with log of real GDP per capita. With the SPI data services scores, all GCC economies underperform considerably with respect to their income peers. In fact, no GCC economy matches the openness score for 5 developing MENA economies. For the narrower ODIN openness score that is constrained to the NSO online access, Oman is the only GCC economy that outperforms its income peers. West Bank and Gaza, Jordan, and Tunisia have higher ODIN openness scores than Saudi Arabia and Qatar. Given that much of the measure covers online features of NSO websites, it is surprising that countries with more resources are performing worse than economies with fewer resources. Note that data openness, as defined here, should not be confused with access to micro data. Morocco ranks high in the data openness scores below, despite the lack of access to recent micro data.

⁴ <https://dsbb.imf.org/e-gdds/overview>

⁵ <https://nada.ihsn.org/>

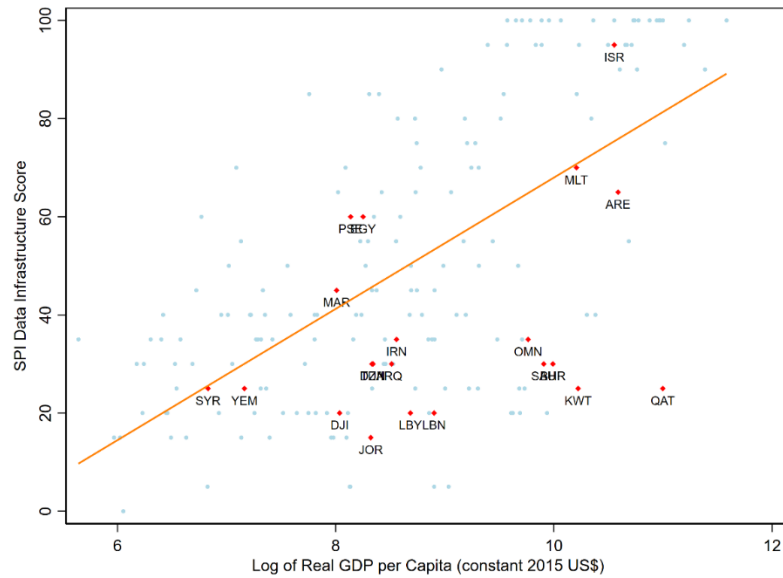
Figure 5: Data Openness Scores



Stylized Fact 5: The regulatory framework for data (data infrastructure) is poor across the region, especially in GCC economies

In its current form, the SPI score on data infrastructure largely covers legislation. The indicator is supposed to also capture data skills/literacy, but this information is thus far unavailable. Figure 6 shows a positive correlation between the SPI data infrastructure score and the log of real GDP per capita. Economies across the MENA region underperform with respect to the presence of data legislation. Only West Bank and Gaza, Egypt, Israel, and to some extent Morocco outperform their income peers. Advanced economies such as Qatar and Kuwait do no better than fragile and conflict economies such as the Syrian Arab Republic and Yemen in the data infrastructure score. There is considerable room for improvement in this indicator.

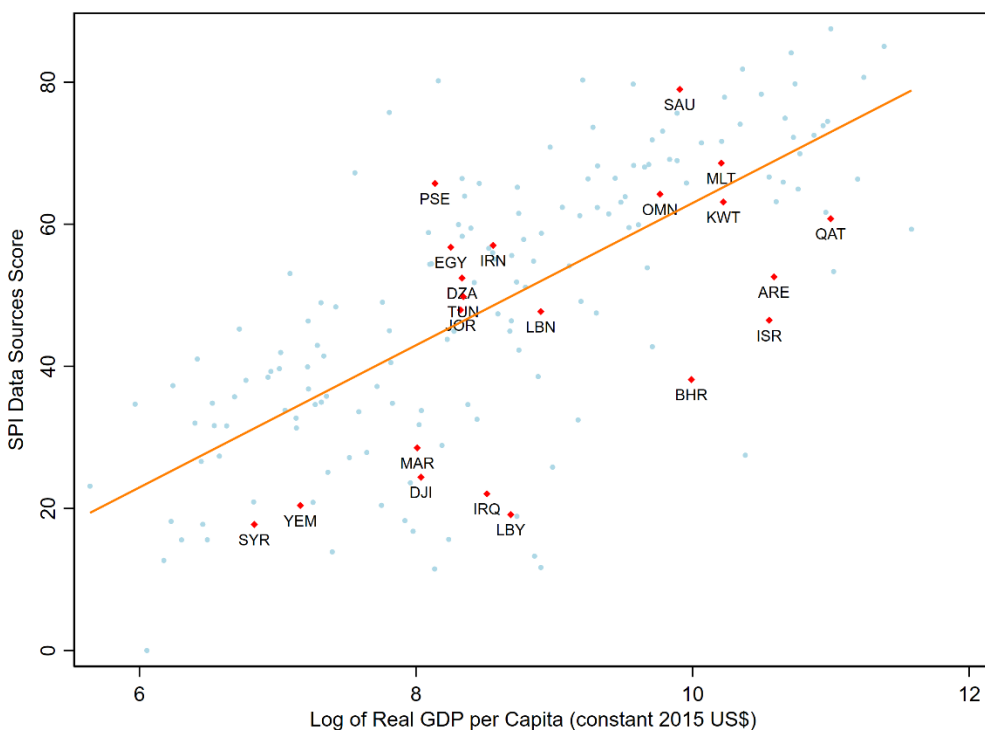
Figure 6: SPI Data Infrastructure (legislation) Scores



Stylized Fact 6: Dispersion of source data scores in the region suggests that NSOs in the region could learn from each other

There is much heterogeneity in the region with regards to the production of micro data (figure 7). At the lower end are conflict economies such as the Syrian Arab Republic, Yemen, Iraq, and Libya. Morocco, surprisingly, falls close to this group of economies. On the other hand, economies such as Saudi Arabia, Oman, Egypt, the Islamic Republic of Iran and West Bank and Gaza considerably outperform their peers. Given that the production of micro data is largely in the domain of NSOs, there are opportunities for NSOs across the region to learn from each other on ways to improve and increase collection of foundational micro data.

Figure 7: SPI Source Data Scores



In-Depth Exploration of Accessibility and Availability of Data

The data transparency indicators thus far have provided a holistic description of the data systems. This section provides an in-depth exploration of the accessibility of data by focusing on specific indicators and surveys. These include (i) micro survey data; (ii) macroeconomic data on employment, GDP, Production, and debt; and (iii) public health data. This is achieved using a “mystery client” approach whereby various government websites including NSOs are visited and the availability and accessibility of data at that point in time are evaluated.

(i) Availability and Accessibility of Micro Survey data

The availability of a wide range of survey data is documented by Ekhtor Mobayode and Hoogeveen (2021) for 20 MENA countries (including Malta). Several types of censuses and surveys are explored including establishment surveys, price surveys, consumption surveys, labor force surveys, health surveys, population census, and economic census (table 2). Census and survey data are often out of date. Only 13 of the 20 countries are current on their population census and nine out of 20 are up to date on their economic census. Only five of the 20 countries carried out an establishment survey recently and about half are up to date on their health, labor force, and consumption surveys. Furthermore, public accessibility of microdata is limited. Of the 140 potential microdata sets explored (seven data categories in 20 countries), 78 had been collected and 22 were accessible. Moreover, about a third of these 22 datasets were not accessible through the website of the National Statistics Office but had to be found on international microdata repositories, such as the World Bank, International Household Survey Network (IHSN) microdata library, IPUMS, Eurostat, Demographic and Health Surveys (DHS), and Multiple Indicator Cluster Surveys (MICS).

Table 2. Status of survey data (with year collected) on public facing website of MENA NSOs

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 economy
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)	No	No	2/7
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)	No	No	No	No	No	1/7
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	No	No	No	Yes (2014)	Yes (2014)	2/7
Total economies indicating collection	5/20	19/20	12/20	11/20	9/20	13/20	9/20	

of recent data for
each data category

Total recent data
indicated to have
been collected
across all economies

78/140

Source: Ekhtor Mobayode and Hoogeveen (2021). 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>.

(ii) Frequency of Macro Economic Data in MENA

While the SCI, SPI, and ODIN indicators capture the quality of the data ecosystem to a large extent, they do not necessarily encompass the extent and availability of high-frequency data. Table 3 presents the availability of high-frequency data on GDP, industrial production, and unemployment for 19 economies in the MENA region. The date of the most recent data available is also indicated. This information is compiled from the websites of various statistical offices or data portal initiatives, central banks and ministries of planning, economy, or finance across the MENA region. The findings are benchmarked against Mexico, which serves as a good comparator for the MENA region given its upper-middle-income economy status and a well-functioning data ecosystem. Of the 19 economies in the MENA region, 15 report quarterly data on GDP. Some economies lack information for the year 2020 entirely. Economies in conflict such as Libya (2014) and Yemen (2017) have outdated data. Only 10 of the 19 MENA economies have monthly or quarterly information on industrial production—for the remaining nine economies, information is not readily available. Only eight economies report quarterly unemployment data, while none have monthly data. The benchmark country, Mexico, reports unemployment data monthly. A caveat is that the table does not consider quality. For example, definitions of unemployment may be inconsistent with international standards (Arezki et al., 2020).

Table 3 MENA Macro Data Availability (as of January 2022)

Economy	GDP Data	GDP Data latest (calendar year)	Industrial Production Index Data	Industrial Production Index latest (calendar year)	Unemployment Data	Unemployment Data latest (calendar year)
Algeria	Quarterly	Q2 2021	Quarterly	Q3 2021	Semi-Annually	May 2019
Bahrain	Quarterly	Q3 2021	n/a	n/a	Only Census	2020
Djibouti	Annually	2018	n/a	n/a	n/a	n/a
Egypt, Arab Rep.	Quarterly	Q3 2021	Monthly	September 2021	Quarterly	Q3 2021
Iran, Islamic Rep.	Quarterly	Q3 2021	Quarterly	Q3 2021	Quarterly	Q3 2021
Iraq	Quarterly	Q2 2021	Quarterly	Q2 2021	Only 2 years	2016
Jordan	Quarterly	Q3 2021	Monthly	October 2021	Quarterly	Q3 2021
Kuwait	Quarterly	Q4 2020	n/a	n/a	Annually	2021

Lebanon	Quarterly	Q4 2019	n/a	n/a	only census	2018-2019
Libya	Annually	2014	n/a	n/a	Labor force survey	2013
Morocco	Quarterly	Q3 2021	Quarterly	Q3 2021	Quarterly	Q3 2021
Oman	Quarterly	Q3 2021	n/a	n/a	Annually	2020
Qatar	Quarterly	Q3 2021	Monthly	November 2021	Quarterly	Q2 2021
Saudi Arabia	Quarterly	Q3 2021	Monthly	November 2021	Quarterly	Q3 2021
Syrian Arab Republic	Annually	2019	n/a	n/a	Annually	2019
Tunisia	Quarterly	Q3 2021	Monthly	September 2021	Quarterly	Q3 2021
United Arab Emirates	Quarterly	Q2 2020	n/a	n/a	Annually	2020
West Bank and Gaza	Quarterly	Q3 2021	Monthly	November 2021	Quarterly	Q3 2021
Yemen, Rep.	Annually	2017	n/a	n/a	only census	2013
MENA Total: 19	Quarterly: 15/19		Monthly or Quarterly: 10/19		Quarterly: 8/19	
Mexico	Quarterly	Q3 2021	Monthly	October 2021	Monthly	November 2021

Source: Gatti et al., (2022). Information obtained from country statistical office websites; government data portal websites; central bank websites and Ministry of Planning/Economy/Finance websites

Notes: N/a means information is not readily available. All data accessed as of January 2022. This table is not exhaustive, it has only surveyed country Statistical Office Websites, Government Data Portal Websites, Central Bank Websites and Ministry of Planning/Economy/Finance websites. Third-party websites were excluded from this survey. The table has been updated and verified by World Bank country economists. Mexico was selected as a comparator because it is an upper-middle-income economy with a well-functioning data ecosystem.

On top of the lack of published monthly unemployment data in the MENA region, there are also challenges regarding the definitions employed. For examples, countries around the world usually follow the ILO's definitions of employment and unemployment, which are consistent with definitions adopted by other developed countries, such as the United States (see Table 4).

Table 4 Definitions of Employment and Unemployment from the U.S. BLS, the French INSEE and the ILO

<i>Definitions</i>	Employment	Unemployment	Source
United States -Bureau of Labor Statistics (BLS)	Someone, aged 16 or over, who has either (1) worked at least 1 hour as a paid employee or (2) in their own business, profession, trade, or farm, or (3) was not temporarily absent from their job, business, or farm, whether or not they were paid for the time off, or (4) worked without pay for a minimum of 15 hours in a business or farm owned by a member of their family.	Someone, aged 16 or over, who (1) does not have a job, (2) has actively looked for one in the past 4 weeks, and (3) is available for work.	https://www.bls.gov/cps/definitions.htm

France-Institut Nationale de la statistique et des études économiques (INSEE)	Individuals who worked for any amount of time, even if only for an hour in the course of the reference week. Only individuals in the working-age population (between 15 and 64 years of age) are considered.	All people aged 15 and older who do not have a job and are looking for one.	https://www.insee.fr/en/metadonnees/definitions
International Labor Organization (ILO)	All those of working age (15 years and over) who, during a short reference period, were engaged in any activity to produce goods or provide services for pay or profit. They comprise employed persons "at work", that is, who worked in a job for at least one hour; and employed persons "not at work" due to temporary absence from a job, or to working-time arrangements (such as shift work, flextime and compensatory leave for overtime).	All those of working age (15 years and over) who were not in employment, carried out activities to seek employment during a specified recent period, and were available to take up employment given a job opportunity.	https://www.ilo.org/ilostat-files/Documents/Statistical%20Glossary.pdf

The definition of official employment and unemployment rates in MENA appear to be inconsistently reported across countries. Many MENA countries do not comply with the ILO's definition of employment or unemployment—or do not provide sufficient information to indicate whether the definitions are followed (see Table 5). Oman, which has one of the highest ODIN scores in the region, does not specify the definition of employment and unemployment. Countries in conflict, such as the Syrian Arab Republic and Yemen, do not provide any information on unemployment. Changes in definitions may entail shifts in the employment and unemployment numbers reported (Arezki et al., 2020).

Table 5 Definitions of Employment and Unemployment

	Follow ILO unemployment definition	Follow ILO employment definition	Age of working population
Morocco	YES	unspecified	15 and above
Algeria	YES	YES	unspecified
Tunisia	YES	unspecified	unspecified
Libya	YES	unspecified	15 and above
Egypt, Arab Rep.	YES	YES	15 and above
Lebanon	YES	unspecified	YES
West Bank and Gaza	YES	unspecified	15 and above
Jordan	YES	unspecified	15 and above
Saudi Arabia	NO	unspecified	15 and above
Oman	Unspecified	unspecified	unspecified
United Arab Emirates	YES	unspecified	YES
Qatar	YES	unspecified	15 and above
Bahrain	YES	YES	unspecified
Kuwait	YES	unspecified	YES
Iran, Islamic Rep.	YES	YES	NO
Djibouti	NO	NO	unspecified
Iraq	unspecified	unspecified	unspecified

Syrian Arab Republic and
Yemen, Rep.

unavailable

unavailable

unavailable

Source: Arezki et al., 2020. Summary based on information from national statistics websites. Information collected in March 2020.

Global shocks such as the pandemic and the war in Ukraine have elevated debt in developing economies both in the region and the world. A complete documentation of debt is essential to plan for potential financial vulnerabilities. Gross public debt is not reported consistently across MENA countries (table 6). While the central government is largely reported, the challenge lies largely with reporting of different categories of public debt such as state and local government, extra budgetary funds and State-Owned Enterprise (SOE) liabilities. Of all the MENA economies analyzed, Tunisia is the only country with the full array of debt reporting.

Table 6 Debt Reporting in MENA Economies

	Algeria	Bahrain	Djibouti	Egypt Arab Rep.	Iran Islamic Rep.	Iraq	Jordan	Kuwait	Lebanon	Libya	Morocco	Oman	Qatar	Saudi Arabia	Syria	Tunisia	United Arab Emirates	West Bank and Gaza	Yemen
Central government	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	
State and local government	n/a		✓	✓	×	✓	✓	n/a			✓	n/a	n/a	n/a		n/a	✓	×	×
Other elements of the general government	×	✓	×	✓				×								✓	×	×	
o/w social security fund	n/a	✓	×	✓			✓	×								✓	×	n/a	
o/w extra budgetary funds	×	✓	×	n/a				×						×		✓	×		×
Guarantees (to other public and private sector, including to SOEs)	×	×	✓	✓		✓	✓			✓	✓	×	×	×		✓	✓	×	n/a
Central bank (borrowed on behalf of the government)	n/a		n/a		✓	✓	×	✓		n/a	✓			n/a		✓		n/a	
Non-guaranteed SOE debt	×	×	×	×	×						×	×	×	×		✓	×	n/a	n/a

Source: Gatti et al., 2021a

Note: Table follows the public debt reporting template of World Bank-IMF's Debt Sustainability Framework. ✓ indicates the country reports the type of debt (for both domestic and external debts); × indicates the country has the type of debt but does not report it; n/a = not applicable and indicates that the country might not have this type of debt; blank cells indicate that World Bank economists do not have information regarding whether the country has the type of debt but does not report it, or that the country does not have the type of debt, or that the debt might be included in total government debt. Debt reporting is as of 2020. Red cells indicate changes between 2019 and 2020

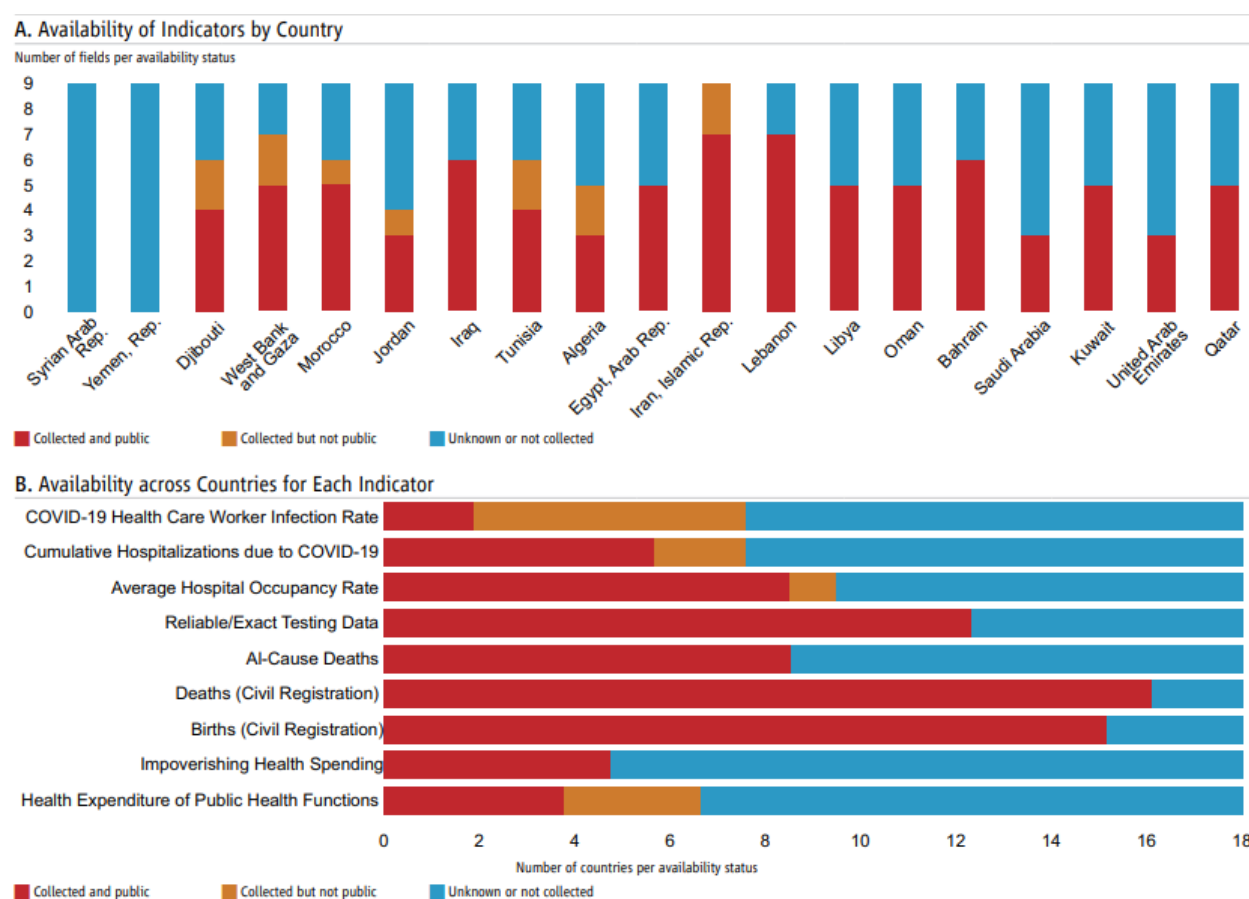
(iii) Public Health Data

The challenges of data overall in the region also translate to data deficiencies in the public health systems. Public health data is an important tool to deal with health crises such as the COVID -19 pandemic. Research by Gatti et. al (2021b) found that health systems were wanting and overconfident before the pandemic. There are also several data challenges with health systems in the region. Figure 8 details the difficulty of

finding and accessing 9 common indicators used to assess pandemic response for 18 countries in the region. Many indicators are either not publicly available or not collected, and in many cases, it was hard to verify which was the case. The indicators cover basic information about key aspects of public health and the conditions facing frontline health workers during the pandemic. These indicators include health system data (such as health worker infection rates and hospital occupancy, spending on public health functions such as contact tracing) and population-level information (such as the total number of births and deaths, and the percentage of households that incurred catastrophic levels of spending in response to an emergency).

The top panel shows the availability of indicators by country. The bottom panel shows availability across countries for each indicator. Essential indicators such as key indicators – such as COVID hospitalizations, COVID testing rates, the infection rate of health care workers and even hospital occupancy rates – were publicly available only about half the time in countries across the region. In many cases, it was not possible to ascertain whether the data was or was not available. This shows that even if data is collected, it is often difficult to find and access.

Figure 8: MENA Public Health Data Availability Assessment



Source: Gatti et al., 2021b

Note: Panel A lists the availability/status of nine fields for each country. Panel B lists the availability status for each field by country. The fields in Panel B are key to a better understanding of the spread, severity and impact of Covid-19 in countries and can be used to conduct comparisons. This assessment aims to determine the status of collection and public availability of these fields per country in the MENA region. Data is either collected and publicly available, collected and not publicly available, or not collected/collection status.

Forging a Social Contract for Data

The note has detailed the data challenge rampant in the region. Bridging the MENA gap requires a multipronged approach to develop sustainable data ecosystems. Where governments lack capacity to generate data, investments are needed to build capacity. Where governments are unwilling to share data, agreements and understanding need to be developed in concert with a clear agenda highlighting the benchmarks of good data ecosystems and the crucial role of data in generating good policies and social harmony. And when important topics are under-researched in the region and require specific data, investments could be made for one-off data collection activities to set up a baseline of knowledge in the region. There is also the promise of digital technologies. If data and digital technologies are combined effectively, there is scope to improve vital services and allow for policies that consider feedback from citizens and businesses. These are key aspects of strengthening the social contract between the state and its citizens.

The 2021 World Development Report on Data for Better Lives calls for a new social contract on data around the world. This is much needed in the MENA region. This is only possible through the existence of an enabling environment. Such an enabling environment has several components. It consists of a mix of legal “safeguards” - ensuring protection of personal data and cybersecurity; and “enablers” that include e-transactions, open data and access to information. Independent, capable, and well-resourced institutions are needed to enforce laws effectively. Part of the process is building the right infrastructure and harmonizing technical standards. The trust engendered could incentivize people and firms to participate in the data economy. To ensure effective and active participation, data literacy and the capacity of civil servants and the broader population will have to be scaled. The evolving social contract also includes the responsible use and reuse of data for development purposes, built on trust that encourages individuals and firms to contribute that essential data. When used in an effective and trustworthy way, data and digital technologies effectively can be crucial enablers for improving service delivery, the accuracy of decision-making, fighting corruption, and increasing transparency.

There are important roles for governments, the private sector and civil society to enable data to create value. Governments play a key role in developing strategies and rules to govern data, with input from civil society and the private sector. Governments also play a key role in producing the data and statistics to inform about developmental progress. The private sector’s role as a producer and user of data is also vital to enable the region to compete within the global digital economy. For that to happen, bottlenecks must be removed to enable access to quality and freely available traditional data sources such as surveys and censuses with open licenses for reuse. Investments in the production of statistics are much needed. There needs to be strong commitment to traditional data (which the WDR 2021 terms “public intent data”) – the kind of data that is used for public good, that improves service delivery and that measures progress specifically for the most vulnerable. This kind of data holds governments accountable and has the potential to empower individuals and groups in fragile and conflict-affected settings where traditional micro data has been sorely lacking. Empowering civil society organizations and academia to research and advocate for the better use of data and the protection of data rights may increase trust in the data economy.

International institutions such as the World Bank can play an important role in developing a social contract around data. One way forward for the World Bank, especially with improving traditional sources of data, would be to create a [Statistics Compact](#) for the region.⁶ The goal of the Statistics Compact is twofold: (i) close the microdata gap (data collection and sharing) and (ii) invest in statistical modernization. The Statistics Compact would support investments in a minimum set of microdata sets comprising of population

⁶ <https://blogs.worldbank.org/opendata/scaling-support-official-statistics-middle-east-and-north-africa-mena-data-compact>

and economic censuses (collected every decade), establishment (formal & informal sector), consumption/income, health and labor force surveys (collected every 3 years) and price surveys (collected monthly) along with their public dissemination at the lowest level of disaggregation allowed for while maintaining data anonymization.⁷ These investments are not limited to supporting data collection, but also include capacity building to assure data quality, harmonization of data collection methods and tools across surveys and countries, data anonymization and dissemination and technical assistance (TA) aimed at improving data integration and interoperability.

The microdata component is complemented with investments in statistical modernization aimed at increasing the capacity to combine survey and non-survey (administrative; transaction, big) data for the production of statistics. In a world that is digitizing rapidly, NSOs need to start to exploit the opportunities for statistics production by utilizing available data from administrative processes (such as population and business registries), transactions (e.g. scanning data from supermarkets; financial transactions; VAT and tax records) or other big data sources (satellite imagery; social media etc.) and combining these with available microdata sets. Utilizing more secondary data sources requires improvements in data governance. Data governance includes establishing the infrastructure and technology, setting up and maintaining the processes and policies, and identifying the individuals and organizations that have both the authority and responsibility for handling and safeguarding specific types of data. At the national level, data governance also covers data interoperability, standards and quality control. A MENA Statistics Compact could support statistical modernization by offering capacity building and technical assistance aimed at reshaping data production processes and improving data governance systems.

⁷ This is just a minimalistic number of microdata sets that would have to be collected regularly. A strong case can also be made to include income surveys (particularly in countries with a large formal sector), perception surveys and wealth surveys.

References

- Arezki, Rabah, Daniel Lederman, Amani Abou Harb, Nelly El-Mallakh, Rachel Yuting Fan, Asif Islam, Ha Nguyen and Marwane Zouaidi. 2020. "How Transparency Can Help the Middle East and North Africa." Middle East and North Africa Economic Update (April), Washington, DC: World Bank. Doi: 10.1596/978-1-4648-1561-4.
- Armand, Alex, Alexander Coutts, Pedro C. Vicente, and Ines Vilela (2020). "Does Information Break the Political Curse? Experimental Evidence from Mozambique." *American Economic Review* 110 (1): 3432-53
- Bac, Mehmet (2001). "Corruption, Connections and Transparency: Does a Better Screen Imply a Better Scene"? *Public Choice* 107:87-96.
- Binswanger, Johannes and Manuel Oechslin (2015). "Disagreement and Learning about Reforms." *The Economic Journal* 125:853-886.
- Brada, Josef C., Zdenek Drabek, Jose A. Mendez, M. Fabricio Perez (2019). "National Levels of Corruption and Foreign Direct Investment." *Journal of Comparative Economics* 47(1): 31-49.
- Brunetti, Aymo and Beatrice Weder (2003). "A Free Press is Bad News for Corruption." *Journal of Public Economics* 87(7-8):1801-1824.
- Brunetti, Aymo, Gregory Kisunko, and Beatrice Weder (1998). "Credibility Rules and Economic Growth: Evidence from a Worldwide Survey of the Private Sector." *World Bank Economic Review* 12(3):353-384.
- Campos, J. Edgardo, Donald Lien, and Sanjay Pradhan (1999). "The Impact of Corruption on Investment: Predictability Matters." *World Development* 27(6):1059-1067.
- Choi, Sangyup and Yuko Hashimoto (2018) "Does Transparency Pay? Evidence from IMF Data Transparency Policy Reforms and Emerging Market Sovereign Bond Spreads." *Journal of International Money and Finance* 88:171-190.
- Cordis, Adriana S. and Patrick L. Warren (2014). "Sunshine as Disinfectant: The Effect of State Freedom of Information Act laws on Public Corruption." *Journal of Public Economics* 115: 18-36.
- De Soto, Hernando (1989). *The Other Path: The Invisible Revolution in the Third World*, New York: Harper.
- Djankov, Simeon, Edward Glaeser, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2003). "The New Comparative Economics." *Journal of Comparative Economics* 31:595-619.
- Ehlich, Isaac and Francis T. Lui (1999). "Bureaucratic Corruption and Endogenous Economic Growth." *Journal of Political Economy* 107(6): S270-S293.
- Ekhtor-Mobayode, Uche Eseosa, and Johannes Hoogeveen. 2021. Microdata Collection and Openness in the Middle East and North Africa (MENA). August 24. <https://doi.org/10.5281/zenodo.5241685>
- Gatti, Roberta, Daniel Lederman, Rachel Yuting Fan, Arian Hatefi, Ha Nguyen, Anja Sautmann, Joseph Martin Sax, and Christina A. Wood. 2021b. Overconfident: How Economic and Health Fault Lines Left the Middle East and North Africa Ill-Prepared to Face COVID-19. MENA Economic Update. Washington, DC: World Bank. October.

Gatti, Roberta, Daniel Lederman, Ha M. Nguyen, Sultan Abdulaziz Alturki, Rachel Yuting Fan, Asif M. Islam, and Claudio J. Rojas. 2021a. "Living with Debt: How Institutions Can Chart a Path to Recovery for the Middle East and North Africa." Middle East and North Africa Economic Update (April), Washington, DC: World Bank.

Gatti, Roberta; Lederman, Daniel; Islam, Asif M.; Wood, Christina A.; Fan, Rachel Yuting; Lotfi, Rana; Mousa, Mennatallah Emam; Nguyen, Ha. 2022b. "Reality Check: Forecasting Growth in the Middle East and North Africa in Times of Uncertainty" Middle East and North Africa Economic Update (April), Washington, DC: World Bank. Doi: 10.1596/978-1-4648-1865-3

Grajalez, Carlos Gómez, Eileen Magnello, Robert Woods, and Julian Champkin (2013). "Great Moments in Statistics." *Significance* 10 (6): 21–28.

Hollier, James, B. Peter Rosendorff, and James Raymond Vreeland (2011). "Democracy and Transparency." *Journal of Politics* 73 (4): 1191–205.

Islam, Roumeen (2006). "Does More Transparency Go along with Better Governance?" *Economics and Politics* 18 (2): 121–67.

Islam, Asif Mohammed and Daniel Lederman. 2020. "Data Transparency and Long-Run Growth." Policy Research Working Paper No. 9493. Washington, DC: World Bank.

Khemani, Stuti, Ernesto Dal Bo, Claudio Ferraz, Frederico S. Finan, Johnson Stephenson, LouiseCorinne, Adesinaola M. Odugbemi, Dikshya Thapa and Scott D. Abrahams (2016). "Making politics work for development: Harnessing Transparency and Citizen Engagement." Policy Research Report 106337. World Bank, Washington, DC.

Kosec, Katrina and Leonard Wantchekon (2020). "Can Information Improve Rural Governance and Service Delivery." *World Development* 104376. <https://doi.org/10.1016/j.worlddev.2018.07.017>

Lederman, Daniel, Norman V. Loayza, and Rodrigo Soares (2005). "Accountability and Corruption: Political Institutions Matter." *Economics and Politics* 17(1): 1-35

Mauro, Paolo (1995). "Corruption and Growth." *Quarterly Journal of Economics* 110(3):681-712.

Mauro, Paolo (1998). "Corruption and the Composition of Government Expenditure." *Journal of Public Economics* 69:263-279.

McMillan, John, and Pablo Zoido (2004). "How to Subvert Democracy: Montesinos in Peru." *Journal of Economic Perspectives* 18 (4): 69–92.

Prat, Andrea (2005). "The Wrong Kind of Transparency." *American Economic Review* 95(3): 862-877.

Rodrik, Dani (2010). "Diagnostics before Prescription." *Journal of Economic Perspectives* 24(3):33-44

Shleifer, Andrei, and Robert W. Vishny (1993). "Corruption." *Quarterly Journal of Economics* 108 (3): 599-617.

Varvarigos, Dimitrios and Panagiotis Arsenis (2015). "Corruption, Fertility, and Human Capital." *Journal of Economic Behavior and Organization* 109:145-162.

Wei, Shang-Jin (2000). "How Taxing is Corruption on International Investors?" *Review of Economics and Statistics* 82(1), 1–11.

Williams, Andrew (2009). “On the Release of Information by Governments: Causes and Consequences.” *Journal of Development Economics* 89: 124–38.

World Bank (2021). *World Development Report 2021: Data for Better Lives*, forthcoming. Washington, DC: World Bank.

Annex: Detailed Definitions of Data System Indicators

Table A1: SCI, SPI and ODIN Summary Table

	Statistical Capacity Indicator (SCI)	Statistical Performance Indicator (SPI)	Open Data Inventory (ODIN)
Aim and Scope	3 main pillars: (i) Methodology (ii) Data Source (iii) Periodicity and Timeliness	5 main pillars: (i) Data use (ii) Data services (iii) Data products (iv) Data sources (v) Data infrastructure	Two main pillars: (i) Coverage (ii) Openness These two categories are explored for three groups of statistics (a) Social statistics (b) Economic and Financial Statistics and (c) Environmental Statistics.
Coverage (years and countries)	Developing economies (2004-2020)	Developing and developed economies (2016-2019)	Developing and developed economies (2017-2018, 2020)
Notes	To be discontinued	Incomplete (not all elements collected yet)	Only data in NSOs or other official country website considered. No microdata. No standards
Source	World Bank	World Bank (WDR 2021)	Open Data Network

Table A2: Statistical Capacity Indicator (SCI) Components

Dimension	Definition
Statistical Methodology	The statistical methodology dimension measures a country's ability to adhere to internationally recommended standards and methods. This aspect is captured by assessing guidelines and procedures used to compile macroeconomic statistics and social data reporting and estimation practices.
Source Data	The source data dimension reflects whether a country conducts micro data collection activity in line with internationally recommended periodicity, and whether data from administrative systems are available and reliable for statistical estimation purposes.

Periodicity and Timeliness	The periodicity and timeliness dimension measures the availability and periodicity of key socioeconomic indicators, of which nine are Millennium Development Goals (MDG) indicators.
----------------------------	--

Source: <https://datatopics.worldbank.org/statisticalcapacity/scidashboard.aspx>

Table A3: Statistical Performance Indicator (SPI) Components

Dimension	Definition
Data use	A successful data system produces data that are highly used by a wide variety of users (given the lack of data, this indicator is just limited to international institutions – poverty World Bank, WHO, ILO etc)
Data services	A successful data system has highly valued and well-used data services that connects users and producers. This basically includes data release (the IMF SDSS subscription) online access (ODIN openness score), and NADA micro data catalogue
Data products	A successful data system produces high quality statistical indicators that track the SDGs (this is largely availability of SDG data)
Data sources	A successful data system draws on all types of data relevant for indicators (same as SCI source data sub-score but also includes labor force and establishment census and surveys. Admin data is lacking – just civil registration)
Data Infrastructure	A successful data system has good hard (legislation) and soft (skills and partnerships) data infrastructure, as well as the financial resources to deliver. Note that as its stands, skills and partnerships are not included. Only legislation is included, and thus similar to SCI methodology indicator with some inclusions and exclusions.

Source: <https://www.worldbank.org/en/programs/statistical-performance-indicators>

Table A4: Open Data Inventory (ODIN) Components

Dimension	Definition
-----------	------------

Openness	Openness scores are based on whether data can be (i) Machine readable (ii) Non-proprietary formats (iii) Download options (iv) Metadata available and (v) Terms of use These are explored for (a) Social statistics (b) Economic and Financial Statistics and (c) Environmental Statistics.
Coverage	Coverage is based on the availability of key indicators over time and geographical disaggregation. Time dimensions include (i) Data available last 5 years and (ii) Data available last 10 years. Geographical dimensions include (i) First administrative level and (ii) Second administrative level These are explored for (a) Social statistics (b) Economic and Financial Statistics and (c) Environmental Statistics.

Source: <https://opendatawatch.com/publications/open-data-inventory/>

Table A5: Comparison between SPI and SCI

SCI	SPI	SPI Details and Additions over SCI
Methodology	Data Infrastructure (Pillar 5)	In accordance to SPI pillar 5, a successful data system has good hard (legislation) and soft (skills and partnerships) data infrastructure, as well as the financial resources to deliver. Note that as its stands, skills and partnerships are not included. Only legislation is included, and thus similar to SCI methodology indicator with some inclusions and exclusions. <i>Additions over SCI</i> Classification of national industry Classification of household consumption Classification of status of employment Central government accounting status Compilation of monetary and financial statistics Business process <i>Removed from SPI but in SCI</i> Industrial production index Import/Export Prices GDP growth absorbed into periodicity of SDGs
Source Data	Data sources (Pillar 4)	According to SPI pillar 4, A successful data system draws on all types of data relevant for indicators. <i>Additions over SCI:</i> Establishment surveys Establishment census Labor force surveys

Periodicity and Timeliness	Data products (Pillar 3)	A successful data system produces high quality statistical indicators that track the SDGs (this is largely availability of SDG data) Not in SCI although there is overlap. SCI captures periodicity of MDGs, while SPI captured SDG periodicity.
Not in SCI	Data services (Pillar 2)	As indicated by SPI pillar 2, a successful data system has highly valued and well-used data services that connects users and producers. This basically includes (a) data release (the IMF SDSS subscription) (b) online access (ODIN openness score), and (c) NADA micro data catalogue Only the IMF SDDS subscription is included in SCI, but there is a difference. Here the variable captures IMF SDDS/e-GDDS subscription as opposed to SDDS alone. Online access obtained from ODIN. NADA cataloguing of data in NSOs also included. Both are an addition over the SCI.
	Data use (Pillar 1)	According to SPI pillar 1, a successful data system produces data that are highly used by a wide variety of users (given the lack of data, this indicator is just limited to international institutions – poverty World Bank, WHO, ILO etc). More importantly these data should meet certain standards to be comparable This includes whether comparable data that meet certain standards are produced over time for the following indicators. - Poverty estimates (not in SCI) - Child mortality (not in SCI) - Debt reporting (in SCI) - Drinking water (somewhat in SCI) - Labor force participation (not in SCI)