UNITED NATIONS E

ECONOMIC AND SOCIAL COUNCIL

Distr.
LIMITED
E/ESCWA/C.1/2024/5
6 August 2024
ORIGINAL: ENGLISH

Economic and Social Commission for Western Asia (ESCWA)

Statistical Committee
Sixteenth session
Manama, 22-23 October 2024

Item 6 of the provisional agenda

Technological developments in areas that influence the modernization of official statistics

Summary

The present document provides a snapshot of technological developments in the areas that influence the modernization of official statistics in the member States of the Economic and Social Commission for Western Asia (ESCWA). It presents an overview of the work of the High-Level Group for the Modernization of Official Statistics of the Bureau of the Conference of European Statisticians, and the United Nations Committee of Experts on Big Data and Data Science for Official Statistics. The document showcases the application of those technological developments in the Arab countries, including cloud computing for official statistics, data science applications, artificial intelligence and machine learning in official statistics, use cases in price statistics, transport and trade statistics, standards for statistical data and metadata exchange (SDMX), open data and geospatial applications.

The document also touches on partnerships and means for pooling resources and expertise for modernization, and provides examples of successful collaborations from the region. It presents a set of recommendations for the Statistical Committee on the way forward.

2400488E

Contents

Introduction

1. Official statistics constitute a key element of national data infrastructure, particularly when they are supported by public trust, good data quality and independence from political influence. They support the population's well-being and promote economic growth through data-driven public and private decision-making. Innovation and digital technologies offer official statistics the opportunity to be more relevant, interoperable, granular and timely.[1]

2. Data has been gaining importance in recent years. In 2021, the number of people connected to the Internet exceeded 4.6 billion, compared to only 2.3 million in 1990. In the Arab region, the percentage of individuals using the Internet increased from 47.2 per cent in 2017 to 54.6 per cent at the end of 2019.

3. The Fourth Industrial Revolution is enhancing the connection of "things" to the Internet and to people. All types of devices have become sources of data. The total number of "things" connected to the Internet is expected to reach 30.9 billion by 2025. As a result, data production has reached an unprecedented volume and speed. By 2025, the volume of data is expected to reach 175 zettabytes. These data flows have led to the creation of new infrastructure, new businesses, new monopolies, new economies and new business models for the public sector.

4. The realization of the Sustainable Development Goals (SDGs) necessitates robust official statistics to track progress and inform policy decisions. In this era of increasing global interconnectedness, vast amounts of data are continuously generated from diverse sources. This massive flow of data represents a great opportunity for statisticians and policymakers. New technologies offer opportunities for more relevant, granular and timely official statistics.

5. Decision 53/124 of the United Nations Statistical Commission is evidence of this transformative potential. This decision supports the integration of big data and data science within national statistical offices (NSOs). It fosters collaboration between the United Nations Committee of Experts on Big Data and Data Science for Official Statistics (UNCEBD) and the geospatial community, which promotes the creation of a network of data science leaders within NSOs, with the goal of modernizing statistical production.

I. Global process on modernization of official statistics

A. The High-Level Group for the Modernization of Official Statistics

6. The present document highlights the technological developments in the areas that influence the modernization of statistical production, as provided by the High-Level Group for the Modernization of Official Statistics (HLG-MOS) of the Bureau of the Conference of European Statisticians.[2] These developments include cloud computing for official statistics, data science tools, and artificial intelligence (AI) and machine learning (ML) in official statistics, with examples of ESCWA applications in areas such as price data collection and case studies for the use of the Automatic Identification System (AIS) in shipping and trade. Other important elements for modernization presented in the paper are the standards for data exchange, open data and geospatial applications.

[1] Groshen, Erica, The future of official statistics, 2021.

[2] United Nations Economic Commission for Europe (UNECE), HLG-MOS workshop on the modernization of official statistics 2023: report of the expert meeting, 2024.

B. The United Nations Committee of Experts on Big Data and Data Science for Official Statistics

7. The present document also highlights the work of the UNCEBD, established in 2014 by the Statistical Commission in Decision 45/110. The Committee undertook a 10-year review of big data and data science for official statistics, and presented it to the fifty-fifth session of the Statistical Commission in 2024.[3]

8. All NSOs and international organizations under the Statistical Commission were invited to participate in an online survey on the 10-year review of big data and data science for official statistics. A total of 84 NSOs, 55 of which were from developing countries, and 15 international organizations completed the survey. The overall results revealed the gradual emergence of data science in the work of statistical offices over the course of the last 10 years, as well as the increasing willingness of statistical offices to use data from the private sector. The survey questions were structured according to the main components of the Generic Activity Model of Statistical Organizations (GAMSO), namely: strategic vision; legislation; institutional arrangement and partnerships; data sources, methodology and quality assurance; communication and stakeholder consultations; human resources; and information technology (IT) management. For each of those areas, questions were included regarding big data, data science and modernization in general.

9. In the Arab region, there were participants from nine different organizations and countries, holding different positions such as director general, analyst and chairman of the Board of Directors. Their responses revealed a strong engagement from organizations in strategic innovation, legislative support for data access, the utilization of diverse data sources, robust IT and human resource strategies, and active participation in UNCEBD activities. This engagement underscores the commitment of these organizations to modernize and innovate their statistical processes and collaborate internationally to enhance their capabilities.

10. The main results from the responding Arab countries on the different questions are reported in the annex to the present document, and the main recommendations are provided below:

(a) Strengthen legislative support and encourage countries that lack legislation to facilitate access to privately held data to develop and implement such laws;

(b) Support organizations in the planning stages of their innovation strategies by providing frameworks and best practices;

(c) Expand training and capacity-building programmes focused on data science, data engineering and the use of alternative data sources;

(d) Foster stakeholder collaboration and promote partnerships between statistical organizations and various stakeholder communities, including the private sector and civil society;

(e) Improve IT infrastructure upgrades, particularly for organizations with limited resources;

(f) Encourage greater participation in UNCEBD conferences and task teams by highlighting the benefits and providing financial or logistical support, where needed;

(g) Support organizations in exploring and integrating alternative data sources such as satellite data, web scraping and mobile phone data;

(h) Assist countries in developing comprehensive national data strategies that encompass data governance, sharing and quality assurance;

(i) Enhance awareness and communication through targeted communication strategies and outreach programmes.

[3] E/CN.3/2024/6. Item 3 (d).

11. To respond to some of the recommendations, the next chapter presents the elements needed for the modernization of national statistical systems using technology in official statistics.

II. Technological developments that influence the modernization of statistical production

A. Cloud computing for official statistics

12. Cloud computing has emerged as a transformative force in various fields, including statistics, by offering scalable and accessible computing resources over the Internet. In disaster-prone and conflict-affected countries, where traditional infrastructure may be lacking, compromised or damaged during a disaster, leveraging cloud computing can significantly enhance statistical analysis, aiding informed decision-making and policy formulation. This section explores the concept of cloud computing, its modes of usage and how it can be effectively utilized in statistical endeavours, particularly in conflict-affected regions.

Concept of cloud computing

13. Cloud computing refers to the delivery of computing services, including storage, databases, servers, networking, software and analytics, over the Internet (the cloud) of many Governments for free or on a pay-for-use basis when it involves the private sector cloud. Cloud computing eliminates the need for organizations to invest in and maintain costly infrastructure, offering flexibility, scalability and cost-effectiveness. It operates on three main service models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS), each providing varying levels of control and management to users.[4]

- **IaaS:** Users can rent virtualized computing resources, such as virtual machines, storage and networking, from cloud service providers. This mode offers the highest level of control, allowing users to deploy and manage their own applications and software.
- **PaaS:** PaaS provides a platform that allows users to develop, run and manage applications without dealing with the underlying infrastructure. It offers tools and services for application development, database management and business analytics.
- **SaaS:** SaaS delivers software applications over the Internet on a subscription basis. Users can access and use applications hosted on cloud infrastructure without worrying about installation, maintenance or updates.

14. Cloud computing revolutionizes statistical analysis by providing:

(a) Scalability: Cloud resources can be scaled up or down based on demand, enabling statisticians to handle large datasets and complex computations efficiently;

(b) Accessibility: Statistical tools and software hosted on the cloud are accessible from anywhere with an Internet connection, facilitating collaboration among researchers and analysts globally;

(c) Cost-effectiveness: Cloud computing eliminates the need for upfront infrastructure investments, allowing organizations to pay only for the resources they use and reducing operational costs;

(d) Data security: Cloud service providers implement robust security measures to protect data, ensuring confidentiality, integrity and availability;

(e) Advanced analytics: Cloud platforms offer a wide range of analytics tools and services, including ML, data visualization and predictive modelling, empowering statisticians to derive meaningful insights from data.

[4] Ishtiaq, Mariam and others, Edge computing in IoT: a 6G perspective, 2022.

15. In disaster-prone and conflict-affected countries where conventional infrastructure may be inadequate or disrupted, cloud computing offers a lifeline for statistical analysis:

(a) Infrastructure resilience: Cloud infrastructure is inherently resilient and distributed, ensuring continuity of operations even in volatile environments;

(b) Remote access: Cloud-based statistical tools enable researchers, government employees and analysts to work remotely, overcoming geographical barriers and security concerns prevalent in conflict zones;

(c) Collaboration and capacity-building: Cloud platforms facilitate collaboration and knowledge sharing among government entities, and local, national and international stakeholders, fostering capacity-building and skill development in statistical analysis;

(d) Disaster recovery: Cloud providers offer robust disaster recovery mechanisms, ensuring data backups and continuity of operations in the event of natural disasters or conflicts.

16. An example of a country that recently introduced cloud computing is the Sudan. ESCWA provided technical assistance to central bureaus of statistics (CBS) in the Sudan in January 2024 on cloud computing and recommended to enact robust legislation to facilitate the integration of cloud computing within the country. This resulted in crafting a comprehensive national strategy dedicated to cloud computing, conducting a meticulous assessment of the current state of CBS data according to a nationally endorsed framework and establishing a coordinated governance mechanism.

B. Data science, artificial intelligence and machine learning in official statistics: specific applications

17. Data science is a broad field of study about data systems and processes aimed at maintaining datasets with tools, applications, principles and algorithms to derive insights from data.

18. AI, as per the Eurostat definition, refers to systems that use technologies such as text mining, computer vision, speech recognition, natural language generation, ML and deep learning to gather and/or use data to predict, recommend or decide the best action to achieve specific goals, with varying levels of autonomy. AI is a powerful tool that can accelerate progress towards the 2030 Agenda for Sustainable Development, improving economic productivity and driving innovation.

19. ML is a "field of study that gives computer the ability to learn without explicitly being programmed".[5] ML algorithms can be grouped into two different types: (1) supervised ML where a dataset has known target values, and machines are instructed to learn the relationships between features and the target; and (2) unsupervised ML where there is no target in the dataset, and the algorithm needs to figure out any patterns on its own. ML can be used to increase the relevance and quality of official statistics in an environment of growing demands for trusted information, rapidly developing and accessible technologies, and numerous competitors.[6]

20. Both AI and ML are fundamentally transforming statistical organizations along the General Statistical Business Process Model (GSBPM) framework, which regulates the generation of credible statistics through higher efficiency of automated repetitive processes, extraction of valuable insights from data and enhancement of analytical capabilities.

21. The integration of AI and ML in official statistics is still a work in progress to adhere to formal statistical standards and mitigate biases in technology, in both data and algorithms. NSOs need proper funding, human resources and technical skills for this purpose.

[5] Brown, Sara, Machine learning, explained, 2021.

[6] UNECE, Machine learning for official statistics, 2021.

22. Pilot studies have been conducted by several countries using the coding and classification theme. Examples include: the National Institute of Statistics and Geography (INEGI) in Mexico on classifying occupation and economic activity; Statistics Canada on classifying industry and occupation; and Statistics Norway on classifying standard industrial code classification. The United Kingdom Office of National Statistics recently used large language models (LLMs) to classify deidentified free text from labour market surveys to the Standard Industrial Classification (SIC), the Standard Occupational Classification (SOC) and the Classification of Individual Consumption According to Purpose (COICOP). LLMs seemed to distinguish spelling mistakes, sentiment and other nuances, better than traditional ML approaches.

23. The new tool of the United Kingdom Office of National Statistics, "ClassifAI", was developed as an experimental text classification pipeline using Retrieval-Augmented Generation (RAG), with the aim of improving on existing approaches in combining ML with manual and rules-based techniques in classifying industry or occupational classifications in terms of flexibility and accuracy. The code base was released in July 2024 for a proof-of-concept pipeline that tests this new approach on GitHub. Further assessments and validation are being made available for research purposes.

24. Initial results have been promising, demonstrating a marginal improvement in agreement with manual human classification compared with existing approaches, at different levels of classification hierarchies.[7]

C. Price data collection and case studies

25. Traditionally, gathering price data has been a complex and labour-intensive task, requiring significant manual effort and resources. However, advancements in data science and technology have created a growing imperative to modernize and simplify this process. ESCWA has responded to this need by utilizing both scanner data and the Automatic Price Collection Tool (APCT) to streamline the extraction of price data. Scanner data provide a wealth of detailed and real-time information, further enhancing the precision of price statistics. The APCT stands out as a groundbreaking innovation, designed to harness technology efficiently, and aiming to improve the accuracy and efficiency of price data collection, thereby enhancing the reliability of price indices. These tools complement each other to ensure reliable statistics and scalable solutions, representing a state-of-the-art approach to addressing the challenges of traditional manual data collection techniques.

Objective

26. The APCT utilizes sophisticated data scraping techniques and advanced programming to automate the extraction of price information from online sources. This method not only ensures comprehensive coverage of online retail platforms but also enhances efficiency and accuracy by overcoming the limitations of manual data collection. The APCT adaptability to different economic contexts and its integration capabilities with existing systems is what defines it as an innovative and modern tool, capable of meeting diverse data collection needs in today's dynamic economic landscape.

27. Scanner data, on the other hand, are derived from the digital transactions recorded at the cash registers of retail shops. Scanner data systems leverage cloud computing, data science and automated data processing to extract transactional price data from physical retail stores in real time. These systems interface directly with retail outlets, capturing detailed pricing information across a wide range of products and locations. Automating the extraction process not only reduces the labour-intensive nature of traditional data collection methods but also enhances data accuracy, coverage and timeliness. The integration between NSOs and retail environments ensures that price indices reflect current market dynamics accurately, supporting robust economic analysis and policymaking decisions.

[7] Data Science Campus, ClassifAI – Exploring the use of Large Language Models (LLMs) to assign free text to commonly used classifications, 2024.

28. Both the APCT and scanner data are pivotal in modernizing the collection of price data. The primary objective of the APCT is to automate the collection of online price data, particularly focusing on items listed in the national Consumer Price Index (CPI) basket. The tool is expanding its functionalities to support the collection and construction of the International Comparison Programme (ICP) list. By automating online data collection processes, the APCT aims to enhance efficiency, accuracy and timeliness in compiling essential economic indicators. On the other hand, scanner data automate the extraction of data from physical stores, ensuring that they are complementary in enhancing the efficiency of price statistics and indices. The integration of the APCT alongside scanner data ensures the reliability of price indices by providing detailed, real-time information from both online and physical retail environments. Together, these tools significantly improve the reliability and scalability of economic data collection, ensuring the production of high-quality statistics.

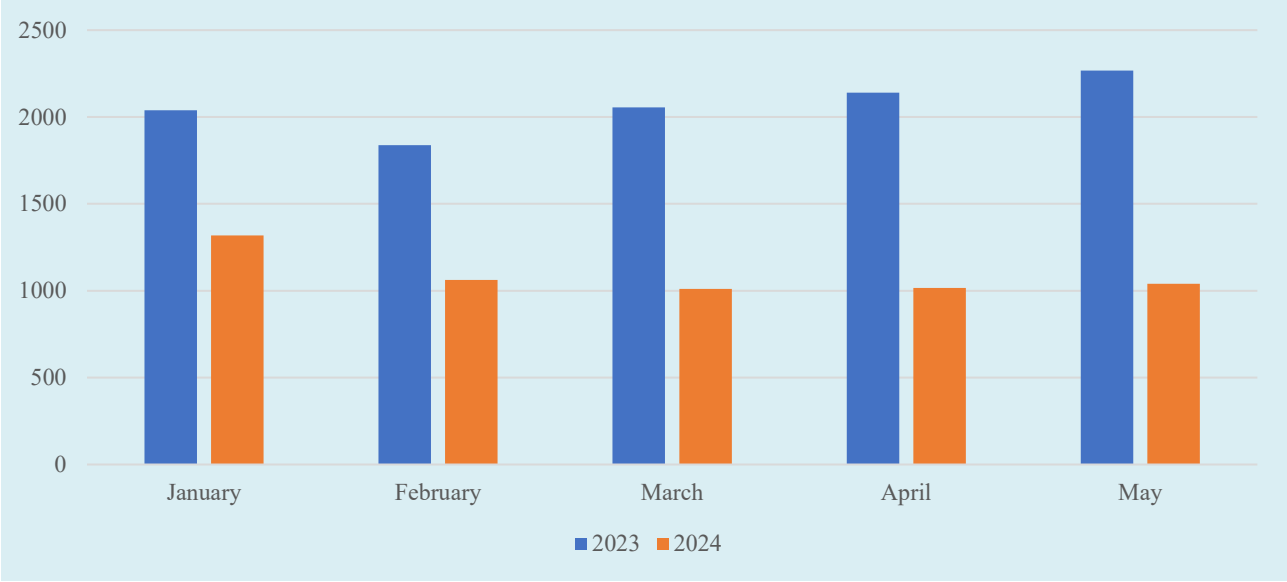
D. Automatic identification system for shipping and maritime transportation

29. The AIS is a signalling (sensor) system used for capturing information about a ship's location, speed and status. It enables near real-time analysis of maritime activity and traffic patterns for ports and passageways, in addition to how they are affected by local and global events or other phenomena. Given the high volume, high frequency and global coverage of the AIS, timely insights can be extracted for any area of interest across the globe, providing valuable insights into maritime and vessel behaviour, trade flows and shipping dynamics. A framework and data processing methods to derive statistical indicators from AIS data were developed by the Asian Development Bank.[8]

30. The data are made available by the United Nations Global Platform infrastructure, where AIS data can be accessed, processed and analysed to produce the indicators used in this report.[9]

31. Utilizing AIS data, ESCWA conducted a case study to estimate the impact of the attacks in the Red Sea on vessel movements in the Suez Canal. The study involved calculating the number of ships passing through the canal each month from January to May 2024, and comparing these figures to the corresponding months in the previous year.

Figure 1. Cargo and tanker distribution in the Suez Canal on a monthly basis, 2023 and 2024



[8] Asian Development Bank, Methodological Framework for Unlocking Maritime Insights Using Automatic Identification System Data, 2023.

[9] UN Big Data, AIS data.

32. The findings revealed a significant decline in the number of vessels traversing the Suez Canal. This reduction led to an escalation in shipping costs as vessels were compelled to navigate longer routes. It also resulted in a loss of revenue for Egypt, as fewer vessels were paying the customary crossing fee to traverse the canal.

E. Standards for data exchange: a modernization tool for statistics

33. The statistical data and metadata exchange (SDMX) initiative aims to modernize the statistical process by streamlining the statistics workflow, including the collection, validation, computation, aggregation, standardization, dissemination and exchange of data such as education and labour statistics, national accounts and SDG data. It also enables working with microdata, administrative data, geospatial data and non-traditional data sources, including big data. Moreover, SDMX enables interoperable implementations within and between systems concerned with the exchange, reporting and dissemination of statistical data and related metadata and improves the coordination of statistical activities across the national statistical system and among development partners.

1. *Typical uses of SDMX*

34. SDMX could be used in the following domains:

(a) Governance and management of data and metadata: Efficient data exchange and dissemination of statistical products rely heavily on internal inputs. SDMX offers various tools to maintain and manage these input files such as mapping files, code lists files, etc.;

(b) Dissemination of data: SDMX facilitates data and metadata reporting. It automates the loading of data and reference metadata into databases, automates the creation of database tables, enables a database to be compatible with SDMX web services, and implements data and metadata warehousing;

(c) Exchange of data: SDMX provides a standardized way of organizing and exchanging data, enabling interoperable implementations within and between systems concerned with the exchange, reporting, and dissemination of statistical data and related metadata.

2. *ESCWA activities on SDMX*

35. ESCWA SDMX Converter for SDGs: ESCWA established an automated system for SDG data exchange with its member States to facilitate the timely exchange and dissemination of quality national SDG data to the Arab SDG Monitor national reporting platforms. By the end of 2023, 16 member States from the Arab region reported their SDG indicators using the SDMX standard to update the national reporting platform, thus reducing the reporting burden and enabling the dissemination of high-quality and standardized SDG data.

36. Building the capacities of Arab countries: Demand for capacity-building has grown substantially in recent years. In response to member States' request during the fifteenth session of the Statistical Committee, ESCWA collaborated with the United Nations Statistics Division (UNSD) to implement seven workshops on the use of SDMX for reporting on SDGs. These workshops covered basic to advanced topics since 2020, and were held either in-person, virtually or in a hybrid mode.

37. Examples of SDMX use cases in the Arab region:

(a) United Arab Emirates (Federal Competitiveness and Statistics Centre): With the aim of fostering the production and dissemination of high-quality data, the United Arab Emirates implemented the .Stat Suite, which is an SDMX native standard-based, open-source platform for the efficient production and dissemination of statistical data;

(b) GCC-Stat: With the objective of automating the collection, storage and dissemination of data, and as part of the MARASA system, GCC-Stat adopted the use of SDMX for data preparation. This involves building SDMX artefacts to extract, store, validate and disseminate the data in SDMX structured formats;

(c) Jordan (Department of Statistics): The implementation of SDMX at the Department of Statistics will facilitate data exchange and integration in the newly created data centre for Jordan.

38. SDMX is a business choice that is making significant progress in countries that decided to integrate it in their statistical workflow, allowing for the improvement of data exchange quality through standardization, automation, validation, data sharing and facilitation of the transferability of skills. Member States are thus invited to adopt SDMX and benefit from its free-of-charge tools to modernize their statistical systems.

F. Open data: overview and socioeconomic impact

39. Governments possess and continue to generate vast amounts of data, some of which are stored in paper formats or legacy systems (and therefore need to be digitized), and others are available digitally through e-government services. All of this is a potential source of wealth and public value creation, provided that data collection, storage, dissemination and sharing methods are modernized to make data easily accessible and reusable. However, these efforts face significant challenges, such as lack of skills and expertise needed for this purpose.

40. Open data refer to data that are freely available for anyone to use, reuse and redistribute without restrictions. They encompass a wide range of information, including government data, scientific research and other publicly accessible datasets. The concept of open data is rooted in the principles of transparency, accountability and innovation. The movement gained significant momentum with the advent of the digital age and the proliferation of the Internet, which made the distribution and accessibility of data easier than ever before.

41. In 2018, the Statistical Commission, in its decision 49/105, created a subgroup on open data under the Friends of the Chair group on the Fundamental Principles of Official Statistics to work on the principles, guidance and support for the implementation of open data in countries. The subgroup presented its work at the fiftieth session of the Commission, which included a mapping of open data principles to the Fundamental Principles, a synthesis of existing open data work and an analysis of other matters relating to the implementation of open data in official statistics.

42. At its fiftieth session, the Statistical Commission, in its decision 50/105, agreed on the establishment of an open data working group to continue the work on open data, including the development of guidance for the assessment and practical application of open data in the context of official statistics, and requested the group to present its work at the fifty-first session of the Commission and every two years thereafter.[10]

1. *Socioeconomic impact*

43. Globally, the socioeconomic impact of open data is profound. They facilitate innovation and entrepreneurship, leading to the creation of new industries and job opportunities. For instance, open data in the environmental sector can help develop new technologies for monitoring and combating climate change. Open data can also support education and research by providing students and academics with valuable resources, fostering a knowledge-based economy.

44. In the Arab region, open data have the potential to address specific socioeconomic challenges and unlock significant benefits. Despite varying levels of adoption across countries, there is a growing recognition of the value of open data for the following reasons:

[10] E/2018/24-E/CN.3/2018/37; E/2019/24-E/CN.3/2019/34.

- **Economic growth:** Open data can stimulate economic development by enabling startups and small businesses to leverage data for creating innovative solutions. For example, in the United Arab Emirates, open government data have been used to develop applications that enhance tourism and service delivery, boosting economic activity.
- **Transparency and governance:** In many Arab countries, open data initiatives are enhancing government transparency and accountability. When Governments disclose their budget, spending and financial data, and administrative data, and make them available and accessible for the public, citizens can better understand and engage with their Governments, fostering a more participatory political environment. Some examples from Arab countries are shown in the table.
- **Social impact:** Open data can also address social issues such as education, health care and urban development. For instance, open education data can improve the quality and accessibility of education by highlighting areas that need attention. Similarly, open health-care data can enhance public health responses and policymaking.
- **Crisis management:** In regions prone to conflicts and natural disasters, open data can play a critical role in crisis management and recovery. By providing real-time information, open data initiatives can improve the efficiency of humanitarian aid and disaster response efforts.

45. There are risks, however, with open data that need to be addressed. The issues of data location, jurisdictional control and data sovereignty have emerged with the growth of cloud service deployments and are key fundamental policy considerations for both Governments and commercial operators.[11]

Examples of open data initiatives in Arab countries

Country	Initiative	Short description	More information
Oman	Bayanat Analytical Portal	The portal was developed to support the communication and exchange of financial and non-financial data and information between listed and licensed companies. It relies on data collected according to the Extensible Business Reporting Language (XBRL) - an open standard for digital business reporting. Companies submit their financial results, quarterly and annually, and the results get verified by accredited auditors. The system is integrated with various authorities, including Muscat Stock Exchange (MSX) and Muscat Clearing & Depository (MCD); and a single sign-on is implemented for all reporting companies. The portal ensures the availability of accurate and reliable financial information to all relevant stakeholders with the support of analysis tools and dashboards (Microsoft Power BI).	https://opengov.unescwa.org/index.php/node/1253 https://bi.bayanat.gov.om/
Tunisia	National Register of Government Data	The website aggregates the list of government data inventoried by public institutions and made available and accessible for users. It allows users to consult the data and metadata, and offers participation mechanisms to contribute to the enrichment of the inventory.	http://www.ogptunisie.gov.tn/en/?p=3448 https://registre.data.gov.tn/fr/
United Arab Emirates	Bayanat.ae	The official Open Data Portal of the United Arab Emirates provides access to financial datasets,	https://bayanat.ae/ https://opengov.unescwa.org/node/1217

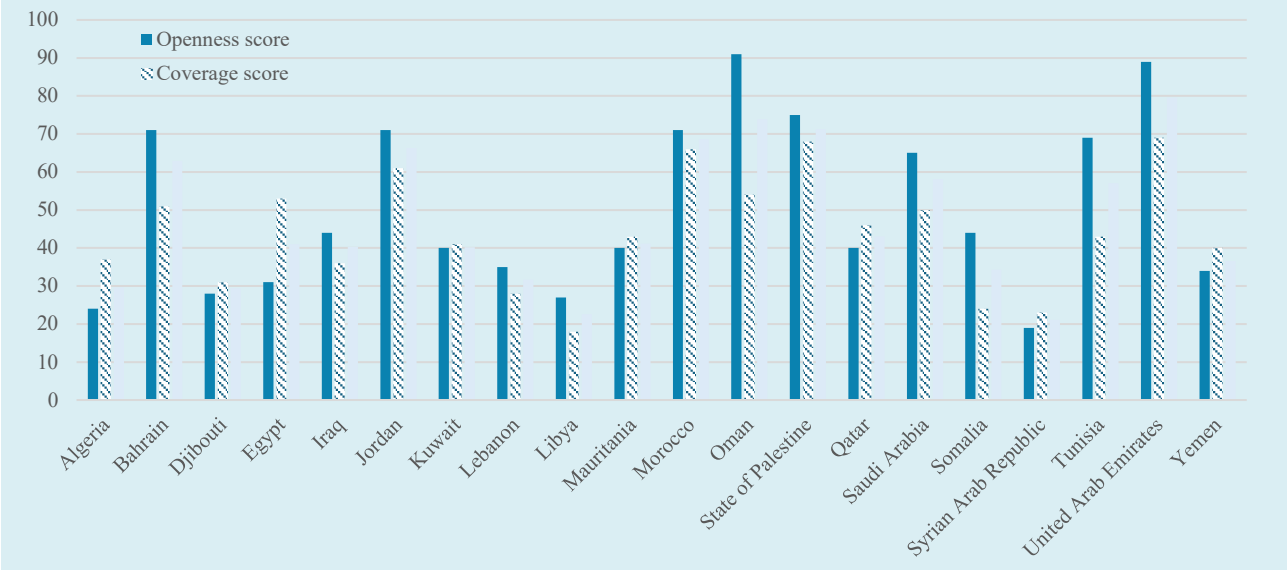
[11] Carnegie Endowment for International Peace, Cloud governance challenges: a survey of policy and regulatory issues, 2020.

Country	Initiative	Short description	More information
		including finance and banking, government budget and open procurement.	
	Ajman Open Data	It is the official data portal of the Government of Ajman, displaying free datasets from nine different entities in the emirate, covering eight main topics/subjects (business and industry; economic and finance; environment; health, well-being and care; housing; leisure and culture; public order, justice and rights; transport and infrastructure). Datasets can be used by researchers, application developers, startups, decision makers, business owners and developers.	https://opengov.unescwa.org/index.php/node/1189 https://data.ajman.ae/pages/homepage/

2. *Open data status and trends in the Arab region*

46. The Open Data Inventory (ODIN), produced by Open Data Watch, assesses the coverage and openness of official statistics, and is adopted by member countries and international observers of the Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs). The latest regional analysis report showed a slight decline in West Asia (-0.45) and North Africa (-0.4) in terms of coverage, and a significant improvement in West Asia (8.25) and slight decline in North Africa (-0.4) in terms of openness.[12] Arab countries' scores are reported in figure 2. ODIN values are available on the ESCWA platform "Index Simulator for Policymakers in the Arab Region" (ISPAR).[13]

Figure 2. Openness and coverage of official statistics in Arab countries



Source: Open Data Watch, ODIN biennial report, 2022.

47. Examples of open data initiatives include the following:

- The United Arab Emirates has been a frontrunner in adopting open data policies. The Dubai Data Law, enacted in 2015, mandates that government entities publish their data in an open and

[12] Open Data Watch, ODIN biennial report, 2022.

[13] Index Simulator for Policymakers in the Arab region. Available at https://ispar.unescwa.org/about.aspx.

accessible manner. The United Arab Emirates federal portal, Bayanat.ae, provides a wide range of datasets covering various sectors such as health, education and transportation.

- Saudi Arabia launched its National Data Management Office (NDMO) to oversee the implementation of open data policies as part of its Vision 2030 initiative. The Saudi Open Data Portal provides access to datasets from multiple government agencies, promoting transparency and innovation.
- The Ministry of Transport and Communications of Qatar (now called Ministry of ICT) has established the Qatar Open Data Portal, which offers datasets from various sectors including health, education and environment to enhance transparency, support decision-making and stimulate economic growth.
- The open data initiative of Morocco is governed by the Agency for Digital Development (ADD). The open data portal, data.gov.ma, provides access to datasets from different government departments, aiming to foster transparency and public participation.
- ESCWA has provided support to the State of Palestine since 2020 to build the capacities of government officials to implement the open government framework, including open data strategy, legal aspects of open data, participation strategies, open data platforms and the reuse of data. ESCWA also peer-reviewed the Open Data Strategy for the State of Palestine.

3. *ESCWA activities on open data*

48. ESCWA has been promoting open data initiatives across the Arab region. As part of a regional project on open government in the Arab region that ESCWA has been implementing since 2019, a framework on open government was developed, and several toolkits[14] were produced to build national capacities. Through its technical cooperation programme, ESCWA continues to support its member States in developing national open data policies, raising awareness and building national capacities on best practices for implementing open data initiatives and projects, and deploying the toolkits to guide Arab officials to develop open data policies and to assist them in the implementation phase.

G. Geospatial tools and applications for official statistics

49. Mandated to support the modernization of NSOs in the Arab region, ESCWA is promoting the integration of geospatial and statistical information, which is critical to advancing the United Nations Global Geospatial Information Management (UN-GGIM) agenda, particularly in supporting the SDGs. The convergence of these data streams enhances the capacity for comprehensive analysis, informed decision-making and effective policy implementation. This paper synthesizes insights from key studies, reports and initiatives to reveal the current state, challenges and future directions for geospatial information management in the Arab region.

1. *Global geospatial information management and global statistical geospatial framework*

50. UNSD and UN-GGIM have established a dedicated Expert Group on the Integration of Statistical and Geospatial Information, tasked with developing and advancing this global standard. The Global Statistical

[14] E/ESCWA/CL4.SIT/2021/MANUAL.2.

Geospatial Framework (GSGF)[15] was adopted at the ninth session of UN-GGIM[16] as an important bridge to linking statistical and geospatial information.

51. The GSGF sets forth principles and guidelines for the integration of geospatial and statistical data. It underpins efforts to harmonize data collection, improve data quality, and ensure interoperability across diverse geographic and statistical systems. The framework emphasizes the need for standardization and collaboration among national and international stakeholders to leverage geospatial information effectively for global development initiatives.

52. The geo-statistical outputs and tools resulting from integrating geospatial and statistical information are increasingly recognized as key drivers for improving the quality of both statistical and geospatial information. The benefits of these geo-statistical results extend beyond enhancing the quality of official statistics to include substantial support for subnational and national development priorities.

53. When adopted by member States, such an integrative process further supports governmental plans for the modernization of national statistical systems and improves the abilities of NSOs to cope with the many policy support demands brought forth by the 2030 Agenda for Sustainable Development.

2. *Integration of geospatial information with statistical information in support of SDG indicators*

54. ESCWA presented the study on the **"**Integration of geospatial information with statistical information in support of the SDG indicators" in 2021 in response to the recommendations of the ESCWA Statistical Committee.[17] The study provides comprehensive information on integrating geospatial and statistical information to support SDG indicators, showcasing best practices from around the world and the Arab region. It defines the pillars of the geospatial statistical information infrastructure and outlines policy principles for effective collaboration between NSOs, planning authorities and those concerned with geospatial information. The study highlights the pivotal role of integrated data in monitoring and achieving SDG indicators, presents methodologies for the seamless fusion of these data types, and addresses the technical and institutional challenges encountered in data integration, offering strategic recommendations for overcoming these barriers.

55. The United Nations Economic Commission for Europe (UNECE) working paper, entitled "Geospatial information services based on official statistics",[18] explores the development and deployment of geospatial information services that rely on official statistics. It illustrates the transformative potential of integrating new data sources, such as administrative and big data, within the statistical system. However, this integration poses challenges for the use of geospatial data. Small geographic level data are increasingly needed to support initiatives like the post-2015 development agenda and the SDGs.

56. While frameworks provided by UN-GGIM and the GEOSTAT 2 project facilitate this, administrative and big data can often only be aggregated to higher geographic levels due to disclosure and quality concerns. This results in a mismatch between the use of point-based data for statistics aggregation and the availability of datasets at larger geographic areas, which are insufficient for monitoring sustainable development. NSOs must therefore adopt disaggregation methodologies for providing detailed statistics. The implications of these methodologies on official statistics need thorough exploration through UN-GGIM and other initiatives.

[15] United Nations Statistical Commission, The Global Statistical Geospatial Framework, 2019.

[16] Ninth Session of the United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM), held on 5–9 August 2019. Available at https://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/.

[17] E/ESCWA/CL4.SIT/2021/TP.1.

[18] UNECE, Geospatial information services based on official statistics, 2016.

Increasing geospatial content in statistical processes raises the risk of disclosure, with overlapping geographies and grid references potentially identifying sensitive data.

57. While NSOs have developed disclosure control methodologies, new data sources and process changes could heighten these risks. The use of technical and non-technical standards is aiding the integration of statistics and geospatial information, though statistical and geospatial standards remain distinct with little harmonization effort. Recognized by the Open Geospatial Consortium[19] and the International Organization for Standardization, this gap requires engagement from the statistical community. The goal is to integrate more geospatial data into the statistical process, fostering a greater understanding of geospatial information's value and increasing spatial analysis alongside statistical analysis in reporting, thereby supporting the modernization of official statistics and adding value to the data produced by NSOs.

3. *UN-GGIM: Arab States*

58. UN-GGIM: Arab States[20] is a committee that was established in accordance with Resolution No. 4/111 issued by the fourth session of the United Nations Committee of Experts on Geospatial Information Management, held in New York in August 2014. It serves as the formal Arab chapter of the UN-GGIM initiative, with the core objective of coordinating, facilitating and enhancing regional cooperation in geospatial information. It also collaborates with other regional and international entities to promote best practices in this field.

59. At its tenth plenary meeting,[21] held in Abu Dhabi in February 2023 and organized by the Federal Geographic Information Center,[22] UN-GGIM: Arab States took the initiative to empower Arab women in geospatial information. Members agreed to develop a digital platform, managed by the secretariat of UN-GGIM: Arab States, to enhance the role of Arab women in geospatial information, build their capacities and competencies, and facilitate communication with experts through the UN-GGIM: Arab States website.[23]

60. In this regard, UN-GGIM: Arab States has provided home use licenses to 500 Arab women in geospatial fields across the Arab region over the course of three years. This initiative is a step towards gender equality in the technology sector, particularly in geospatial specializations. Recipients of these licenses will have full access to training materials through the Esri Academy e-learning platform, enabling comprehensive self-paced learning and keeping professionals updated on the latest developments in geospatial technology.

4. *ESCWA Geo-Statistical Laboratory and case studies*

61. The ESCWA Geo-Statistical Laboratory uses earth observation and remote sensing data, geospatial information systems and data science tools to compile the most comprehensive geospatial information on Arab countries from official and open data sources. Several case studies contribute to the compilation and analysis of the SDGs, especially indicators 11.3.1 (land consumption rate) and 9.1.1 (proportion of rural population with access to roads), which are difficult to compile from other traditional sources.

62. The ESCWA Geo-Statistical Laboratory developed two regional land cover accounts (cross-sectional and time series) to assess the changing shares of different land cover classes across the entire Arab region. Understanding these changes sheds light on the impacts of urbanization, the intensity of crop and animal production, afforestation and deforestation, the use of water resources and the changing ecological capacities of land. Land cover change is a surrogate measure for gauging the degree of decoupling between economic

[19] Open Geospatial Consortium website, available at https://www.ogc.org/.

[20] UN-GGIM: Arab States website, available at https://un-ggim-as.org.

[21] UN-GGIM: Arab States website, available at https://www.un-ggim-as.org.

[22] Federal Geographic Information Center website, available at https://fgic.gov.ae/.

[23] UN-GGIM: Arab States website, available at https://www.un-ggim-as.org/.

and ecological potential capacities of natural resources. Aggregated land cover across the Arab region is processed to detect change in land cover classes. The aggregated land cover classes used in this study are: (1) built-up area; (2) agriculture-crop; (3) inland water; and (4) natural cover. The estimates are non-official statistics that cover the entire region. Country-specific estimates (national or subnational) are available upon request via the ESCWA Geo-Statistical Laboratory.

5. *Results*

63. Based on the cross-sectional analysis (1992 versus 2020) methodology,[24] the results showed that the built-up area in the Arab region has increased from 12,418 km² to 32,365 km². Most of this increase is due to the conversion of natural land cover (11,631 km²), agricultural-crop land cover (8,154 km²) and inland water cover (162 km²) to built-up areas.

64. Based on mid-resolution satellite data (300 m x 300 m), between 1992 and 2020, the ecological potential capacity of natural landscapes (such as barren, tree cover and inland water) across the Arab region has depreciated at a rate equal to the current area of Cairo and Dubai combined, year over year. The above trend estimates are based on non-parametric regression on 29 annual data points. The statistical intervals exclude the reported remote sensing processing accuracy of 75 per cent on land cover type, classifications and identification. The actual interval estimates should thus be wider. Such estimates shed light on the ecological potential capacities of landscapes and their sustainability across the Arab region. The ESCWA Geo-Statistical Laboratory is conducting detailed analyses on land cover mapped over socioeconomic statistics and coupled with a suite of environmental, geographical, demographic and location-based attributes towards disseminating a set of geo-statistical indicators for the region, from streets to neighbourhoods, using higher resolution data with related statistics.[25]

65. The integration of geospatial and statistical information is vital for achieving the SDGs, and provides comprehensive analysis, informed decision-making and effective policy implementation. By fostering a greater understanding of geospatial information's value and increasing spatial analysis alongside statistical analysis, the global geospatial community can significantly contribute to sustainable development and the modernization of official statistics.

66. Through various initiatives, such as the GSGF and the collaborative efforts of organizations like ESCWA, UNECE and UN-GGIM: Arab States, significant strides have been made in harmonizing data collection, enhancing data quality, and ensuring interoperability between geographic and statistical systems. These efforts have not only improved the quality of official statistics but also provided substantial support for national and subnational development priorities.

67. Key challenges remain, such as the need for small geographic level data, the risks of data disclosure, and the harmonization of statistical and geospatial standards. Addressing these challenges requires continued innovation in data integration methodologies, robust policy frameworks, and effective collaboration among NSOs, planning authorities and geospatial information experts.

[24] Cross-sectional analysis (1992 versus 2020) and time series analysis (1992 to 2020) are conducted to estimate land cover change across the Arab region. The advantage of the time series analysis over the cross-sectional analysis is the ability of the former to produce statistical confidence intervals on the estimates, while the latter provides a snapshot of the estimates. The script for non-parametric regression detects and estimates trend in time series of annual values for distribution-independent regression. The statistical methods used are the non-parametric Mann-Kendall test for testing the presence of the monotonic increasing or decreasing trend and the non-parametric Sen method for estimating the slope of a linear trend. The Mann-Kendall test requires at least 4 values, and the calculation of the confidence intervals for the Sen slope estimate requires at least 10 values in the time series. These threshold limits are satisfied in the current analysis.

[25] For further details, tables and maps, see https://geoportal.un.org/arcgis/apps/storymaps/stories/8bd78fadc81d49d580f05ff23e0a9219. Other case studies on floods and droughts in the Arab region using Google Earth Engine and time series data from different data sources can be found on https://data.unescwa.org/GIS?country=#.

68. Developing case studies like the land cover change in the Arab countries, and floods and droughts geospatial tools are key to better planning and risk management.

III. Pooling resources and expertise for modernization

69. NSOs are recently extending partnerships with government agencies, universities and private companies to face significant challenges in meeting the increasing demand for complex, up-to-date and transparent data. These strategic collaborations empower NSOs to tap into specialized expertise, cutting-edge technology and innovative data collection methods. For example, a project in the Arab region used big data from mobile phone records to gather socioeconomic information that traditional methods could not capture. This transformative impact of collaboration enhanced data quality and addressed evolving information needs.

70. Modern data landscapes are forcing NSOs to adapt and partner with other entities according to modern partnership frameworks, shifting from one-off collaborations to partnerships built on shared risks, co-creation and cultural transformation. A strong partnership framework acts as a governance tool, fostering collaboration and leveraging the distinct strengths of partners, i.e. NSOs' data expertise and private sector's technological prowess. NSOs are moving from being data providers to data stewards, responsible for quality and compliance, and promoting best practices within the wider statistical ecosystem.

71. Building a successful partnership requires stepwise implementation: starting with meticulous planning to identify stakeholders and define issues, followed by implementation through the establishment of clear governance structures and project initiation. In the adaptation phase, progress is monitored, and strategies are adjusted. Finally, the evolution stage prompts decisions to continue, scale up or end the partnership based on its effectiveness.

72. NSOs have a range of partnership options, each suited for specific goals. Data sharing agreements enrich datasets by granting access to information from government agencies or private companies. While these agreements broaden the data spectrum for robust analysis, they require clear protocols to address legal and ethical concerns. Joint projects, which combine expertise from diverse partners—such as NSOs, academia and others—tackle complex challenges. While these projects maximize skillsets and resources, effective project management remains crucial for seamless coordination. Additionally, capacity-building initiatives enhance staff skills through training programmes, empowering them to adapt to modern data science demands. However, this requires a long-term commitment as benefits may not be immediate. Selecting the optimal collaboration type hinges on factors like project objectives, resource availability, urgency and legal considerations. By strategically evaluating these criteria, NSOs can make informed partnership choices to achieve desired outcomes.

73. Successful partnerships for NSOs rely on strong principles. Clearly defined goals, roles and expectations set the foundation, and are formalized in agreements or plans. Each partner leverages their strengths to contribute to the shared vision, while a designated leader ensures smooth operation. Trust and commitment develop over time through open communication. Consistent information-sharing and transparency in decision-making foster understanding and collaboration while adapting to changing dynamics. A well-defined exit strategy ensures a smooth conclusion at the end of the partnership.

Case studies of successful NSO collaborations

- **Private sector partnership:** ESCWA established a partnership with the Central Administration of Statistics (CAS) in Lebanon and Data-Pop Alliance to leverage call detail records generated by mobile network operators to examine the dynamics of Syrian refugees in Lebanon. As part of this collaboration facilitated by ESCWA, the Lebanese Ministry of Communications, which is the mobile call records owner, entrusted CAS with the responsibility of storing and making the data available for case study analysis, while CAS was committed to data confidentiality.

- **Academic and research centres collaboration:** ESCWA collaborated with experts at the Hamad Bin Khalifa University Qatar Computing Research Institute to utilize Facebook as a source of big data for official statistics - from data extraction and cleaning to analytics, retrieval, inference and prediction - for various applications, including image and video manipulation, traffic signal optimization and inferring road graphs from satellite imagery. This collaboration demonstrated a commitment to solving fundamental and challenging problems in impactful domains.
- **International organization partnership and multi-stakeholder collaboration:** CAS partnered with the United Nations High Commissioner for Refugees (UNHCR), which acts as the custodian of data derived from the Vulnerability Assessment of Syrian Refugees in Lebanon (VASyR) Survey that feed into the annual report published by UNHCR since 2013. These data provide an overview of the geographical variations in vulnerabilities of refugee populations at the district and governorate levels in Lebanon. Another collaborative project facilitated by ESCWA on "leveraging big data and alternative data sources to respond to crises and disasters in Lebanon and Jordan" included CAS, UNHCR, and the Council for the Orientation of Development and Ethics (CODE).

IV. Conclusion and recommendations

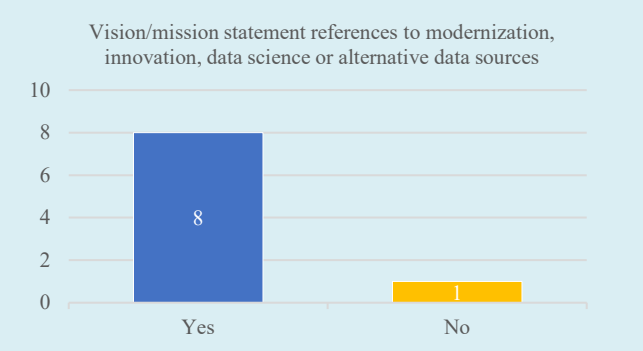
74. Based on the cases and studies in the present document, the following recommendations are made to NSOs: (a) adopt cloud computing for official statistics; (b) experiment with specific applications of data science tools in official statistics, especially those provided as case studies in the present document; (c) implement data standards such as SDMX for all possible domains in statistics; (d) adopt open data strategies; (e) integrate geospatial tools and applications for official statistics; (f) pool resources and expertise for modernization; and (g) leverage different types of partnerships and collaborations.

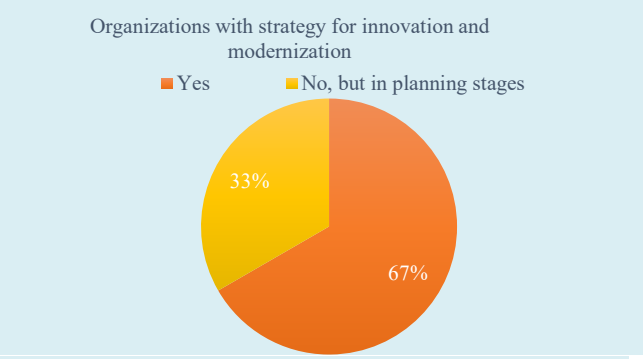
75. The Statistical Committee is invited to provide comments on the content of the present document, especially its recommendations.

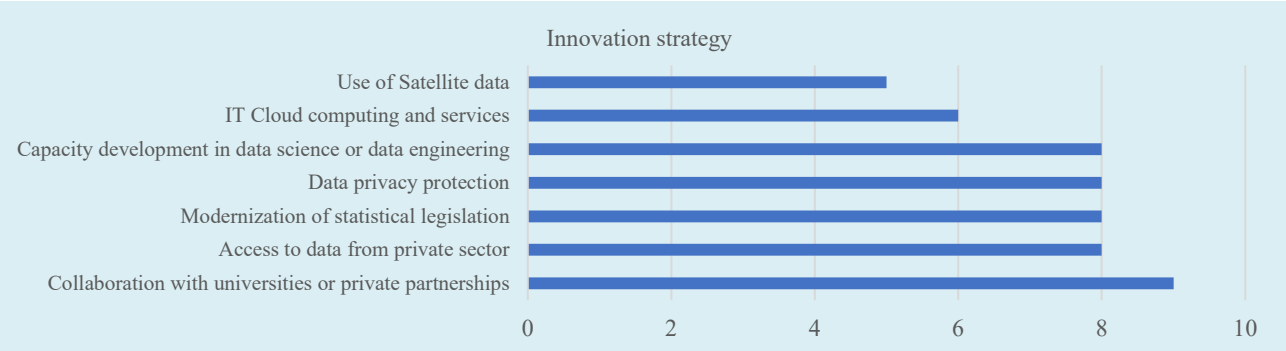
Annex

Results of the United Nations Statistics Division - United Nations Committee of Experts on Big Data and Data Science for Official Statistics (UNCEBD) survey in the Arab region

Strategic vision







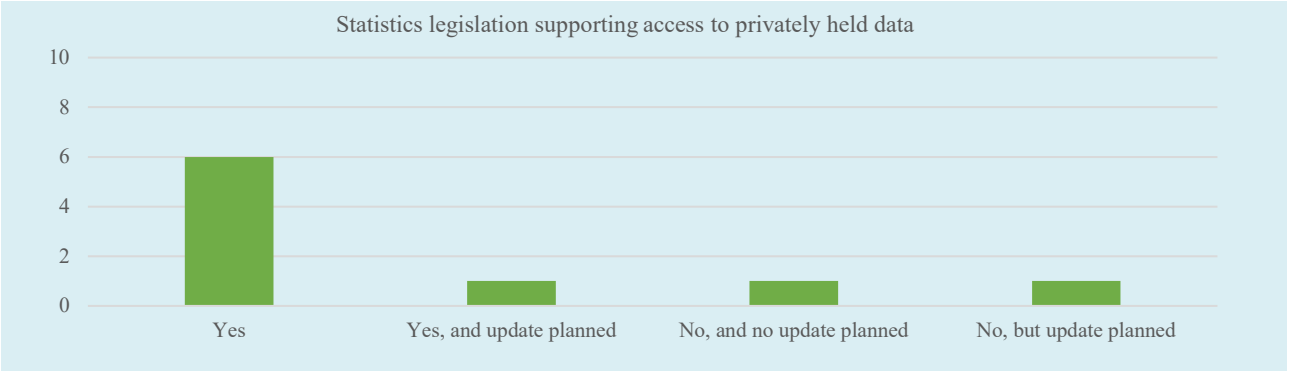
Vision/mission references:

- Eight out of nine organizations have references to modernization, innovation, data science or the use of alternative data sources in their vision or mission statements.
- Only one organization did not have such references.

Innovation strategy:

- Six organizations have an established strategy for innovation and modernization.
- One organization is in the planning stages of developing an innovation strategy.
- Two organizations do not have an innovation strategy.

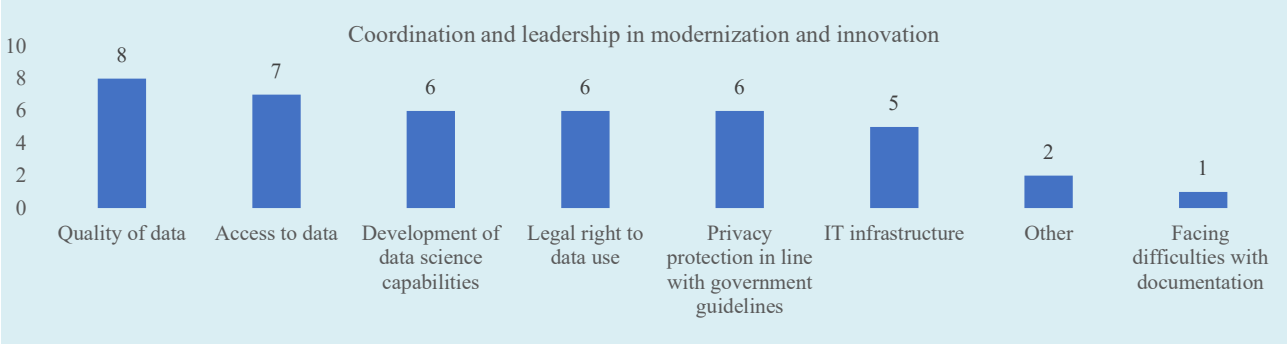
Legislation



Specific legislation:

- Six organizations have specific legislation supporting access to privately held data.
- One organization has plans to update its legislation.
- One organization tried and failed to implement such legislation.
- One organization does not have such legislation, and no updates are planned.

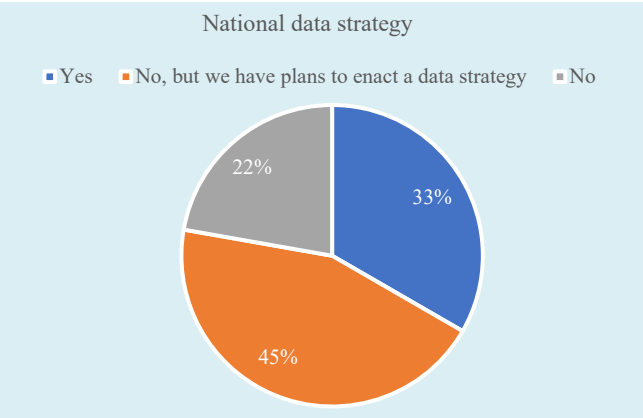
Institutional arrangement and partnerships

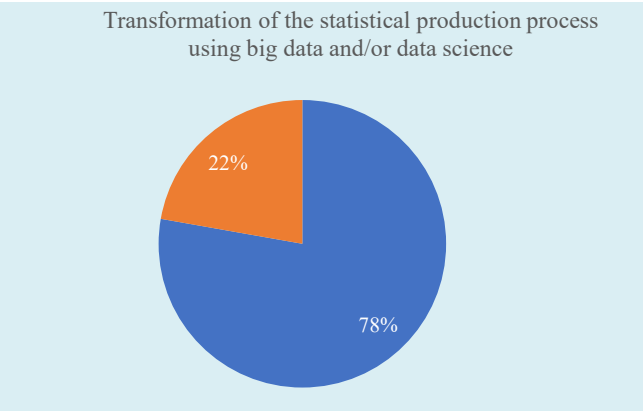


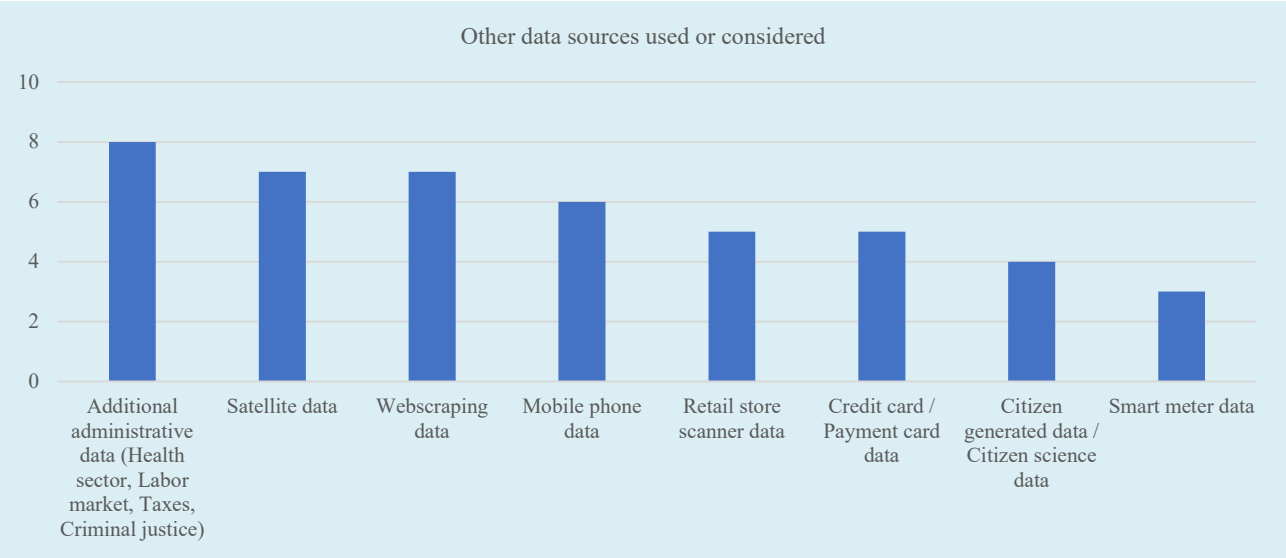
Stakeholder relations:

- Organizations maintain strategic relations with various stakeholder communities, including government agencies, international agencies and private sector companies. These relationships are essential for their data and statistical activities.

Data sources, methods and quality assurance







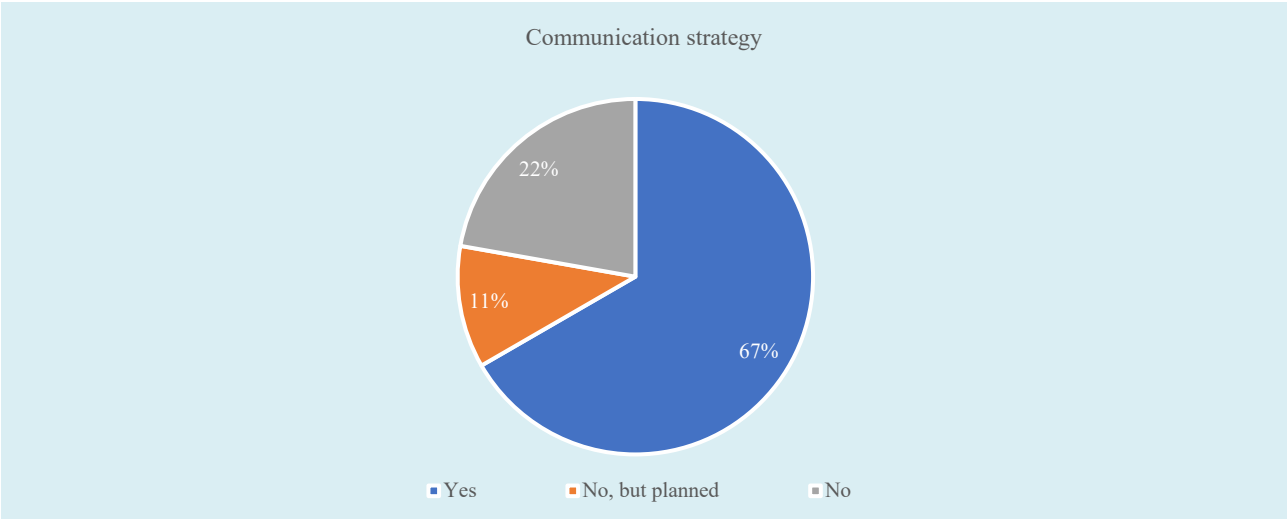
National data strategy:

- Four organizations have a national data strategy in place.
- Four organizations have plans to enact a national data strategy.
- One organization does not have a national data strategy.

Alternative data sources:

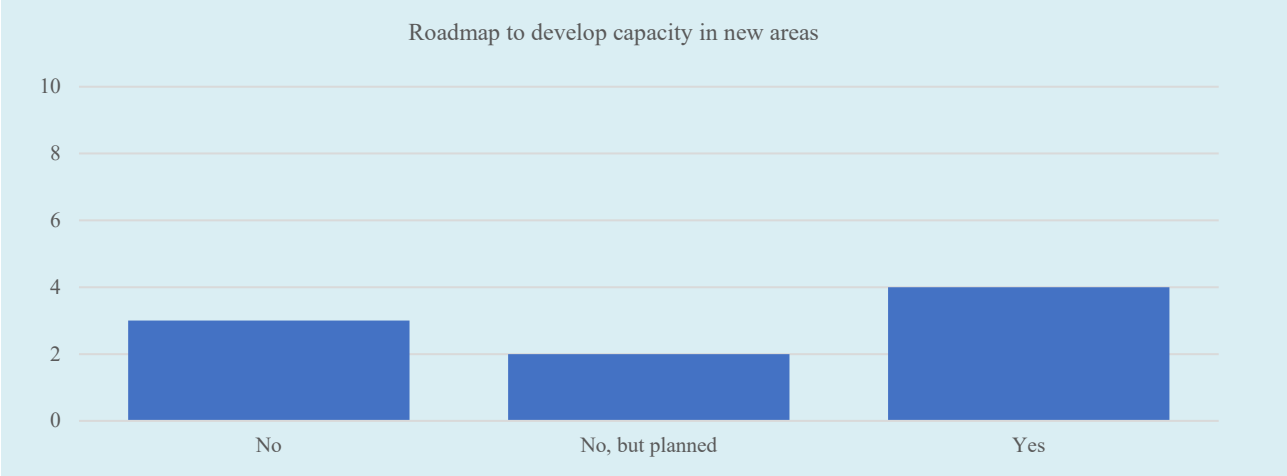
- Commonly used data sources include satellite data, mobile phone data, web scraping data and additional administrative data from various sectors.
- These data sources are pivotal to enhancing the statistical capabilities of organizations.

Communication strategy



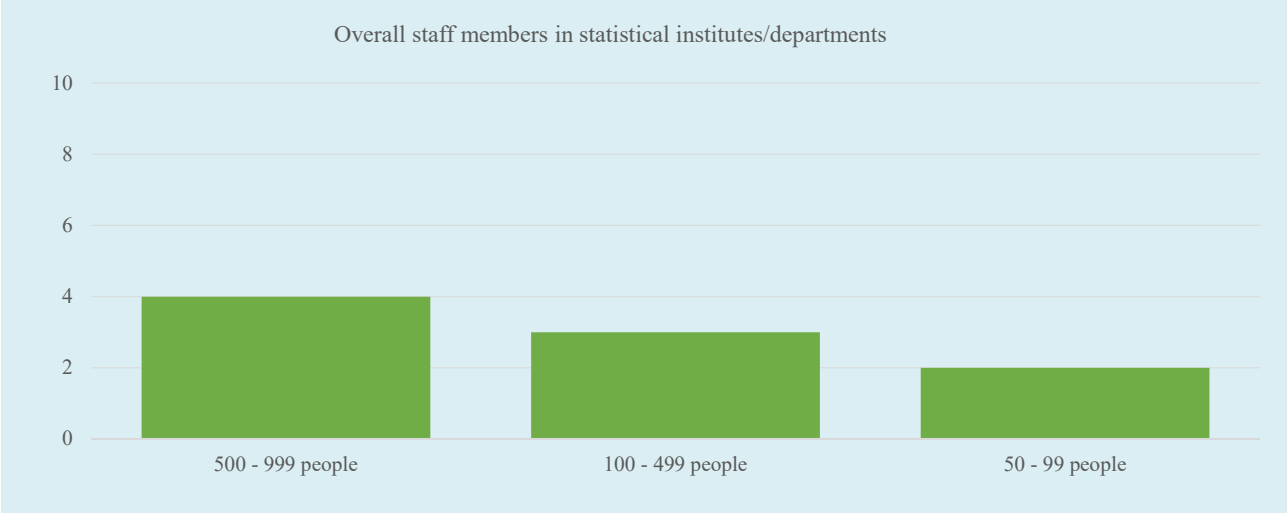
- Countries were asked if they had a communication strategy and whether they were communicating results using new technologies or new data sources.

Human resources

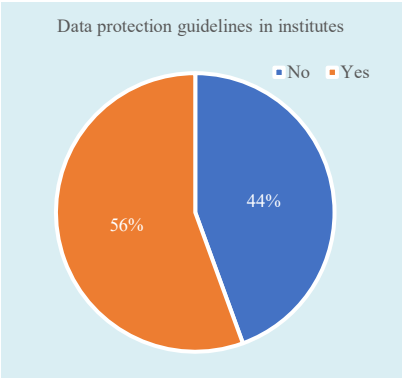


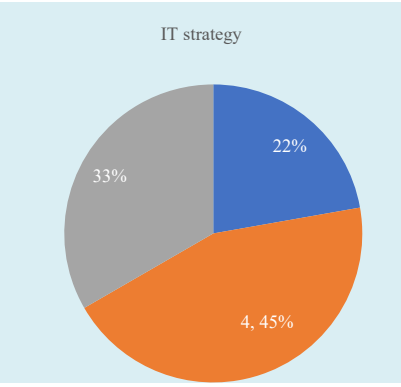
Staffing:

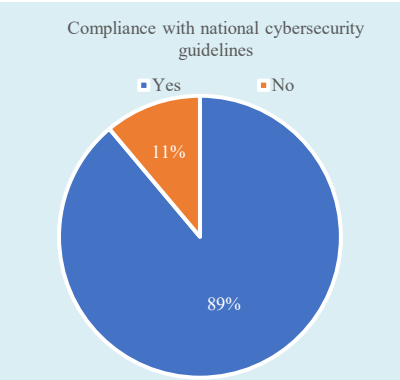
- The number of staff members in statistical institutes varies widely, with some organizations having between 500 to 999 staff members.
- The number of data scientists employed also varies, ranging from none to over 20.



IT management

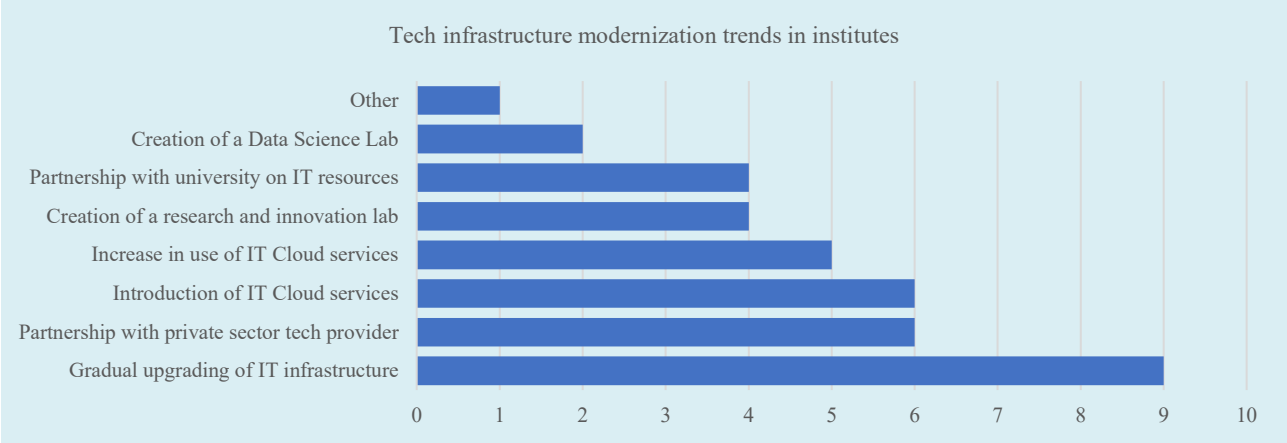






Technology infrastructure changes:

- Many organizations have introduced or are planning to introduce significant changes to their IT infrastructure.
- Common initiatives include creating data science labs, research and innovation labs, and forming partnerships with private tech providers and universities.



Participation in UNCEBD activities

Participation in UNCEBD conferences:

- Most organizations have participated in at least one UNCEBD conference.
- Participation is spread across multiple conferences held from 2014 to recent years.

Participation in UNCEBD task teams:

- Organizations are actively involved in various task teams, especially those focused on satellite data, mobile phone data, scanner data, web scraping and privacy-enhancing technologies.

Interest in UNCEBD regional and sector hubs:

- There is significant interest in participating in regional hubs, particularly those based in Rio de Janeiro, Hangzhou and Kigali.
