PARiS 21

Partnership in statistics
for development
in the 21st century

OPEN DATA WATCH

DATA DISSEMINATION IN THE DIGITAL AGE

A review of data portals for official statistics in IDA-eligible countries

PARIS21: The Partnership in Statistics for Development in the 21st Century (PARIS21) promotes the better use and production of statistics throughout the developing world. Since its establishment in 1999, PARIS21 has successfully developed a worldwide network of statisticians, policy makers, analysts and development practitioners committed to evidence-based decision making. With the main objective of achieving national and international development goals and reducing poverty in low- and middle-income countries, PARIS21 facilitates statistical capacity development, advocates for the integration of reliable data in decision making and co-ordinates donor support to statistics.

ODW: Open Data Watch (ODW) is an international, non-profit organisation working at the intersection of open data and official statistics. ODW work supports the implementation of change in the production and management of official statistical data. Concentrating efforts in three areas – policy advice, data support and monitoring – ODW seeks to make development data better and more accessible for increased use and impact. An example of the interaction between these workstreams, the Open Data Inventory, provides in-depth annual assessments of coverage and openness that countries use to identify and address data gaps. Only with high-quality and open data can international organisations, governments and citizens solve the challenges of measuring and achieving the Sustainable Development Goals.

Please cite the report as: PARIS21/ODW (2021), *Data Dissemination in the Digital Age: A review of data portals for official statistics in IDA-eligible countries*

PREFACE

Data portals form a critical part of a data system's architecture. They are one of the principal links between data users and producers. As gateways for data access, they serve a vital function in today's digital data ecosystem, facilitating the use of data for evidence-based decision making. The information technology and digital revolution, alongside citizen movements for transparency and open data, have contributed to the proliferation of data portals.

Since the outbreak of the COVID-19 pandemic in early 2020, many new data portals and dashboards from private and public actors have appeared, providing information for monitoring the pandemic. Data, statistics, graphs and tables on these portals serve as a primary window into the rapidly evolving status and impact of the crisis across the globe. However, the proliferation of data portals for disseminating data and statistics predates the pandemic.

National statistical offices (NSOs) have embraced data portals as an instrument for official data dissemination – especially in the wake of the data-driven 2030 Agenda for Sustainable Development. They have been supported in their efforts by international organisations, bilateral donors and civil society organisations that have sponsored the development of portals for the wider dissemination of development indicators. However, data portals need to be designed and implemented in a sustainable manner to be beneficial.

Previous work by Greenwell et al. (2016) noted that design, deployment and technology considerations were key determinants for the adoption of data portals by NSOs. However, more systematic evidence is required to understand the implementation status and performance of NSO data portals. While many institutional reviews of data portals of official statistical agencies may have taken place in the past to inform capacity development efforts, reports in the public domain remain scant. An assessment of such portals against a comprehensive set of principles and guidelines would be a welcome addition to the existing literature. A broad assessment of data portals is particularly important in the context of aiding low-income countries with limited statistical capacities to harness the gains from the data revolution equitably.

This report provides a first account of the state of data portals containing indicator data maintained by national statistical offices in IDA-eligible (International Development Association) countries.

Section 1 gives the context for this study, describing why a critical examination of national statistical offices' data portals is required.

Section 2 presents an overview of the literature on data portal assessments.

Section 3 provides the overall methodology, including the definition of data portal used in this report. It also describes the 11 categories of data portals used in the analysis, how the portals were identified, the criteria used to evaluate them and caveats to consider. The PARIS21-ODW methodological framework operationalises evaluation of data portals based on the United Nations Statistical Division's *Principles of SDG Indicator Reporting and Dissemination Platforms and Guidelines for their Application* (UNSD, 2019[1]).

Section 4 provides findings from data portal assessments. Readers interested in the findings can skip directly to this section, which provides findings for each guideline separately, alongside the guideline-specific methodology, analysis and implications.
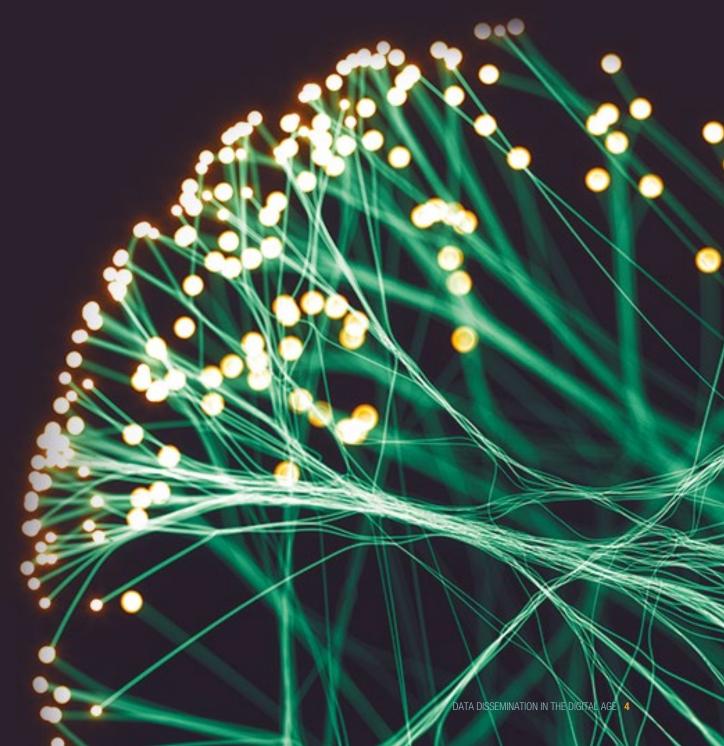
Section 5 provides five recommendations for the development data community based on the findings

Section 6 introduces ideas for future research and development on data portals of official statistical agencies.

ACKNOWLEDGEMENTS

CONTENTS

FIGURES

TABLES AND BOXES

ABBREVIATIONS AND ACRONYMS

API	Application programming interface
FAO	Food and Agriculture Organization
IDA	International Development Association
IT	Information technology
NMDI	National Minimum Development Indicators
NSO	National Statistical Office
ODIN	Open Data Inventory
ODW	Open Data Watch
PARIS21	Partnership in Statistics for Development in the 21st Century
RSS	Really simple syndication
SDGs	Sustainable Development Goals
SDMX	Statistical Data and Metadata eXchange
UNICEF	United Nations Children's Fund
UNSD	United Nations Statistics Division
URI	Universal Resource Identifier

EXECUTIVE SUMMARY

This study by the Partnership in Statistics for Development in the 21st Century (PARIS21) and Open Data Watch (ODW) proposes a holistic methodology to evaluate data portals that disseminate statistical indicators, based on the *Principles of SDG Indicator Reporting and Dissemination Platforms and Guidelines for their Application* proposed by the United Nations Statistics Division in 2019. This study provides a concrete way to operationalise the principles and guidelines to enable an extensive assessment of NSOs' data portals. This assessment evaluates data portals of the 74 IDA-eligible countries from December 2019 to September 2020 using criteria drawn from the principles and guidelines, including, but not limited to, national ownership, accessibility, user-centricity[1], data communication and scalability.

The study includes portals that:

▶ Allow interaction for the user to search or filter data through a search bar or a drop-down menu of data options

▶ Have the option to download data in machine-readable formats

▶ Provide access to statistical macrodata or indicators (as opposed to microdata or other types of geospatial data, for instance)

▶ Are found on or linked to an NSO website.

78 macrodata portals were found in 53 out of 74 IDA-eligible countries using the criteria described above.

KEY FINDINGS

Almost a third of the IDA-eligible countries in the study did not have a data portal. In countries where data portals are present, adherence to the principles and guidelines is uneven. Although portals perform better on some of the foundational requirements of the guidelines, like data communication and data disaggregation, many of the more advanced guidelines are far from being implemented.

Specific findings include:

▶ Extensive implementation of data communication features: 94% of portals had tables; 77% had graphs or charts; 73% had maps; one portal had a data subscription feature available.

▶ Extensive implementation of data disaggregation features: About 90% of portals included disaggregated data features.

▶ Moderate implementation of multilingualism and accessibility features: Only about 10% of portals have a 'good' Google Lighthouse accessibility score and 60% of portals were only available in one language.

▶ Moderate implementation of metadata features: 19% of portals were missing all metadata and 75% had incomplete metadata.

▶ High reliance on externally developed portals: 86% of data portals were developed with support from international agencies. While the data portals are often developed with the necessary support of development partners, this kind of centralised approach can reduce an NSO's ability to customise the portal according to their local needs.

▶ Inadequate implementation of standardised interfaces: 45% of portals supported application programming interfaces (APIs), but only 3% had documentation for their APIs.

▶ Inadequate implementation of open data features: Less than 3% of portals had a "terms of use"; only one had a "terms of use" that conformed to CCO, CC-BY or other open data licenses; less than 8% of portals had a bulk download option available.

---

1 User-centricity refers to the solution design approach in which extensive attention is given to the intended user's characteristics, including the environment and workflows. User needs drive the resulting functionalities, leaving aside secondary issues like aesthetics (Norman, 1986)

KEY RECOMMENDATIONS

1. Prioritise a back-to-basics approach

This study demonstrates that a wide variety of functions expected of well-run data portals are not adequately implemented. A pragmatic approach should be adopted in implementing the principles and guidelines, prioritising features that will provide a higher return on investment. For example, improving metadata availability and multilingualism can be prioritised over advanced features such as linked data and APIs.

2. Adopt user-centric design to account for the needs of end users

User-centric design (Guideline 4) can assist NSOs to better understand and respond to the needs of their end users. This is critical to promoting the use of their data by prioritising data portal developments and data releases that will have the highest impact on users. Quantitative feedback mechanisms, like Google Analytics, along with more qualitative methods for gaining feedback, such as focus groups and interviews, can produce a holistic picture of user needs along with success stories and statistics on the importance of data portals to assist NSOs in advocating for increased funding.

3. Advance national ownership of data portals
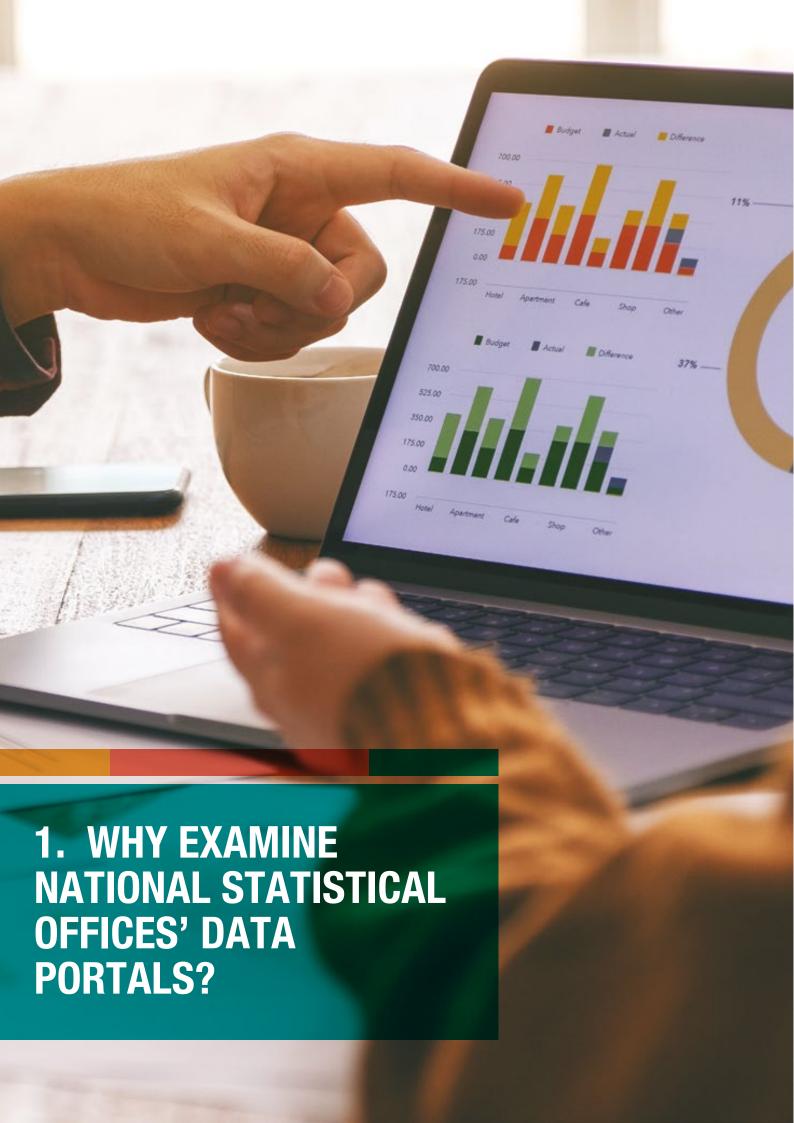
A significant proportion of NSOs in developing countries rely on external partners to perform key aspects of the design, implementation and management of their data portals. Nevertheless, NSOs should be fully engaged in the selection of a portal and the dissemination of their national data. Development partners should support workshops, training and regular consultations to bring relevant country actors together to assess priorities and address the issues highlighted in this study. They should co-ordinate their work, putting data standards at the heart of their initiatives so that they complement (and do not duplicate or compete with) each other and support countries' national strategies for the development of their statistical systems.

4. Improve upstream data management practices for a sustainable data dissemination infrastructure

A well-designed data portal is one part of the larger data dissemination infrastructure. The long-term sustainability of a data portal rests on robust data management practices, including an overarching digital process orientation of NSOs; for example, data modelling and standards that propel a shift from paper to digital methods and enable repurposing and synergies within the national statistical system.

5. Streamline data dissemination processes to reduce the reporting and management burden for maintaining data portals

Strategies for creating integrated, yet modular, "all-in-one" data portals and streamlining data dissemination processes could reduce the reporting and management burden on NSOs. Some countries have no portals; others have three or four that may have duplicate content and functions. While in some cases it may be unavoidable to have multiple data portals geared to different audiences, if these are not connected to one another through well-designed (and to the extent possible, automated) data flows, the NSO and other agencies within the national statistical system may find themselves manually maintaining separate but parallel dissemination processes, increasing their burden. Data portals should be interoperable and integrated as much as possible to minimise the burden on data providers and ensure that users do not have to search many portals to find the data they need.

1. WHY EXAMINE NATIONAL STATISTICAL OFFICES' DATA PORTALS?

1. WHY EXAMINE NATIONAL STATISTICAL OFFICE DATA PORTALS?

1.1. DATA PORTALS ARE KEY FACILITATORS OF DATA ACCESS AND USE

Demand for better data systems from both users and producers of data has made data portals one of the most prominent mechanisms for data dissemination and access in the modern data ecosystem. Movements for government transparency and open data originating from civil society and citizen groups have been an important driver of the creation of data portals to open access to government statistics and data. With the launch of the international Open Data Charter, the Open Data for Development Network (OD4D) and the Open Government Partnership, the political will to publish open data became mainstream. On the producers' side, rapid changes in information and digital technologies galvanised widespread efforts to modernise statistical business processes in national statistical systems (Vale, 2021[2]), accelerating the adoption of data portals to disseminate data.

The adoption of the Sustainable Development Goals (SDGs) in 2015 and its accompanying, data-driven agenda, renewed global attention and action to leverage the digital data revolution to serve both data producers (hereafter "national statistical agencies") and users. Especially in the context of low- and middle-income countries (OECD, 2017[3]), NSOs embraced an active role, championing openness, transparency, and access to data and methods by disseminating official data using digital platforms and leveraging statistical standards. This is reflected in the Cape Town Global Action Plan for Sustainable Development Data (CT-GAP), adopted in 2017 by the United Nations Statistical Commission. Strategic Area 4 of the plan outlines key actions for the "dissemination and use of sustainable development data." These include: 1) promoting the development of technological infrastructure for better data dissemination; 2) leveraging the experience of the Millennium Development Goals in using online methods for the dissemination of SDG statistics,

including the use of Statistical Data and Metadata eXchange (SDMX); and 3) developing effective communication and data dissemination strategies and guidelines for public and private dialogue oriented to policy makers, legislators, the media, the general public and the economy.

Today, data portals form a critical part of the infrastructure of official statistics: they are the principal link connecting data users and data producers. They perform a range of functions as platforms for data publication, data discovery, monitoring of public policies, and engagement with data users and citizens, among others. Data portals are key enablers of data access and use. They serve a vital function in the overall virtuous data cycle by which good data encourages greater use leading to improved policies and outcomes, and which, in turn, increases demand for data. But the existence of a data portal does not guarantee that the portal or the data it contains will be used. Fit-for-purpose data portal design and architecture are critical to facilitating access to and use of data portals.

1.2. REALISING THE BENEFITS OF DATA PORTALS NECESSITATES A CRITICAL REVIEW OF THEIR IMPLEMENTATION

Data portals provide several benefits and opportunities for NSOs. They enable easier and efficient use and reuse of official statistics at scale, leading to greater return on investment than ad hoc or one-off dissemination methods. Well-designed data portals are integrated with an underlying database, allowing seamless, error-free updating. They feature easily accessible, machine-readable data that can be presented in tables or visualisations (charts, graphs) ready for consumption. Data portals also attract more and different kinds of users than websites that lack these features, which can lead to wider use and recognition of official statistics and greater public trust in them. Further, data portals provide NSOs with the opportunity to leverage their

institutional role as a co-ordinator of the national statistical system to provide technical assistance to other data providers through a centralised national data portal, thereby increasing their political capital and influence (World Bank, 2014[4]).

However, data portals must be well designed and implemented in a sustainable manner so that NSOs are able to realise the opportunities and benefits they provide. Previous work by Greenwell et al. (2016[5]) observed that in the wake of the United Nations' data-driven 2030 Agenda for Sustainable Development (2015) and the SDGs, the enthusiasm and push for data portals have created sustainability and implementation challenges. Further, design and deployment considerations during data portal development and maintenance were found to be key determinants of their quality and usability. The study by Greenwell et al. sparked an important international consultation on the state of data portals (especially in aid-dependent countries) at the 2018 UNSD Conference on National Platforms for SDG Reporting, leading to the identification of best practices and solutions for developing and implementing national reporting and dissemination platforms.

Today, the open data movement finds itself at a critical juncture. While significant progress towards open data has been made – as measured by open data indexes like the Open Data Inventory – progress has been unevenly distributed along the data value chain (Open Data Watch, 2018[6]).

Because attention has been primarily focused on the production side, factors that facilitate data use remain underexplored. If we evaluate and reflect on the data for the decision-making process, we know that just making data available does not necessarily mean they will be used. As Hidalgo (2016[7]) points out, "The goal of open data should not be just to open files, but to stimulate our understanding of the systems that this data describes." Further, while substantial efforts have been made to make data open and accessible, progress in data dissemination practices and openness has been unevenly distributed in different parts of the world. *The World Bank's World Development Report 2021: Data for Better Lives* provides a comprehensive overview of the state of development data globally and finds that the perspective of lower income countries is often missing from international development data discussions. These countries also tend to perform less well on open data and data dissemination practices and standards (World Bank, 2021[8]).

The true potential of the digital data revolution can be equitably achieved when modern statistical production architectures (including data portals) mobilise data access and use. Hence, realising the returns to the digitalisation of statistical processes and unlocking the benefits of the open data movement warrants a deeper examination of the implementation and outcomes of data portals, particularly in capacity constrained countries.

1.3. THIS STUDY PROVIDES THE FIRST ACCOUNT OF THE STATE OF PLAY OF NSO DATA PORTALS IN IDA-ELIGIBLE COUNTRIES

Applying a systems view to data dissemination entails understanding, monitoring and updating the digital data practices of NSOs and assessing their data portals for factors that affect data use. However, there has been little to no effort to systematically review the data portals of official statistical agencies against a set of comprehensive criteria. A broad assessment of data portals is particularly important in the context of aiding low-income countries with limited statistical capacities to harness the gains from the data revolution equitably. Evidence of the benefits of data transparency on long-term economic growth further reinforces the need to examine the state of data access in developing countries (Islam and Lederman, 2020[9]).

This joint study by the Partnership in Statistics for Development in the 21st Century (PARIS21) and Open Data Watch (ODW) aims to fill this gap. It builds on the work by Greenwell et al. (2016[5]) and expands the parameters of data portal assessment beyond design and technology considerations.

In particular, the study proposes a holistic methodology to evaluate data portals, based on the UNSD's Principles of SDG Indicator Reporting and Dissemination Platforms and Guidelines for their Application (UNSD, 2019[1]). This assessment evaluates data portals of 74 IDA-eligible countries using criteria drawn from the principles and guidelines, including but not limited to national ownership, accessibility, user-centricity, data communication and scalability. The study provides a concrete way of operationalising the principles and guidelines to enable an extensive assessment of NSOs' data portals and, implicitly, a test of the applicability of the principles and guidelines to data portals currently in use. The evaluation was performed between December 2019 and September 2020 and provided insights into the state of data portals for that time. However, the methodology can be used to update these baseline findings in the future to track how portals are progressing in their application of the principles and guidelines.

Based on the results of the evaluation, the study provides recommendations to actors in today's development data ecosystem – such as NSOs, development partners and data portal developers – to improve their data portals, integrate user-centric design and adopt a systems lens to sustainable data dissemination practices. The findings of the study contribute to increasing returns to the digitalisation of statistical business processes and closing the last-mile gap of the open data movement.

2. WHAT DO WE KNOW ABOUT DATA PORTALS TODAY?

2. WHAT DO WE KNOW ABOUT DATA PORTALS TODAY?

Although NSO data portals are a critical component of national statistical infrastructure, not much is known about their use and performance, especially in low- and middle-income countries. Most assessments of data portals have examined portals that provide data from line ministries or subnational data sources rather than data from NSOs. Even then, the evidence from low- and middle-income countries is limited. This section reviews recent assessments of data portals at the global, regional and national levels.

At the global level, Kubler et al. (2018[10]) conducted a comparison of metadata quality in data portals using their Analytic Hierarchy Process that integrates various data quality dimensions and end user preferences. They used the framework to assess over 250 e-government data portals from organisations across 43 countries. They concluded that most organisations do not pay sufficient heed to dataset management, resources or the associated metadata they disseminate on their portal.

At the regional level, Bello et al. (2016[11]) evaluated 22 data portals from 17 African countries on the following parameters: implementation technology, data formats, licensing, major datasets, functionality and Berners Lee's 5-star model (2012[12]) for linked open data. Steinberg and Castro (2017[13]) analysed data portals in 20 Latin American countries and found high variance in the amount of data published, the data formats and the popularity of data portals in the region. Berends et al. (2020[14]) evaluated selected data portals from seven European countries. Their report makes recommendations to enhance the sustainability of data portals in five key areas – governance, finance, architecture, operations and metrics – based on interviews, secondary research and practical experience

At the national level, Kumar et al. (2020[16]) assessed the accessibility, openness, usability, and technical functionality of 22 publicly available data portals that feature data related to Nepal. The review found that many data portals in the country are not using the best search engine optimisation practices and dissemination strategies and have load time and technical accessibility issues. Zhu and Freeman (2018[17]) developed a User Interaction Framework with criteria in five dimensions: 1) access; 2) trust; 3) understand; 4) engage☐integrate; and 5) participate. They deployed the framework to evaluate the open government data sites created and maintained by 34 municipal government agencies in the United States. Their review found that the portals perform well on data access, but their focus on user engagement and participation was inadequate.

BOX 2.1. SPOTLIGHT: THE OPEN DATA MATURITY MODEL

The European Data Portal, launched in 2015, monitors the development of national open data policies and portals in Europe. Since it was launched, the annual **Open Data Maturity study** serves as a benchmark for the development of open data in Europe. It assesses the level of maturity against four dimensions: 1) policy; 2) portal; 3) impact; and 4) quality. The study clusters countries into four groups: 1) trendsetters; 2) fast trackers; 3) followers; and 4) beginners, from the most to the least mature.

Figure 2.1. Open Data Maturity dimensions and dimension-specific metrics

DIMENSION	METRICS
Open data policy	Policy framework Governance of open data Open data implementation
Open data impact	Strategic awareness Political impact Social impact Environmental impact Economic impact
Open data portal	Portal features Portal usage Data provision Portal sustainability
Open data quality	Currency Monitoring and measures DCAT-AP compliance Deployment quality and linked data

Source: Blank (2019[15])
Note: DCAT-AP Data Catalog Application Profile used to describe public sector datasets.

Going beyond government data portals, Neumaier et al. (2016[18]) performed an automated quality assessment of metadata across data portals of both private and public institutions. Their findings are based on monitoring of 260 data portals with 1.1 million datasets. They included a discussion of general quality issues: for example, the retrievability of data and the analysis of our specific quality metrics.

Looking specifically at data portals for official statistics, Open Data Watch with AidData and PARIS21 invited seven NSOs in low- and middle-income countries to participate in a study analysing web traffic on their principal websites or data portals using Google Analytics (Open Data Watch, 2018[6]). Among other results, the study found that data portals separated from the NSO website receive less web traffic than portals on NSOs' websites. NSO websites tend to be better optimised for search engines, and portals on those websites, consequently, gain more traffic from search engines.

3. WHAT ASSESSMENT METHODOLOGY IS USED TO EXAMINE DATA PORTALS?

3. WHAT ASSESSMENT METHODOLOGY IS USED TO EXAMINE DATA PORTALS?

This study examines data portals from IDA-eligible countries using the 4 principles and 12 guidelines proposed in the Principles of SDG Indicator Reporting and Dissemination Platforms and Guidelines for their Application (UNSD, 2019[1]). Although it is aimed at the development of SDG reporting platforms or portals, the document advises, "Given that the scope of the platform is wider than reporting and dissemination for the SDGs, the platforms can also be referred to as [a] 'national data platform'" (p2), thus clarifying the generalisability of these principles and guidelines. The current study applies these principles to propose a methodological framework to assess any NSO's data portal.

The main products of this research are the findings from the evaluations of data portals against the principles and guidelines. However, the PARIS21-ODW methodological framework for the assessment of data portals is a useful secondary output. Creating actionable recommendations from the principles and guidelines can provide countries with a clearer path to improving their portals and adhering to the guidelines. It can also enable the international community to better support progress toward the implementation of the principles and guidelines.

3.1. WHAT IS A DATA PORTAL AND WHAT TYPES EXIST?

This study used the definition of data portal from the forthcoming *Handbook on Statistical Organization* (UNSD, 2021[19]) as a guide: "A data portal is a web-based, interactive data and metadata platform with databases modelled for specific data types and domains such as microdata, macrodata or geospatial data" (see Table 3.1 for alternative definitions).

Applying one interpretation of the elements in the above definition of "web-based, interactive data" and "modelled databases", this study includes data portals that:

▶ Allow the user to search or filter data through a search bar or a drop-down menu of data options. Static pages that contain only preselected data (in PDF, Excel, or other formats) or that do not pull data from a database or an API through a search, filter or other user-input method are not considered to be data portals.

▶ Have the option to download files in machine-readable formats. Dashboards that contain features like search, modifiable data tables and visualisations but do not allow downloading in machine-readable formats are not included in this study.

The above steps for filtering data portals operationalise the UNSD's definition to differentiate between data portals and other websites that host data. To further narrow the focus of inquiry, the following criteria were also considered in the selection of data portals:

▶ The study evaluates data portals that contain only indicators. This was primarily because the technical considerations (including underlying processes) for implementing microdata or geospatial data portals differ from portals disseminating indicators (or statistical data). Portals containing indicators are also arguably the most prevalent portals used by NSOs today. The research team, however, recorded any microdata portals and geospatial data portals associated with the NSO.

Table 3.1. Data portals: Alternative definitions

Definition	Source
A data portal is a web application, website or page of a website that holds data from different sources, organised under subsets or categories to make it simple for the users of the site to find.	Zuar, 2020
A data portal is a data exploration tool with a customised public web interface that allows scientists, managers and the general public to discover and access public data. The data portal has three major components: 1. data catalogue 2. data map 3. data views.	Secoora
A data portal is a web platform for publishing data. The aim of a data portal is to provide a data catalogue, making data not only available, but discoverable, for data users, while offering a convenient publishing workflow for publishing organisations. Typical features are web interfaces for publishing and for searching and browsing the catalogue, machine interfaces (APIs) to enable automatic publishing from other systems, and data preview and visualisation.	Open Data Handbook
A data portal is a gateway to data, which can be big or small, open or restricted. Its core purpose is to enable the rapid discovery and use of data. However, as a flexible, central point of truth on an organisation's data assets, a data portal can become essential data infrastructure and be extended or integrated to provide many additional features: • data storage and APIs • data visualisation and exploration • data validation and schemas • orchestration and integration of data • data lake co-ordination and organisation.	Datopian Tech

Note: API: application programming interface.

▶ The study only includes data portals that are found on or linked to an NSO website. One of the purposes of this study is to provide recommendations to NSOs on how to improve the implementation of their data portals and to understand how these portals are managed. Consequently, limiting the study to these portals will provide the most relevant results. Linking to the NSO website is interpreted as a proxy for the NSO's endorsement of that data portal.

3.2. HOW WERE THE PORTALS IDENTIFIED?

The process of finding the data portals linked to the NSO websites from the 74 IDA-eligible countries was intended to replicate the steps that a user would go through to find a portal. Although this method may not find all the data portals linked to an NSO website, it should find those that users are most likely to encounter and should be representative of the official portals available in a country.

Based on the approach and understanding of data portals described in Section 3.1, the following steps were taken to identify data portals in this study:
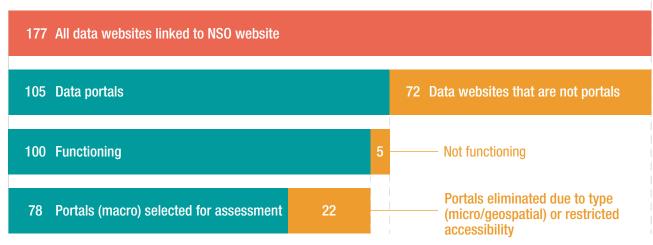
▶ **Step 1:** Only sites that were directly hyperlinked to the main NSO website were evaluated. For this, the principal NSO website was searched for data-hosting sites by checking the following sections:

1. Data portals or data-hosting websites directly linked on the home page.

2. Tabs on the menu bar of the site that say: data, statistics, data portal or other data-related terms.

3. A site search for the terms "data portal", "database" and "data". Sites that are not in English were searched for using the corresponding translations.

▶ **Step 2:** Data portals were then carefully differentiated from other data-hosting websites (such as dashboards or other platforms) based on our interpretation of "web-based, interactive data" and "modelled databases," outlined in Section 3.1.

▶ **Step 3:** Only functioning data portals were then selected for further evaluation.

▶ **Step 4:** Macrodata portals were then filtered out from other kinds of data portals (such as microdata or geospatial portals), as per the scope of the study.

The identification process is illustrated in Figure 3.1.

Seventy-eight data portals in 53 out of 74 IDA-eligible countries were identified through the method described above and recorded in the research database. Data portals from Step 4 that did not qualify were also recorded to obtain an overall picture of the data dissemination landscape (see Section 4.1), but they were not included in the final assessments against the principles and guidelines (see Section 4.2).

Figure 3.1. How this study identifies data portals

177 All data websites linked to NSO website

| 105 Data portals | 72 Data websites that are not portals |

| 100 Functioning | 5 — Not functioning |

| 78 Portals (macro) selected for assessment | 22 — Portals eliminated due to type (micro/geospatial) or restricted accessibility |

Source: Authors' elaboration based on PARIS21-ODW desk assessment of data portals

3.3. WHAT CRITERIA WERE USED TO EVALUATE THE DATA PORTALS?

The UNSD's principles and guidelines provide a standardised benchmark for evaluating data portals and are intended to be universally applicable. The principles, slightly adapted, are as follows:

1. Clear institutional arrangements and management
2. Fit for purpose
3. Sustainability
4. Interoperability and statistical standards.

The 12 guidelines address the development and evaluation of data portals consistent with the principles (UNSD, 2019[1]).The relationship between the 4 principles and 12 guidelines are presented in Annex A. In the current study, a set of criteria, primarily posed as questions about observable characteristics of the data portal, were used to assess whether a data portal broadly complies with the guidelines, and by extension, the principles. These questions were developed using portions of Open Data Watch's Data Site Evaluation Toolkit (DSET) (Open Data Watch, 2021[20]), an evaluation system used to identify the elements of a well-functioning and open data site. These are presented in the complete PARIS21-ODW methodological framework in Annex B.
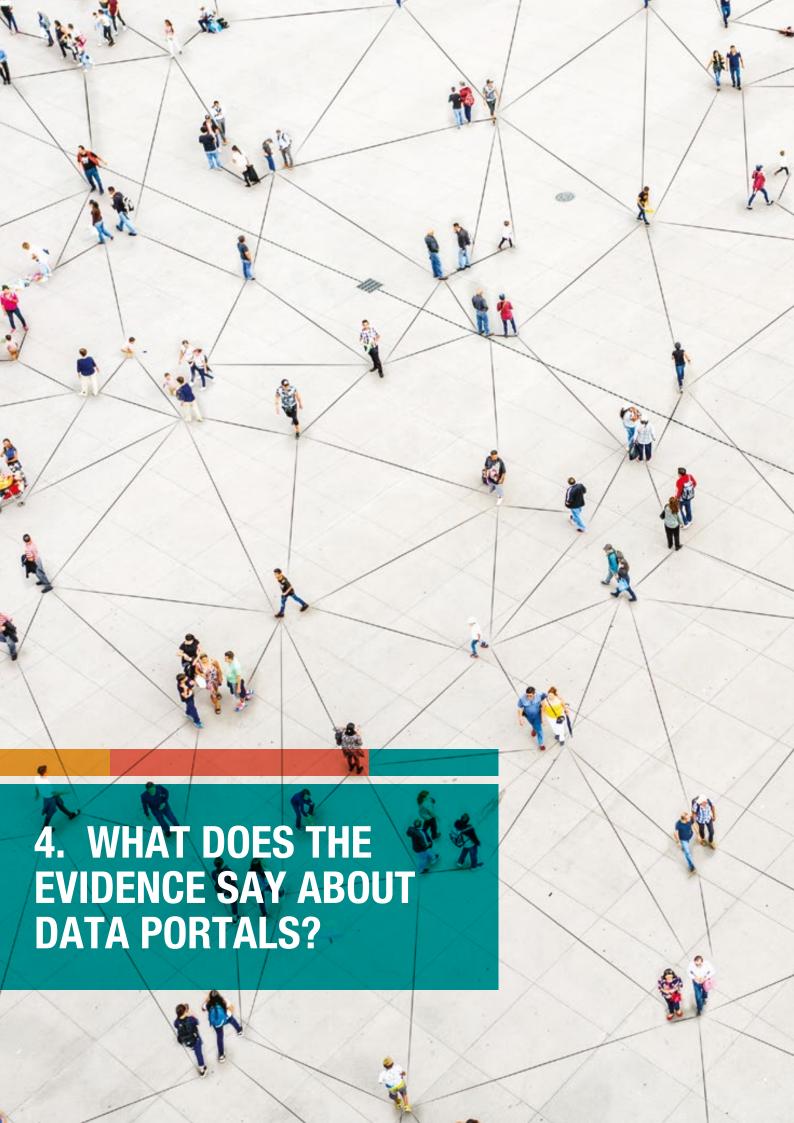
3.4. CAVEATS: WHAT THIS STUDY DOES NOT DO

When reviewing the findings, the following caveats should be kept in mind:

1. The methodological framework consists of a mix of desk research and surveys to enable a holistic assessment of data portals against the wide-ranging set of principles and guidelines. However, the findings in this report pertain to only the desk research portion of the assessment. Guideline 2 on collaboration is hard to evaluate remotely as information about partnerships and co-ordination between stakeholders in the national statistical system is not readily available on NSOs' websites. Likewise, Guideline 9 on scalability requires knowledge of the portals' design that is not apparent to external users. A survey was designed to evaluate these aspects, but there was not a high enough response rate to provide valid findings for all IDA-eligible countries. Consequently, it was not possible to evaluate the implementation of some portions of the principles and guidelines. Further, the desk research portion of the analysis does not constitute detailed findings on data portal traffic or user behaviour, as it cannot be gleaned from publicly available information. (Privileged access to web analytics of the NSOs' websites would be needed to properly assess web traffic.)

2. This research project studies only data portals and not all websites where NSO data may be hosted or from where data can be downloaded from an NSO website. Therefore, it is not a complete representation of all the data dissemination channels that a user might encounter when looking for data. As noted earlier, the research team used an operational definition of a data portal to select the sites to study. Only data portals that are meant primarily for the dissemination of official statistics, particularly those disseminating macrodata, are included in this study.

3. This study only provides insight into how IDA-eligible countries are implementing the principles and guidelines in their data portals. A broader study of non-IDA countries might show greater variation in their implementation or other variations in the number and quality of data portals available in each country.

4. Not all aspects of the guidelines can be measured. The questions that are mapped to the guidelines are the research team's best approximation of how to implement and measure the guidelines. There is, however, room for debate on whether other characteristics of data portals could be measured and mapped to different guidelines.

This study is not designed to be the final word on how to measure adherence to the guidelines, nor does it attempt to evaluate the completeness and sufficiency of the guidelines themselves. Rather, it is a first step towards developing clearer, actionable steps for NSOs, development partners and data portal managers to improve their data portals and better serve their user community.

4. WHAT DOES THE EVIDENCE SAY ABOUT DATA PORTALS?

4. WHAT DOES THE EVIDENCE SAY ABOUT DATA PORTALS?

This analysis of data portals in IDA-eligible countries demonstrates that adherence to the principles and guidelines is uneven. Although their portals perform well on some of the more foundational requirements for the guidelines, like data disaggregation and data communication, many of the more advanced guidelines are far from being implemented. Advanced guidelines in this report are categorised as those that require more technical capacity and resources from the NSO. Furthermore, almost a third of the countries in the study did not have a data portal to evaluate. With such a broad range in the status of data portals, the findings in this report are critical to prioritising the next steps for improving the application of the guidelines and, ultimately, the further development and use of data portals.

4.1. WHAT KINDS OF DATA PORTALS WERE FOUND AND WHAT DO THEY TELL US ABOUT THE DATA DISSEMINATION LANDSCAPE?

Seventy-eight data portals were identified in 53 of the 74 IDA-eligible countries. Twenty-one countries, or 27% of IDA-eligible countries, in the evaluation did not have data portals on or linked to their NSO website. While these countries might have official statistics hosted on their NSO's website, their users do not have access to the functionalities that data portals can provide. Figure 4.1 illustrates the distribution of data portals found in IDA eligible countries, ranging from 0 portals in 21 countries to 4 data portals in 3 countries.
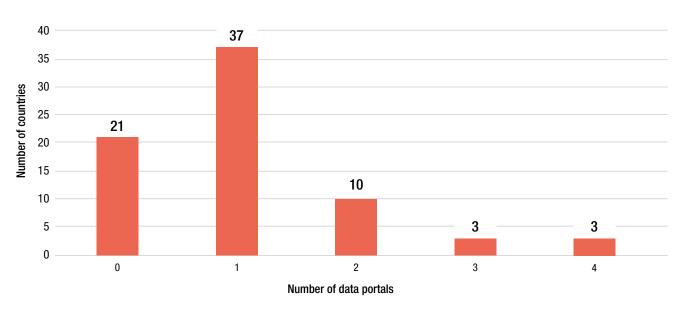
Figure 4.1. Almost a third of IDA-eligible countries do not have any data portals linked to their NSO's website

Source: PARIS21-ODW desk assessment of data portals

As noted in Section 3.2, the research team also noted other websites where data were hosted – even if they did not qualify as data portals based on our definition – and found that such sites were almost as numerous as data portals, possibly signifying a lag in countries' implementation of more advanced data dissemination platforms. For example, the research team found about 80 websites where data were available through static links to PDFs or Excel files. Data posted on these sites can be more challenging to search for and require manual updates to connect the NSO website to new datasets.

Further, some NSO websites did not have any links to data portals; others had broken links (which resulted in a "404 Not Found" error) or were linked to offline portals throughout the evaluation period (December 2019-September 2020). This presents a complex picture of the overall data dissemination landscape, characterised by numerous broken links and out-of-date portals linked from the NSO sites. This also signals the overstretched capacity of the NSO staff, burdened with maintaining multiple databases, platforms and portals.

The research team found that data portals that are integrated into the main NSO website typically perform much better on the principles and guidelines evaluation because these sites can leverage the existing infrastructure and contextual information (like "About us" pages) to provide a better experience to the user. Integrating portals into an NSO website could improve the usability and accessibility of many portals. This type of centralised approach can also streamline data access for users by providing all the data they need without having to visit many different external portals.

Out of the 78 selected data portals, 85% were found to be created or managed by international organisations, such as the Food and Agriculture Organization (FAO) and the African Development Bank. These externally developed portals were hosted on different domains than the NSO's website.

The data portals included in this study fall into the 11 following categories:

Open Data for Africa portals: Developed by the African Development Bank, the Open Data for Africa portals are the most common type of portal found

by the study; they are present on 30 NSO websites. The portals provide a flexible array of options for countries to report on different datasets, such as SDG data, census data and gender indicators, and have SDMX support through the "national summary data pages". Through a link with the Knoema data platform, Open Data for Africa portals also provide visualisation options for datasets. As a one-stop shop for countries to report on many official statistics, the Open Data for Africa portals meet many of the principles' and guidelines' criteria.

NSO-made data portals: Developed internally by the NSO or through a contractor, these portals are typically integrated into the NSO website and disseminate official statistics. Twelve of these portals were found in the evaluation. There is a wide variety of designs and features available on NSO-made data portals. The implementation of NSO-made data portals is more common in higher income countries. Only two were found in countries classified as low income.

CountryStat portals: Developed by the FAO, the CountryStat portals disseminate food and agriculture statistics. Although the portal was designed and implemented by the FAO, the organisation formed partnerships with NSOs and the ministries of agriculture, fisheries, forestry and others to introduce the CountryStat portal and build capacity to use it. The FAO no longer maintains CountryStat portals and has turned their management over to countries. Eight CountryStat portals were found and evaluated in this study.

REDATAM portals: The "REtrieval of DATa for small Areas by Microcomputer" (REDATAM) portals were developed by the Latin American and Caribbean Demographic Centre (CELADE) and are used for the dissemination of statistics and censuses, particularly the population and housing census. These portals were available in seven countries. Although they are designed to disseminate microdata, unit record data were not available on the evaluated platforms, which only allow access to aggregates of the data, so they function as indicator portals to the public.

PopGIS portals: Developed by the Pacific Community's (SPC) Statistical Development Division, the PopGIS portal disseminates macrodata and

geospatial data from the Pacific Islands community. The portal has been implemented in seven countries included in this study.

DevInfo portals: Developed by the United Nations Children's Fund (UNICEF), DevInfo was originally developed for the Millennium Development Goals, but countries have adapted the platform to disseminate a variety of datasets. Although DevInfo is no longer supported by UNICEF, the portal is still being used by seven countries in this study.

Open SDG portals: Developed by the UK Office for National Statistics, the US government, the Center for Open Data Enterprise (CODE) and members of the Open SDG community Sustainable Development Goals (SDG), Open SDG is an open-source platform for managing and publishing data about the SDGs. This portal is being used by three countries in this study.

PxWeb: Developed by Statistics Sweden, PxWeb is used for publishing statistics from a database and has been available free of charge for government agencies and municipalities, international statistical institutions, and international organisations of statistics since 2016. Kosovo was the only country in the evaluation that used PxWeb.

NMDI portals: The National Minimum Development Indicators (NMDI) portals were developed in 2011-12 by the SPC to assist countries using regional and international reporting frameworks, such as the United Nations Millennium Development Goals. Although this portal is no longer supported, it is still used by the Solomon Islands.

.Stat Suite portals: Developed by the Statistical Information System Collaboration Community under the Organisation for Economic Co-operation and Development (OECD) Secretariat, the .Stat Suite is an open source standard-based data platform used for the efficient production and dissemination of official statistics. Cambodia was the only country found implementing a .Stat Suite portal in this evaluation.

Other portals: Two portals found in the evaluation did not fit into any of the above categories and were lumped together as "other." One was a Google Site; the other was a regional platform that also disseminated national data.

4.2. FINDINGS BY GUIDELINE

The evaluation of data portals was based on the response of the data portal to remote queries or information obtained using web tools that analyse the performance of websites. This is referred to as "desk research" because it did not involve direct contact with the NSO or the data portal managers. Although more information could be obtained through interviews or questionnaires sent to portal managers, this proved to be impractical for the present study. All the guidelines, with the exception of Guideline 2: Collaboration, were evaluated by desk research findings mapped to the corresponding questions of the PARIS21-ODW methodological framework (see Annex B).

This section reviews findings from each of the guidelines to provide an analysis of data portals in IDA-eligible countries and their progress implementing the principles and guidelines. For each of the sections, the corresponding guideline is excerpted from the principles and guidelines and the methodology for assessing the guideline is provided. The research team restricted their analysis to only those aspects of each guideline that were directly noted in the text so that the desk research would be manageable. However, there may be aspects of each guideline that are not covered in this assessment.

Guideline 1: National ownership

How do the guidelines describe national ownership?

National statistical offices should have the ability to maintain, adapt, transform and customise their data portals to address their own needs and the needs of their users such as the management of subnational administrative boundaries, country-specific ethnic and language groups, and additional indicator definitions related to national development priorities.

What does the PARIS21-ODW framework measure?

The study team was unable to assess the adaptability or potential for customisation of data portals through desk research. Therefore, to evaluate national ownership, the study team reviewed the following aspects of the data portal:

1. **Does the "About us" or another page on the data portal provide information on who hosts, manages or maintains the site?** If so, is the NSO listed as the managing organisation or the responsible agency?

2. **Is the data portal linked to an international organisation's data portal?** (This can help determine whether data portals were developed externally.)

The research team counted only "About us" pages that were hosted on the same subdomain as the portal and so excluded some "About us" pages from higher level domains. These were excluded because "About us" pages that are not on the same subdomain as the portal were not likely to contain information about the portal but about the site or organisation at large that might not be relevant for the portal.

What do the findings show?

Analysis from the desk research showed that fewer than half of the portals studied had an "About us" page with management information and only 34% of all portals had an "About us" page with the NSO listed as the organisation responsible for managing it.

As shown in Table 4.1, about 85% of the portals found were linked to a data portal that was developed externally by an international organisation. About 93% of low-income IDA countries' data portals were externally developed, but there is a heavy reliance on portals created by international organisations among all IDA-eligible countries.

Table 4.1. Data portals by type and income group

Portal type	Share of portals (%)	Number of portals	IDA-eligible countries (Number of countries)		
			Low income	Lower middle income	Upper middle income
Open Data for Africa portals	38	30	17	13	0
CountryStat portals	10	8	5	3	0
REDATAM portals	9	7	0	4	3
Popgis portals	9	7	0	4	3
DevInfo portals	9	7	3	3	1
Open SDG portals	4	3	1	2	0
Other portals	3	2	0	2	0
.Stat Suite portals	1	1	0	1	0
NMDI portals	1	1	0	1	0
PxWeb	1	1	0	0	1
Externally developed portals	85	66	26	33	7
NSO made portals	15	12	2	9	1
Total portals reviewed	**100**	**78**	**28**	**42**	**8**

Notes: Externally developed portals are calculated by adding the different portal types, apart from PxWeb. PxWeb is counted as an NSO-made portal by the authors. This is because although the national statistical office (NSO) uses PXWeb, it is not mentioned on the web page. It is also possible that the NSO has modified the portal format to fit its needs so it is closer to our classification of an NSO-made portal.

Figure 4.2. Data portal ownership: High reliance on externally developed portals

Source: PARIS21-ODW desk assessment of data portals

What do the results imply?

The lack of "About us" pages and management information reveal a gap in public-facing information about the portals and potentially a lack of ownership control over the portals. Countries' reliance on external portals (such as Open Data for Africa, REDATM, CountryStat and others) may limit their ability to manage them or to adapt them to their needs. While NSOs load data onto these portals and have some options on which features to implement, they do not have as much control over them as they would with an internally developed portal. Further, they are not in control of whether the portal software and user interface are updated and maintained. This has become a problem for DevInfo portals, which are no longer being maintained by UNICEF.

Externally developed portals fill a critical gap in data infrastructure, so there may be a balance to be struck between national ownership and leveraging available resources in the near term. Open-source and modular data portals may be another potential solution to this challenge, as they would provide more options for country ownership. However, the capacity and resources required to manage the code for those portals may still be a difficult hurdle to overcome.

Guideline 2: Collaboration

How do the guidelines describe collaboration?

Data portals should be designed, developed, improved and maintained using a collaborative approach that leverages learning between various stakeholders of the national statistical system as well as technology developers, donors, policy makers, subject-matter experts, business partners, advocacy groups, and both institutional and grassroots users.

What does the PARIS21-ODW framework measure?

Information about collaborations to design, develop or maintain the data portals should come directly from those responsible for the portals. The survey portion of the assessment framework (see Annex B) included a set of questions to elicit information on the internal and external collaboration among data portal developers and stakeholders. Although the response rate was too low for analysis, the questions could be used by NSOs or other stakeholders as part of a self-assessment process.

This study also examined the use of open-source software to build platforms such as Open SDG as evidence of support for collaborative practices.

Open source is defined as code or software that is made available for free for use, modification and distribution. The implementation of open-source software is widespread across the Internet and many sites use open-source libraries or pieces of open-source code. The research team examined if country data portals clearly noted their use of open-source code on the portals (through text on the "About us" page or through a reference of GitHub account) and second, recorded when websites about data platforms stated that they were built with open-source code.

What do the findings show?

The data portals of five countries explicitly noted the use of open-source code; these were all adaptations of the Open SDG portal. Further research found 21 more data portals using open-source code. However, only five instances clearly stated their use of open-source code and invited others to collaborate.

What do the results imply?

The use of open-source software supports collaboration by permitting the use and reuse of code that has been developed by different groups. Open-source portals that can be reused by many countries is a more sustainable and efficient option for supporting data dissemination than creating separate portals for every country. However, creating and implementing open-source solutions requires significant co-ordination across the international statistical system and significant technical capacity at the NSO to implement. Further, as this study found, some portals don't clearly state their use of open-source code on the country portal, which might be a lost opportunity to invite others to collaborate and assist in developing the code for their platforms. But the success of platforms like Open SDG, and the growing adoption of .Stat Suite, shows that open-source portals can be successfully implemented in a variety of countries and could be a viable option, especially with more international support.

Guideline 3: Multilingualism and accessibility

How do the guidelines describe multilingualism and accessibility?

To leave no one behind, ensure national ownership, and promote the use and impact of data for policy and decision making at the local level, data portals should support national languages and implement national and international best practices in terms of accessibility to persons with disabilities, as well

as full access across the range of browsers and devices, including mobile devices.

What does the PARIS21-ODW framework measure?

To evaluate this guideline by desk research, the study reviewed the following aspects of the data portal:

1. What are the languages in which the data portal can be made available?

2. What is the Google Lighthouse accessibility score for the data portal? What are the failing elements of the site as measured by Google Lighthouse?

The provision of multiple language options enables different language speakers to make use of the data portal. For this purpose, multilingualism includes the possibility of localisation of the language. Users can access different languages on a data portal by either using language features that have been programmed into the portal, or by using Google Translate or other apps that generate a translation. Programmed language options on a data portal are typically noted by language names or flags at the top-right of the webpage. Programmed translations for sites are often more reliable than using Google Translate and so were counted as fulfilling the requirement of the guideline; Google Translate was not.

There are many ways that the accessibility of a website can be evaluated on different browsers and devices. The Web Content Accessibility Guidelines (WCAG) (Caldwell et al., 2008[21]) is the most well-known standard for accessibility. Google Lighthouse, used for this evaluation, provides a quick snapshot of many of the aspects of WCAG and does so in an automated manner, which allows easy and standardised assessment across data portals. A full accessibility analysis of each site would be its own project entirely, that could be carried out by the portal managers, if provided with sufficient resources. Google Lighthouse also analyses search engine optimisation and performance.

What do the findings show?

As shown in Figure 4.3, about 60% of portals were available in only one language, 28% offered a second language, and 10% had three or more languages available. English was, by far, the most common language available on the portals and the most likely language to be offered if only one language was available. Further, the languages that were available were often major international languages, not national or local languages.

Figure 4.3. Data portal language options: More than half of data portals are only available in one language

Source: PARIS21-ODW desk assessment of data portals

The data portals had an average Google Lighthouse score of 73.4. For context, the average Google Lighthouse accessibility score from the 2020 Web Almanac, which gathered data from 7.5 million websites, found an average Lighthouse accessibility score of 80% (Niyi-Awosusi and Tait, 2020[22]). Most of the data portals tested were in the 50-89 score range, with a small number performing on the higher and lower ends (Figure 4.4). The portals performing poorly on the accessibility measure tended to be DevInfo, CountryStat and REDATAM, while the Open SDG portals were most likely to score in the higher performing (good) category.

Figure 4.4. Google Lighthouse accessibility scores: About 90% of data portals have inadequate features to facilitate accessibility

Source: PARIS21-ODW desk assessment of data portals

The most common issues found in Google Lighthouse were missing element titles or language attributes, two types of metadata that are integrated into website code. These metadata describe the different parts of a web page so that screen reading software can read the content on the page.

What do the results imply?

The low availability of language options on the portals studied could be a barrier to access for people with a limited linguistic capacity or limited knowledge of English. The lack of smaller regional dialects available on these sites could also be a barrier of use for more rural data users and for many sub-regions in the countries studied. To "leave no one behind", priority should be placed on providing a wider range of languages so that the portals can be used by more people. Support should also be provided to NSOs to assist them in translating their content into multiple languages, as having multilingual functionality on a website is not effective if translations of the underlying content are not available.

The Google Lighthouse results suggest the presence of significant barriers to use for people with disabilities. In comparison to the countries included in the Web 2020 Web Almanac study of data portal accessibility

(Niyi-Awosusi and Tait, 2020[22]), the portals are performing below the average for accessibility across the web. This lower average score on

accessibility could be due to older portal designs (such as DevInfo) used in many countries. Updating portal architecture and design, with a special focus on improving older portals, could be an effective way to improve the accessibility of data portals. Focusing on the most common Google Lighthouse issues – missing element title and language attributes – would have a direct impact on improving the accessibility of the portals for people who are visually impaired.

Guideline 4: User-centred design

How do the guidelines describe user-centred design?

Data portals should be designed for and with users (including both operational and end users, such as data consumers or NSO officers), and project owners should engage them in all phases of development. This includes, inter alia, the analysis of user-platform interaction and the establishment of a permanent feedback loop that will result in an iterative process of continuous improvement in response to user demand. Regular collection and analysis of usage data and online user feedback should further assist in providing guidance for future modifications and enhancements. To facilitate the central role of the user, data platforms should be developed following the Agile principles (Agile Alliance, n.d.[23]) and using strategies such as design thinking.

What does the PARIS21-ODW framework measure?

Because the design process could not be evaluated by desk research substantively, the study reviewed the availability of feedback mechanisms on data portals. This includes checking that data portals had a helpline, contact us form, email address or other helpful features for use on the data portal. The evaluation specifically required that contact us information be for the data portal, not just a generic contact us link for the NSO that might be difficult to use to provide feedback specifically on the data portal. To evaluate the availability of feedback mechanisms, the study reviewed:

1. Is there a facility (chat, email, telephone or other) to give the agency feedback on its data portal?

What do the findings show?

Only about 32% of data portals reviewed have a feedback mechanism (and not just a phone number or contact information) on the platforms to receive user feedback on the design of the portal. Some portals had a phone number or contact information available, but it was not identified as a contact for feedback. Externally developed portals were more than twice as likely to have a feedback mechanism than NSO-made portals.

Figure 4.5. Presence of feedback mechanisms: only in about a third of data portals

68
NO FEEDBACK MECHANISMS

23
FEEDBACK MECHANISMS
(EXTERNALLY DEVELOPED)

9
FEEDBACK MECHANISMS
(NSO-MADE)

Source: PARIS21-ODW desk assessment of data portals

What do the results imply?

Feedback mechanisms to request help using a platform or to provide recommendations for improvements is a core element of a feedback loop with users. Providing this information in an easily accessible place on the portal is also a low-cost improvement to a portal that can be implemented without much technical capacity.

Analysing user behaviour on a portal through Google Analytics or a similar analytics software is another method that can provide more quantitative insight into the use of a data portal. Although the web analytics data were not available for this study, portal managers can use website analytics to learn who the main users of a portal are; how they use the portal (what pages they visit and what data they use); and how they arrived at the portal (from a search engine,

social media, direct links or other traffic sources). All these data from website analytics and feedback mechanisms on the portals can help NSOs respond to user demands through data portal improvements and improved dissemination efforts.

Guideline 5: Data communication

How do the guidelines describe data communication?

Data portals should implement innovative strategies to improve the presentation, communication and use of data for sustainable development. They should support multiple ways to explore, represent and communicate data on statistical indicators, and address the needs and priorities of diverse groups of users, including policy makers, legislators, civil society, the private sector, the media, the public and

academia. This includes innovative data visualisation and data storytelling capabilities. Regular collection and analysis of usage data and online user feedback should guide future modifications and enhancements.

What does the PARIS21-ODW framework measure?

To evaluate data communication by desk research, the study reviewed the following aspects of the data portal to measure its communication capability:
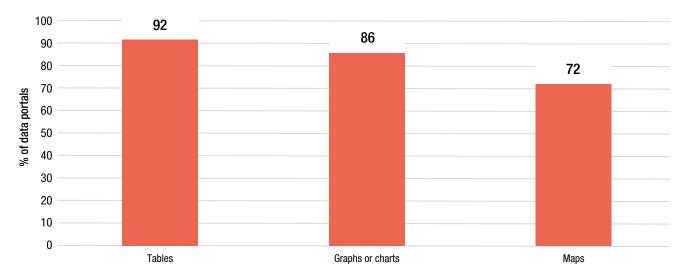
1. Are there options to filter search results by country, year or other variables?

2. What are the types of visualisation options available on the data portal: maps, charts, graphs, scatter plot, tables, others?

3. Is shareability on social media an integrated feature of the portal?

4. Does the portal allow for subscriptions?

5. Does the portal support user-created (end data consumer) web-based charts, tables and maps?

What do the findings show?

Most of the portals offered at least a tabular view of the data and the ability to create visualisations with basic visualisation options. Portals like Open Data for Africa provided many different options for visualisation through the Knoema platform.

Figure 4.6. Data visualisation: tables, graphs, charts and maps are popular options

Source: PARIS21-ODW desk assessment of data portals

The ability to share data on social media and advanced visualisation features such as maps were only available on about a third of the sites reviewed. The ability to modify the available visualisations was available on 94% of the portals. However, many of the visualisations that could be modified were just tables.

Data subscriptions (a more advanced data communication feature) were much less common and were only implemented on Mozambique's Open Data for Africa portal. The assessors searched sites for really simple syndication (RSS) feeds, which can be used to receive updates on new datasets and information posted on websites and are a type of data subscription.

What do the results imply?

Basic data communication features are available on many sites, but more advanced features are rarely implemented. A preliminary evaluation that included non-IDA countries found that their portals were more likely to implement advanced features. Providing geospatial visualisation options, like maps, might be a first step for portals to implement. This would be especially useful for understanding subnational distributions and trends.

To further increase the use, shareability and impact of the data, priority should also be given to integrating shareability on social media. This is helpful for users who want to share interesting datasets through their networks and can improve the search engine rankings for the connected portals. Implementing shareability through social media should be prioritised over implementing data subscriptions, as sharing and accessing content through social media are more popular than using data subscriptions and the RSS feeds that they employ (Lacoma, 2021[24]).

Guideline 6: Data disaggregation

How do the guidelines describe data disaggregation?

Data portals should support improved access to, and use of, disaggregated data to focus on all segments of the population, including the most vulnerable.

Data platforms should allow the management and dissemination of data disaggregated by subnational geographic areas, sex, age group, residence, wealth and income group, disability, ethnicity, migrant status, and other important characteristics.

What does the PARIS21-ODW framework measure?

To evaluate data disaggregation by desk research, the study team reviewed the following aspect of the data portal:
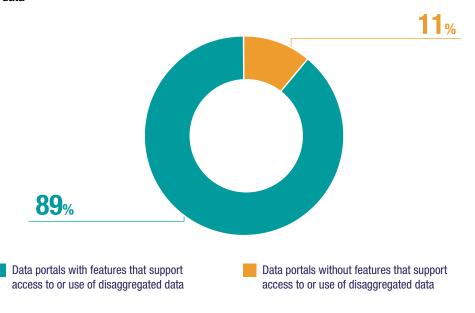
1. **Can the user view, filter, select or download data disaggregated by sex, administrative unit and other relevant characteristics?**

This evaluation only reviewed whether it was possible for the user to view, select, filter or download disaggregated data when they were available on the platform.

What do the findings show?

Eighty-nine percent of portals had the ability to view, filter or select by sex, administrative unit and other relevant characteristics. Portals that did not demonstrate this ability may not have had disaggregated data available or, despite having disaggregated data, lacked the functionality to view or select the data.

Figure 4.7. Data disaggregation: Almost 90% of data portals have features that support access to or use of disaggregated data

11%

89%

■ Data portals with features that support access to or use of disaggregated data

■ Data portals without features that support access to or use of disaggregated data

Source: PARIS21-ODW desk assessment of data portals

What do the results imply?

This is one of the brighter spots in the evaluation: most data portals' connection to their database allows to search and filter by different disaggregations and variables. A further evaluation to review the availability of disaggregated data, including gender disaggregations, would be a useful next step, because the capacity for a portal to display disaggregated data is of no use if the portal does not have disaggregated data. However, if portals can present disaggregated data, then once new disaggregated datasets become available and more data gaps are filled, these data portals will be well placed to help users search for and find the data disaggregations that they need.

Guideline 7: Modularity and extensibility

How do the guidelines describe modularity and extensibility?

Data portals should be modular, composed of modules (sub-systems) and components that interoperate to service the different phases of the data life cycle (such as the Generic Statistical Business Process Model). The data that these modules and components consume as inputs and produce as outputs should, as much as possible, be based on open standards and protocols such as Statistical Data and Metadata eXchange (SDMX). Similarly, the use of Common Statistical Production Architecture may provide a set of standard specifications for services that can be used in a modular way. The systems should support extensibility through the addition of modules or components, upstream or downstream.

What does the PARIS21-ODW framework measure?

To evaluate modularity and extensibility by desk research, the following aspects of the data portal were reviewed:

1. Does the system support major open standards for data documentation and exchange (for example SDMX)?

2. Is there an SDMX registry on the NSO data portal?

The goal of SDMX, according to the SDMX Community, is "standardising and modernising ('industrialising') the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries" (SDMX, 2020[25]). The implementation of SDMX is critical to interoperability, modularity and the functioning of data flows. The presence of SDMX artefacts on data portals signifies the potential of shared services to be created in support of modularity and extensibility of data portals.

An SDMX registry, according to Eurostat, is used to "store all the artefacts (or products) needed for those statistical information exchanges using SDMX… The registry only provides the information necessary to access and interpret the content of the exchanged statistical data and metadata sets" (Eurostat, n.d.[26]). In other words, an SDMX registry serves as a standardised endpoint for users who want to automate access to data. The presence of SDMX registries suggests the availability of SDMX artefacts and potential retrievals by automated systems.

Several other technical elements of a website could be evaluated to understand the interoperability of the different components of a website; support for SDMX can be evaluated remotely.

What do the findings show?

SDMX features appear on data portals of IDA-eligible countries only through portals like Open Data for Africa that are designed and managed by an entity other than the NSO. This points to the benefit of pooling resources to help provide countries with better functioning and more advanced data portals. Not all Open Data for Africa portals had SMDX registries and some datasets in Open Data for Africa portals did not have SDMX implemented.

Figure 4.8. Modularity and extensibility: Less than 40% of data portals support SDMX and have SDMX registries

Note: SDMX: Statistical Data and Metadata eXchange.
Source: PARIS21-ODW desk assessment of data portals

What do the results imply?

This is one of the more advanced guidelines, and it is not a surprise that many countries struggled to implement it. But implementing SDMX is critical to the ability of NSOs to model their data efficiently for data portals and for their data to be interoperable with other data systems. It should be prioritised for the long-term success of data portals. Although this evaluation did not focus on database management and internal practices for implementing standards – as these are too difficult to evaluate through desk research – the successful implementation of a data portal relies on these systems.

Further research is needed to understand the barriers to implementing SDMX along with support to develop capacity in countries to implement the standard. Working with Open Data for Africa and other portal providers to implement SDMX would help to reduce the barriers faced by many countries.

Guideline 8: Standardised interfaces

How do the guidelines describe standardised interfaces?

Data portals should provide standardised APIs in accordance with the industry's best practices, such as the OpenAPI Specification. This facilitates creating and sharing data across global, regional, national and subnational data communities.

What does the PARIS21-ODW framework measure?

To evaluate standardised interfaces by desk research, the study reviewed the following aspects of the data portal:

1. Does the data portal support APIs?
2. If the data portal supports APIs, is there documentation for the API?

What do the findings show?

Thirty-five portals supported APIs, but only two had documentation on how to use the API available. These features appear, again, on data portals for IDA-eligible countries only through portals like Open Data for Africa that are designed and managed by an external entity.
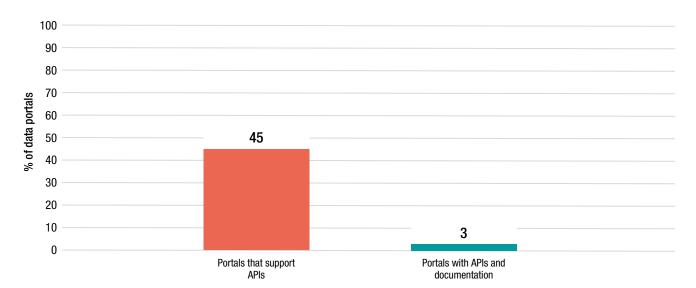
Figure 4.9. Standardised interfaces: While about half of the data portals support APIs, only 3% of portals provided documentation

Note: API: application programming interface.
Source: PARIS21-ODW desk assessment of data portals

What do the results imply?

The existence of APIs means that considerations have been made for data to be used by digital systems and other applications that consume APIs. The best practice involves developing APIs that are consistent and reusable, which can be accomplished by using an API description language to establish a contract for how the APIs are supposed to act.

The lack of instructions and documentation on how to use the APIs could be a significant barrier to use, even though a relatively large number of APIs are available. Providing documentation for existing APIs is low hanging fruit and should be a high priority for data portals to create and make this information easy to find on their websites. As this more advanced feature might be difficult to implement, one next step for the research is to evaluate the costs and benefits of implementing APIs. In the meantime, there are plenty of more basic guidelines that should be prioritised that are lower cost and provide clear benefits to data use.

Guideline 9: Scalability

How do the guidelines describe scalability?

A national reporting and dissemination platform should have an architecture that enables a statistical office to start with a limited scale implementation and iteratively progress towards a full-scale system. Tasks such as adding indicators or breakdowns should be performed directly by an operational user at the statistical office and should not require additional software development. By taking the needs and resources of different NSO departments and other national agencies into account, the design of data platforms should facilitate their adoption across the national statistical system.

What does the PARIS21-ODW framework measure?

Evaluating scalability was not possible via desk research, but the detailed methodological framework in Annex B outlines questions for a survey or interview that can help evaluate this guideline more comprehensively.

However, the research team gathered a few insights about factors that may affect the extent of scalability in the given data portals. One way to assess the design of scalable data portals is to look at whether they were deployed for a limited set of indicators with a sectoral or other strategic focus, thereby limiting their adoption and use across the national statistical system. This limitation of "scalability by design" would increase the need for data producers to design and implement multiple data portals to disseminate a specific set of indicators drawing from particular data sources, sectors, or strategic focus.

What do the findings show?

Most of the portals studied were focused on disseminating official statistics and provided a broad range of different types of data. This is largely

because the Open Data for Africa portal, the most widely used portal in this study, provides the option to disseminate official statistics, SDG data and census data. DevInfo portals developed by UNICEF also disseminate a range of statistics from different sources, sectors, and policy frameworks.

However, the research team also found portals that may have limited scalability, due to a limited focus on the types of data they disseminate. For instance, CountryStat developed by the FAO only disseminates data on food and agriculture statistics.

The second-most commonly found portals (REDATAM portals) in the study disseminate source-focused data, primarily census data in this case.

The scalability of data portals can also be limited by their policy or strategic focus. For instance, the Open SDG platforms are designed to disseminate only SDG indicator data, not other types of official statistics. On the contrary, .Stat Suite portals (such as the one implemented by Cambodia) disseminate both the SDG and national indicator data, demonstrating higher scalability.

What do the results imply?

As the research team was not able to evaluate scalability through desk research, caution should be taken when deriving insights from these findings. The design of data portals to include different types of data is one aspect of determining a portal's scalability. This approach of scalability by design should use an inclusive approach to indicator data dissemination, drawing from multiple sources, sectors and policy frameworks to enable adoption and use by multiple data producers across the national statistical system. Limited scalability by design compels NSOs to develop multiple data portals with emerging data demands and priorities.

A more in-depth evaluation of scalability could be pursued in another iteration of this study to fully evaluate the guideline; however, this would likely require a full survey of data portal managers. The ability for portals to scale according to traffic and resource demands and the implementation of scalable cloud infrastructure was also not evaluated in this research but would be a useful addition to future studies on the topic.

Guideline 10: Metadata

How do the guidelines describe metadata?

Data portals should support statistical metadata at the appropriate level of granularity. This includes structural metadata such as codes and their descriptions; reference metadata such as methodology and quality aspects of published indicators; and other relevant information such as the date of the last update.

What does the PARIS21-ODW framework measure?

To evaluate metadata by desk research, the study team reviewed the following aspects of the data portal:

1. Are all the indicators accompanied by a minimum set of metadata elements?

2. What elements of metadata are missing?

The study team checked for the presence of three metadata elements: 1) the source of the data; 2) the methodology for the data or how they were calculated; and 3) the date of the most recent entry or last update of the dataset. These three core pieces of metadata should provide the basic information for a user to understand the data for use and are consistent with the metadata evaluation of the Open Data Inventory (ODIN), an index managed by Open Data Watch on the coverage and openness of official statistics.

What do the findings show?

Four portals provided all the required elements of metadata. While all but 15 portals demonstrated the ability to present some metadata, 59 (76%) did not present all 3 of the core elements. Methodology was the element that was most often missing from the sites.
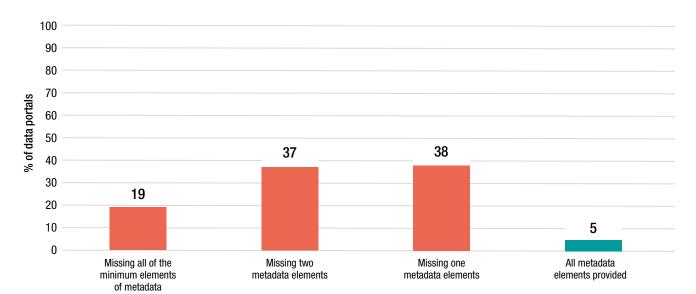
Figure 4.10. Metadata availability on data portals: only 5% of all data portals provide all metadata elements

Source: PARIS21-ODW desk assessment of data portals

What do the results imply?

Providing all three of the core metadata elements is a low-cost step to improve the usability of data and the data portal's adherence to the principles and guidelines. The start of this process is to insert fields in the data portals to display all the critical metadata elements. Once these fields have been added, the required metadata can be systematically inserted alongside the data. Further work will likely be needed to create the appropriate data flows and statistical processes to ensure that when new data are uploaded to the portals they contain the correct metadata. These metadata should be included in the underlying databases on which the data portals are built.

A metadata management system is crucial to ensuring that appropriate metadata describe how data inputs are transformed throughout the statistical production process, leading to the final statistical products that are disseminated through the data platform. In this context, the Generic Statistical Business Process Model can be used as a template for creating the metadata needed to identify and describe each step involved in the production of a dataset, from collection to dissemination.

Missing metadata can make it difficult or impossible to use data hosted on data portals. For example, without information on how the data were calculated, it can be hard to understand whether the data can help answer a research question or to compare the data with similar datasets. Providing the minimum set of metadata reviewed in this study is a low-cost improvement that requires little technical capacity

and can have a large impact on the user. In the long run, countries should consider adopting metadata standards such as SDMX that facilitate data exchange and interoperability.

Guideline 11: Open data

How do the guidelines describe open data?

Data portals should be consistent with open data best practices, summarised as "Open data and content can be freely used, modified, and shared by anyone for any purpose". National reporting and dissemination platforms should include and follow a data license consistent with the open data principles, such as Creative Commons Attribution (4.0) or the Open Database License. Published datasets should be clearly attributed to the originating organisation.

What does the PARIS21-ODW framework measure?

To evaluate open data by desk research, the following aspects of the data portal were reviewed:

1. Is there a terms of use available on the portal?

2. If there is a terms of use on the portal, does it conform to a CC0, CC-BY, CC BY-SA or other open data license?

3. Is a bulk download option available for the data?

4. What are all the available download options for datasets?

To evaluate open data and the use of open data best practices and standards, the study team used results from the ODIN evaluation and reviewed the

availability of machine-readable formats, terms of use and bulk downloads. The full ODIN evaluation on open data has five elements and includes the availability of metadata (covered in the metadata guideline). ODIN also assesses the use of non-proprietary data formats, which was out of the scope of this evaluation.
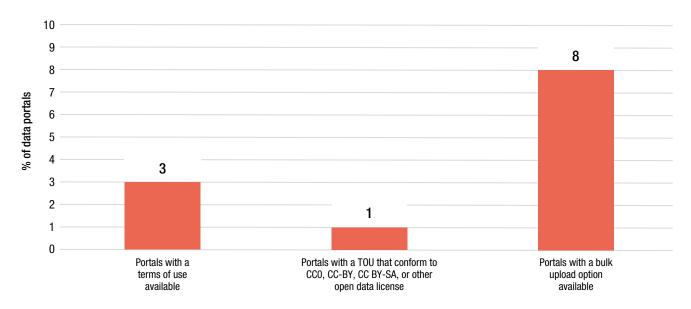
What do the findings show?

All portals studied provide data in machine-readable formats, but many fail to provide terms of use and bulk download options. Machine-readable file formats are formats that can be easily processed by a computer, such as csv, xls, xlsx or JSON. Like the ability to view data disaggregations and filter the data, the high availability of machine-readable formats is likely due to the focus of this study on data portals, which are designed to distribute these formats.

Although there were many available machine-readable download options found on the portals, the other aspects of open data were not as well implemented. Two sites had terms of use available and one was using standard Creative Commons licensing. Three sites had bulk download options available.

Figure 4.11. Open data practices: Only two data portals have a terms of use, one that conforms to a license and six with bulk download

Note:
TOU: terms of use CCO: Creative Commons Public Domain License
CC-BY: Creative Commons Attribution License
CC-BY: Creative Commons Attribution-ShareAlike License
Source: PARIS21-ODW desk assessment of data portals

What do the results imply?

Without a data license or open terms of use, data users may be reluctant to access and use the data because of concerns about legal ramifications of unapproved use. Implementing a Creative Commons license for the site might be a low-cost next step for data portals to implement as it does not require more IT development, only the development and implementation of a policy and a small addition to the website.

The lack of bulk download options is a potential barrier to use by those who are interested in downloading many datasets at once. While APIs can also be used for this purpose, they require more technical expertise and can be more difficult to use. Implementing bulk download, however, can be a complicated feature to add to a website and should probably be addressed after the issues of machine-readable file formats and terms of use.

Guideline 12: Linked data

How do the guidelines describe linked data?

Responding to the increasing demand for open data to leverage global and national investments in data for evidence-based policy and decision making, data portals should comply with a minimum of Level 3 of the 5-star Linked Open Data Principles (Berners-Lee, 2012[12]):

1. Make data available on the web (in any format), under an open license
2. Make data available as structured data (for example, Excel instead of image scan of a table)
3. Make data available in a non-proprietary open format (for example, csv instead of Excel)
4. Use unique resource identifiers to denote data items, so they can be referenced
5. Link own data to other data on the web to provide context.

What does the PARIS21-ODW framework measure?
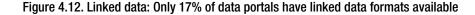
To evaluate linked data by desk research, the study team reviewed the following aspects of the data portal:
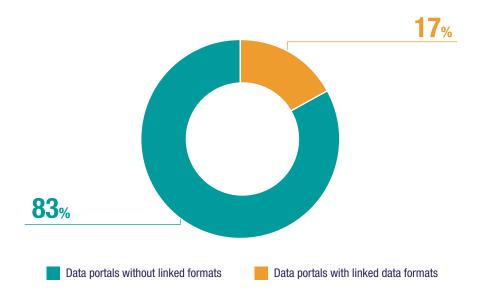
1. Does the portal use linked data formats?
2. If there are linked data formats available, can they be found in Google Dataset Search?

Linked data is an approach for publishing and sharing data on the web, using unique resource identifiers to link information resources with one another. The main implementation technology behind linked data is the general-purpose format called resource description framework. The principles and guidelines refer to the five stars of linked open data. Only the last two stars are relevant for the linked data discussion and were evaluated in this guideline, as the first three stars are covered by the open data guideline. The research team reviewed whether portals had linked data hosted on them and whether those linked data could be found on Google Dataset Search. This evaluation was done by reviewing the source code for the portals for instances of linked data using Google's structured data testing tool.

What do the findings show?

Seventeen percent of the sites had linked data formats available. Most of these are from the Open Data for Africa portals with a connection to a system called Knoema for data visualisation that has linked data formatting. Every portal that had linked data also had that linked data available on Google Dataset Search, which is evidence that Google search engines are finding the datasets that are tagged as linked data.

Figure 4.12. Linked data: Only 17% of data portals have linked data formats available

17%

83%

Data portals without linked formats
Data portals with linked data formats

What do the results imply?

As many portals struggle to provide more basic features, including a minimum set of metadata elements and terms of use, it is not surprising that the more advanced features related to linked data are not widely available. While there has been a lot of talk about linked data in the international arena, it appears that the implementation of semantic web technologies for the dissemination of statistical data is yet not widespread. As there are many other easier to implement changes needed to adhere to the principles and guidelines, these should be prioritised before incorporating linked data features. However, the fact that the linked data that were discovered were also found to be accessible through Google Dataset Search points to the potential promise of this technology. As more NSOs develop capacities to disseminate official statistics as linked data, they can make their data available through the semantic web so that third-party applications can automatically find these resources, access them and operate on them based on their relevance to different use cases.

Other findings: Uptime, website performance and search-engine optimisation

Besides reviewing the data portals against the principles and guidelines, the study team also performed a technical evaluation of the data portals to understand their uptime (percentage of time online) and how their technical features may affect use.

A low uptime was found for many of the portals in the evaluation, which could be causing issues in access and use of the data portals. Several sites were offline for much of the evaluation period or had critical errors that prevented them from loading. Forty-two portals were evaluated for their uptime. (Many of the portals could not have their uptime evaluated because of technical errors from the uptime service, Uptime Robot, the research team used; however, these errors occurred mostly on Open Data for Africa portals that have a high level of uptime and were not found to be offline during the evaluation by the assessors.) There is not a clear industry standard for the uptime of a website. High availability, or an uptime of 99% or more, is one measure. Downtime can lead to real issues for users accessing the data and, for context, even a portal with 99% uptime is down for 87 hours and 36 minutes a year (Haran, 2020[27]). Ten portals did not pass the 99% uptime score, and eight had less than 80% uptime. Open Data for Africa portals were not typically down for the evaluation, so these might have improved the average scores. A site being online is a necessary condition for data use, so uptime should be prioritised for improvement, even though it is not included in the principles and guidelines.

Figure 4.13. Portal uptime percentages: Ten portals did not pass the 99% uptime score

Source: PARIS21-ODW desk assessment of data portals

Many portals had failing Google Lighthouse scores (a score of less than 50 out of 100) for website performance. This is one of the four metrics evaluated by the tool in addition to accessibility – with only 12% of portals receiving a passing score. Failing scores indicate a failure to implement best practices for website performance across the metrics. These sites can be slow to load, which can be particularly difficult for users with slower Internet speeds. This finding was corroborated by the data assessors who reviewed the sites and at times had difficulty performing the evaluations due to a portal's slow load time. This paints a picture of a data landscape with many portals often offline and, even when they are up, performing poorly on metrics of speed.

Fifty-one percent of portals had a passing search engine optimisation score on Google Lighthouse. Sites that do not have search engine optimisation implemented might be difficult to find through Google and maybe underused, as most websites get most of their traffic from search engines.

While these findings do not correspond to any of the principles or guidelines, they are important for the usability and accessibility of the data portals. These technical issues of performance should be a high priority to fix and can have a large impact on the increasing use of the sites.

4.3. SUMMARY OF FINDINGS

Table 4.2 summarises the evaluation results for each guideline. Although the evaluations were not designed to provide passing or failing scores for each guideline, the research team used their qualitative experience of reviewing the guidelines along with the measured observance of the guideline indicators to construct the summary ratings described below.

Guidelines that had an implementation rate higher than 70% were categorised as "well implemented" and marked in green in Table 4.2; those with an implementation between 30% and 70% were categorised as "moderately implemented", marked in yellow; and those with an implementation below 30% were categorised as "poorly implemented", marked in red. These findings are subjective but can still be useful in understanding trends in guideline implementation and for prioritising resources and efforts towards those that are not well implemented.

Inadequately implemented guidelines: Less than 30% implementation

National ownership, standardised interfaces, scalability, open data and linked data all had low rates of implementation of the available indicators and should be prioritised in efforts to implement the guidelines. Although almost 45% of portals had APIs (a part of the standardised interface guideline), only 3% of these had accompanying documentation. Without this documentation, the APIs are very difficult to use, so the guideline is counted as inadequately implemented.

Moderately implemented guidelines: Between 30-70% implementation

Multilingualism and accessibility, modularity and extensibility, and metadata are mid-level performing guidelines and had indicators for implementation that were between 30% and 60%. Improving metadata implementation could be prioritised here as it is a low-cost intervention that could facilitate data use.

Extensively implemented guidelines: Greater than 70% implementation

Data communication and data disaggregation were the only two high-performing guidelines with indicators that had greater than 70% implementation in the data portals studied. Most of the data portals studied demonstrated the ability to provide disaggregated data.

Guidelines without enough information to evaluate

Collaboration, user-centred design and scalability did not have sufficient available indicators or representative measures for this exercise. Although one indicator of user-centred design was included in this study (the availability of a feedback mechanism), this was not representative enough to cover all the substantive elements of user-centred design to conclude how well the guideline is implemented. Similarly, the information found via desk research on collaboration and scalability was inadequate to determine adherence to the respective guidelines.
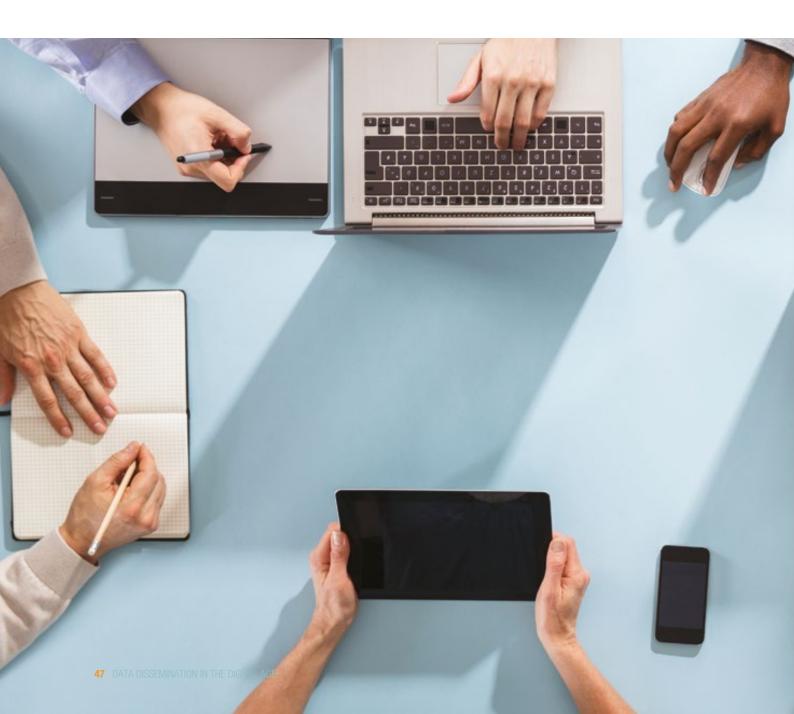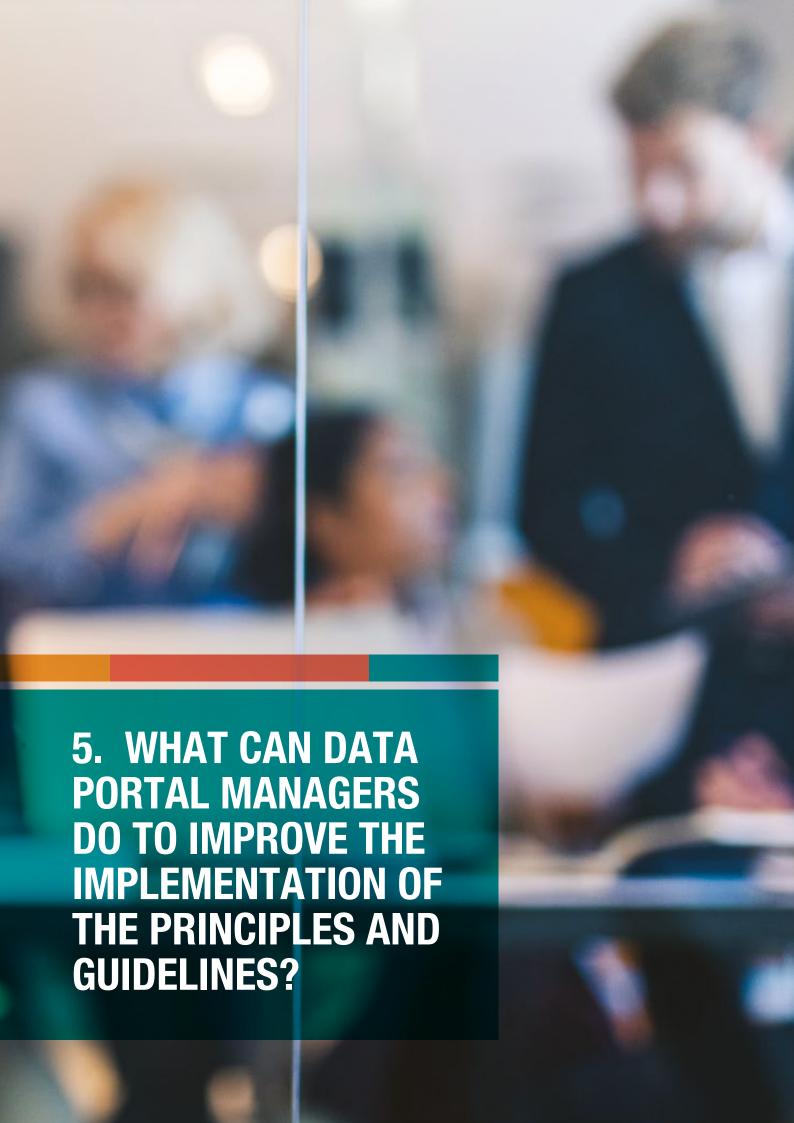
Table 4.2. Top-level findings from desk assessments of guidelines

	Guideline	Key results	Implementation rating
1	**National ownership**	Fourteen percent of portals are not externally developed; 34% of portals have an "About Us" page with the national statistical office listed as the manager.	
2	**Collaboration**	Twenty-one data portals used open-source software in their development, supporting a collaborative approach to platform design. However, only five clearly noted their use of open-source code. The research team could not evaluate the full extent of collaboration in the portals' design, development or management.	
3	**Multilingualism and accessibility**	Thirty-three percent of portals do not pass the Google Lighthouse accessibility test and 60% of portals were only available in one language.	
4	**User-centred design**	Thirty-two percent of portals have a feedback mechanism available.	
5	**Data communication**	Ninety-four percent of portals had tables; 77% graphs or charts; 73% maps; one portal had the data subscriptions feature available.	
6	**Data disaggregation**	Ninety percent of portals can disseminate disaggregated data.	
7	**Modularity and extensibility**	Thirty-eight percent of portals support SDMX; 32% have an SDMX registry.	
8	**Standardised interfaces**	Forty-five percent of portals support APIs; 3% of portals have documentation for their APIs.	
9	**Scalability**	About 60% of portals assessed appeared to disseminate data from multiple sectors, policy frameworks or sources.	
10	**Metadata**	Nineteen percent of the portals are missing all metadata; 37% are missing two metadata elements; 38% are missing one metadata element; 5% had all metadata elements required.	
11	**Open data**	Three percent had a terms of use; 1% had a terms of use that conformed to CC0, CC-BY or other open data license; 8% had a bulk download option available.	
12	**Linked data**	Seventeen percent of the portals had a linked open data format.	

Note: API: application programming interface.

■ Poorly implemented (<30%)   ■ Moderately implemented (30-70%)

■ Well implemented (>70%)   ■ No data

5. WHAT CAN DATA PORTAL MANAGERS DO TO IMPROVE THE IMPLEMENTATION OF THE PRINCIPLES AND GUIDELINES?

5. WHAT CAN DATA PORTAL MANAGERS DO TO IMPROVE THE IMPLEMENTATION OF THE PRINCIPLES AND GUIDELINES?

Many of the data portals studied have not fully implemented the principles and guidelines. A collaborative approach between all stakeholders that create, maintain and support data portals is needed to address this challenge, but NSO staff working directly on managing national data portals should have a central role. The following technical recommendations could be implemented so that data portals can better fulfil the principles and guidelines.

1. PRIORITISE A BACK-TO-BASICS APPROACH

This study demonstrates that basic functions expected of well-run data portals are not adequately implemented. Many of the standards for open data (Guideline 11), such as providing a terms of use and bulk download options, are not fully implemented. Even when more advanced recommendations of the guidelines, like the provision of APIs (Guideline 8 – Standardised Interfaces), have been implemented, the accompanying documentation for them is missing. Phasing out legacy technical and operational systems is a significant challenge, especially in a capacity-constrained context. Therefore, a pragmatic approach should be adopted in implementing the principles and guidelines, wherein features that will provide a higher return on investment are prioritised. For instance, depending on the country's needs, features such as improving metadata availability and multilingualism can be prioritised over advanced features such as linked data.

2. ADOPT USER-CENTRIC DESIGN TO ACCOUNT FOR THE NEEDS OF END USERS

User-centric design (Guideline 4) can help NSOs to better understand and respond to the needs of their end users. This is critical to promoting the use of their data and to understanding users' needs and prioritising data portal developments and data releases based on what will have the highest impact on users. The implementation of feedback mechanisms (which were only found in 50% of the portals studied) is one of the first steps to incorporate user-centric design. The use of analytics programmes to study traffic to websites is another important source of information. Leveraging these technologies along with more qualitative methods for gaining feedback, such as focus groups and interviews, can produce a holistic picture of user needs along with success stories and statistics on the importance of data portals to assist NSOs in lobbying for increased funding.

3. ADVANCE NATIONAL OWNERSHIP OF DATA PORTALS

With 85% of data portals created by external actors, they should be well placed to implement feature updates to better align their data portals with the principles and guidelines. Because most NSOs in IDA-eligible countries do not control all aspects of their country's data portals, the responsibility for change rests primarily on those who design or manage the portals, but NSOs and other agencies of the national statistical system must be involved. Development partners should support workshops, trainings and regular consultations to bring relevant country actors together to assess priorities and address the issues highlighted in the study. And they should co-ordinate their work, putting data standards at the heart of their initiatives so that they complement each other and support countries' national strategies for the development of their statistical systems.

4. IMPROVE UPSTREAM DATA MANAGEMENT PRACTICES FOR A SUSTAINABLE DATA DISSEMINATION INFRASTRUCTURE

A well-designed data portal is part of the larger data dissemination infrastructure. Long-term sustainability of a data portal rests on robust data management practices, including an overarching digital process orientation of NSOs: for example, data modelling that propels a shift from paper to digital methods. Although a complete study of data flows in NSOs is outside the scope of this report, without intentionally designed data flows, NSOs will continue to have problems with their digital processes, such as regular updating and maintenance of their portals. PARIS21's Data Flow Analysis Framework (PARIS21, 2020[28]) provides guidance for NSOs to holistically analyse and improve their data flows that will eventually help improve the management of their data portals.

5. STREAMLINE DATA DISSEMINATION PROCESSES TO REDUCE THE REPORTING AND MANAGEMENT BURDEN FOR MAINTAINING DATA PORTALS

Strategies for creating all-in-one (yet modular) data portals or streamlining data dissemination processes for the NSO could reduce the reporting and management burden on NSOs. Some countries don't have any portals, but others have three or four that may have duplicate contents and functions. If the portals are not connected to their data sources by well-designed data flows, which the low level of implementation of SDMX found in this study suggests, the NSO and other agencies within the national statistical system may find themselves reporting separately to the different portals, increasing their burden. Portal design should be streamlined and the proliferation of portals curtailed to minimise the burden on data providers and ensure that users do not have to search many portals to find the data they need. Co-ordination, collaboration and communities of practice are critical to addressing these challenges. The NSO should be at the centre of these collaborations and should lead the efforts in streamlining data dissemination processes.

A more holistic view of data dissemination is needed to realise a lasting return on NSO investments in IT tools such as data portals. More attention needs to be given to digitalising the data flows to data portals and on improving data portals to increase data use and impact. As data portals become more common, statistical offices should shift the focus from digitalising isolated components or ad hoc interfaces in the statistical infrastructure to a systems approach for transforming data dissemination. This is aligned with emerging approaches to statistical capacity development that go beyond the traditional production-side interventions to include the strengthening of data use and impact. Effective, efficient and sustainable statistical capacity development programmes, as defined by the Capacity Development 4.0 framework, cannot be restricted to one or a few isolated capabilities (PARIS21, 2020[28]).

6. WHAT POLICY AND RESEARCH INITIATIVES ARE NEEDED TO SUPPORT THE IMPLEMENTATION OF THE PRINCIPLES AND GUIDELINES?

6. WHAT POLICY AND RESEARCH INITIATIVES ARE NEEDED TO SUPPORT THE IMPLEMENTATION OF THE PRINCIPLES AND GUIDELINES?

As the development data community turns its attention farther along the data value chain to focus on data use and impact, effective data dissemination strategies and data portals play an important role. The principles and guidelines can serve as a helpful framework for improving these dissemination strategies and encouraging data use. This report presents an initial account of the state of data portals in IDA-eligible countries and presents recommendations to enable a systems approach towards data portal design, deployment and sustainability. However, this is only a starting point for further research and advocacy in this area. Section 5 covered technical recommendations for data portal managers. This final section concludes with broader recommendations on how the development data community can implement new policies and research to support the implementation of the principles and guidelines and the broader use of open data. The following next steps outline how we can achieve these goals and how we can continue to build on the findings of this study.

1. DEVELOP A COMMUNITY OF PRACTICE TO SHARE INFORMATION

The principles and guidelines reflect the global development data community's beliefs about what would improve the usability of data portals (and, by extension, use of their data), but more research to investigate these assumptions is needed. A community of practice to self-monitor the implementation of data portals could spur peer exchange for further improvement of the principles and guidelines as well.

One way to do this is to develop a "self-assessment toolkit" and related repository, based on the

PARIS21-ODW methodological framework. NSOs can use the tools employed here for the desk research to evaluate their data portals against others. The survey questions can, in turn, be used for a qualitative self-assessment which, if shared with other statistical agencies, can provide useful models for data portal governance. Further, as countries improve their data portals, installing and monitoring web analytics can provide information on how these changes facilitate data access. This information could also be used to prioritise the improvements that have the greatest impact on data use.

2. STRENGTHEN EVIDENCE ON HOW WELL DATA PORTALS WORK

This study can be used as a baseline for evaluating the implementation of a majority of the UNSD principles and guidelines, but further evidence is necessary to fill in the gaps. More empirical research, using surveys, interviews or self-assessments, is required to fill the current gaps of knowledge. Here are some outstanding tasks.

▶ How to measure scalability? Guideline 9 on scalability could not be fully evaluated in this study because many aspects of its implementation are only available through surveys or interviews with portal managers. Further research could evaluate the ability of portals to incorporate new indicators and domains to scale the platform according to the changing needs of the NSO.

▶ How are portals implementing data standards? This study largely focused on the availability of features on platforms, not on how platforms are implementing many of those features, such

as SDMX. Future studies could take a deeper dive into API implementation, SDMX and other metadata schemas, and additional data portal features to evaluate how they are implementing the corresponding standards.

▶ How is the architecture of data portals changing? In the past, statistical data platforms often relied on a monolithic architecture in which a tightly coupled software system governed all elements of data management, from database access to delivering data tables and visualisations to the client. However, the spread of new devices and applications has encouraged statistical organisations to implement modern server-to-server data exchange and dissemination technologies, flexibly serving the data from multiple back-end systems to an array of customised front-end applications. This includes, but goes beyond, SDMX to include aspects of modularity and extensibility that could be evaluated in future studies.

3. CLOSE THE LAST-MILE GAP BETWEEN DATA DISSEMINATION AND USE

Data portals are one part of digital dissemination architectures that can stimulate data use. To really enable the uptake and use of data and statistics, we need to close the last-mile gap between data dissemination and use. This presents scope for further exploration of the barriers and levers between data production and data use. Factors affecting data providers such as business processes and back-end digital data flows (Ranjan and Challener, 2021[29]), and users such as data literacy (Misra, 2021[30]), merit further analysis.

4. RESEARCH OTHER MODES OF SUSTAINABLE DATA DISSEMINATION

This study employed a limited definition of a particular digital data dissemination modality: data portals. However, data dissemination channels are wide-ranging, including online applications, dashboards and SMSs. Further, as the study reflected, a significant number of NSOs do not have data portals and use data websites to disseminate statistics. For a broader understanding of data access, dissemination and use, an examination of these alternate digital channels is warranted.

5. UPDATE THE PRINCIPLES AND GUIDELINES TO INCLUDE OTHER FEATURES OF DATA PORTALS THAT AFFECT DATA USE

The principles and guidelines are an important foundational step toward improving data dissemination, but they need to be updated to encompass other data portal features that affect data use and to provide more instruction on the implementation of the guidelines. This research provides a first step towards operationalising the principles and guidelines so that they provide actionable recommendations for countries to improve their portals and adhere to the guidelines. These recommendations also enable the international community to better monitor progress towards the implementation of the principles and guidelines, with this study providing a baseline. On a broader level, there are other aspects of data portal design that affect data use that should be considered for inclusion in the principles and guidelines, either through new guidelines or to be included as implementation steps for existing ones:

▶ **Findability:** Data are only valuable if they can be easily found and, consequently, used. Data findability is, therefore, a critical factor in data use. Findability can be improved by search engine optimisation, active dissemination efforts and limiting the proliferation of redundant portals.

▶ **Manageability:** Maintaining multiple portals with different designs adds to the management burden for NSOs. The NSO should take care to streamline dissemination efforts through the smallest number of data portals that meet the needs of a diverse set of users.

▶ **Technical functionality:** Data portals that are often offline or slow to load discourage use. These factors are especially critical in countries with poor access to the Internet.

▶ **Stakeholders:** data users and data providers should consider revisions to the principles and guidelines to incorporate practical steps that provide methods of tracking progress toward their implementation. This study will hopefully launch a dialogue on these improvements.

Realising returns from data and statistics will require an expanded and inclusive conversation with stakeholders in the data ecosystem, particularly with a focus on country ownership. Catalysing effective

data dissemination and use requires the feedback and work of many partners. Co-ordination and communication between them are critical to ensure that digital data ecosystems work for the needs of data producers, managers and users.

It has been a few years since the creation of the principles and guidelines. It is worth continuing the conversation so that the global community can reflect on their use, applicability and potential improvements. Just as user feedback cycles are essential for developing and managing data portals, they are also crucial for further development and implementation of the principles and guidelines. We look forward to this dialogue and action.

REFERENCES

Agile Alliance (n.d.), "12 principles behind the Agile manifesto", web page, Agile Allaince. https://www.agilealliance.org/agile101/12-principles-behind-the-agile-manifesto [23]

Bello, O. et al. (2016), "Open data portals in *Africa: An analysis of open government data initiatives"*, *African Journal of Library, Archives and Information Science*, Vol. 26/2, pp. 97-106. https://www.ajol.info/index.php/ajlais/article/view/167427 [11]

Berends, J. (2020), *Analytical Report 5: Barriers in Working with Open Data*, Publications Office of the European Union, Luxembourg. https://data.europa.eu/sites/default/files/edp_analytical_report_n5_-_barriers_in_open_data.pdf [14]

Berners-Lee, T. (2012), "5 star open data", web page. https://5stardata.info/en [12]

Blank, M. (2019), *Open Data Maturity Report 2019*, European Commission, Brussels. https://data.europa.eu/sites/default/files/open_data_maturity_report_2019.pdf [15]

Caldwell, B. et al. (2008), "Web Content Accessibility Guidelines", web page, W3C Web Accessibility Initiative, Caldwell, B. et al. (2008), "Web Content Accessibility Guidelines", web page, W3C Web Accessibility Initiative, https://www.w3.org/WAI/standards-guidelines/wcag [21]

Creative Commons (n.d.), *About the Licenses* (webpage, accessed 22.10.21). https://creativecommons.org/licenses/ [32]

Encyclopedia Britannica (2021), *Website* (accessed 22.10.21). https://www.britannica.com/technology/RSS [37]

Eurostat (n.d.), "SDMX Registry", web page, Eurostat. https://ec.europa.eu/eurostat/web/sdmx-infospace/sdmx-explained/who-and-how/registry#~:text=SDMX%20registries%20are%20used%20by,statistical%20information%20exchanges%20using%20SDMX.&text=The%20registry%20only%2provides%20-the,statistical%20data%20and%20m [26]

Greenwell, G. et al. (2016), "Making data portals work for SDGs: A view on deployment, design and technology", Discussion Paper No. 8, PARIS21, Paris**.** https://paris21.org/sites/default/files/Paper_on_Data_Portals%20wcover_WEB.pdf [5]

Haran, N. (2020), "How to reach 99.99% uptime: High availability in practice", *Fyipe blog*. https://blog.fyipe.com/high-availability-in-practice-how-to-reach-99-999-uptime [27]

Hidalgo, C. (2016), "What's wrong with open-data sites – and how we can fix them"**,** *Scientific American Guest Blog.* https://blogs.scientificamerican.com/guest-blog/what-s-wrong-with-open-data-sites-and-how-we-can-fix-them [7]

IBM, C. (2020), *What is an Application Porgramming Interface (API)* https://www.ibm.com/cloud/learn/api [31]

Islam, A. and D. Lederman (2020), "Data transparency and long-run growth", *Policy Research Working Paper*, No. 9493, World Bank, Washington, DC https://openknowledge.worldbank.org/handle/10986/34915 [9]

Kubler, S. et al. (2018), "Comparison of metadata quality in open data portals using the Analytic Hierarchy Process", *Government Information Quarterly*, Vol. 35/1, pp. 13-29. https://doi.org/10.1016/j.giq.2017.11.003 [10]

Kumar, R. et al. (2020), *Assessing the Effectiveness of Data Sites in Nepal*, World Bank, Washington, DC https://documents.worldbank.org/en/publication/documents-reports/documentdetail/960741596604922376/assessing-the-effectiveness-of-data-sites-in-nepal [16]

Lacoma, T. (2021), "What is an RSS feed", web page, Digital Trends, https://www.digitaltrends.com/computing/what-is-an-rss-feed [24]]

Misra, A. (2021), *Advancing Data Literacy in the Post-pandemic World: A Primer to Catalyse Policy Dialogue and Action*, Partnership in Statistics for Development in the 21st Century, Paris, https://paris21.org/sites/default/files/inline-files/DataLiteracy_Primer_0.pdf [30]

Neumaier, S. and J. Umbrich (2016), "Automated quality assessment of metadata across open data portals", *Journal of Data and Information Quality*, Vol. 8/1, http://dx.doi.org/10.1145/2964909 [18]

Niyi-Awosusi, O. and A. Tait (2020), "Part II Chapter 8: Accessibility", *Web Almanac*, web page, HTTP Archive, https://almanac.httparchive.org/en/2020/accessibility [22]

Norman, D. A. (1986), User-Centered System Design: New Perspectives on Human-Computer Interaction

OECD (2017), *Development Co-operation Report 2017: Data for Development*, OECD Publishing, Paris, https://doi.org/10.1787/dcr-2017-en [3]

Open Data Watch (2021), "Data support", web page, Open Data Watch, https://opendatawatch.com/data-support (accessed on 2021). [20]

Open Data Watch (2018), *The Data Value Chain: Moving from Production to Impact*, Data2X, Washington, DC, https://opendatawatch.com/publications/the-data-value-chain-moving-from-production-to-impact [6]

Open Knowledge Foundation (n.d.), "Open definition 2.1 web page (accessed 22.10.21)", Open Knowledge Foundation, https://opendefinition.org/od/2.1/en [34]

opensource.com (2021), *What is open source? (accessed 22.10.21),* web page, opensource.com, https://opensource.com/resources/what-open-source (accessed on 22 June 2021) [36]

PARIS21 (2020), *Guidelines for Developing Statistical Capacity:* A *Roadmap for Capacity Development 4.0,* Partnership in Statistics for Development in the 21st Century, Paris, https://paris21.org/sites/default/files/inline-files/UNV003_Guidelines%20for%20Capacity%20Development%20WEB_0.pdf [28]

Ranjan, R. and J. Challener (2021), *Data Flow Analysis Framework: Guidelines for Analysing Data Flows in National Statistical Offices,* Partnership in Statistics for Development in the 21st Century, Paris, https://paris21.org/sites/default/files/inline-files/DFAF_FINAL_WEB.pdf [29]

SDMX (2020), *About SDMX (accessed 22.10.21),* https://sdmx.org/?page_id=2561 [38]

SDMX (2020), "What is SDMX", web page, SDMX, https://sdmx.org/?page_id=3425 [25]

Steinberg, M. and D. Castro (2017), "The state of open data portals in Latin America", web page, Center for Data Innovation, https://datainnovation.org/2017/07/the-state-of-open-data-portals-in-latin-america [13]

Ubaldi, B. (2013), "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives", *OECD Working Papers on Public Governance*, No. 22, OECD Publishing, Paris, https://dx.doi.org/10.1787/5k46bj4f03s7-en [35]

UNSD (2021), *4th Edition of the Handbook of Statistical Organization*, United Nations Statistical Division, https://unstats.un.org/wiki/display/HSO/Full+Handbook+-+Beta+v2 [19]

UNSD (2019), *Principles of SDG Indicator Reporting and Dissemination Platforms and Guidelines for Their Application*, United Nations Statistical Division, Department of Economic and Social Affairs, New York, NY, https://unstats.un.org/unsd/statcom/50th-session/documents/BG-Item3a-Principles-guidelines-SDG-Monitoring-Reporting-Platforms-E.pdf [1]

Vale, S. (2021), *High-Level Group for the Modernisation of Statistics (webpage, accessed 22.10.21),* https://statswiki.unece.org/display/hlgbas/High-Level+Group+for+the+Modernisation+of+Official+Statistics (accessed on April 2021) [2]

W3 (2004), *Architecture of the World Wide Web*, Volume One, https://www.w3.org/TR/webarch/ [39]

World Bank (2021), "Borrowing countries", web page, World Bank, Washington, DC, https://ida.worldbank.org/about/borrowing-countries [33]

World Bank (2021), *World Development Report 2021: Data for Better Lives*, World Bank, Washington, DC, https://www.worldbank.org/en/publication/wdr2021 [8]

World Bank (2014), *Open Data Challenges and Opportunities for National Statistical Offices*, World Bank, Washington, DC, https://openknowledge.worldbank.org/handle/10986/19984 [4]

Zhu, X. and M. Freeman (2018), "An evaluation of US municipal open data portals: A user interaction framework", *Journal of the Association for Information Science and Technology*, Vol. 70/1, pp. 27-37, https://doi.org/10.1002/asi.24081 [17]

ADDITIONAL READING

Banker, D. (2020), "Findability, accessibility, and usability of data portals in education", *Journal of Education & Social Policy*, Vol. 7/1, pp. 8-15, http://dx.doi.org/10.30845/jesp

Chatfield, A.T. and C.G. Reddick (2017), "A longitudinal cross-sector analysis of open data portal service capability: The case of Australian local governments", *Government Information Quarterly*, Vol. 34/2, pp. 231 243, https://doi.org/10.1016/j.giq.2017.02.004
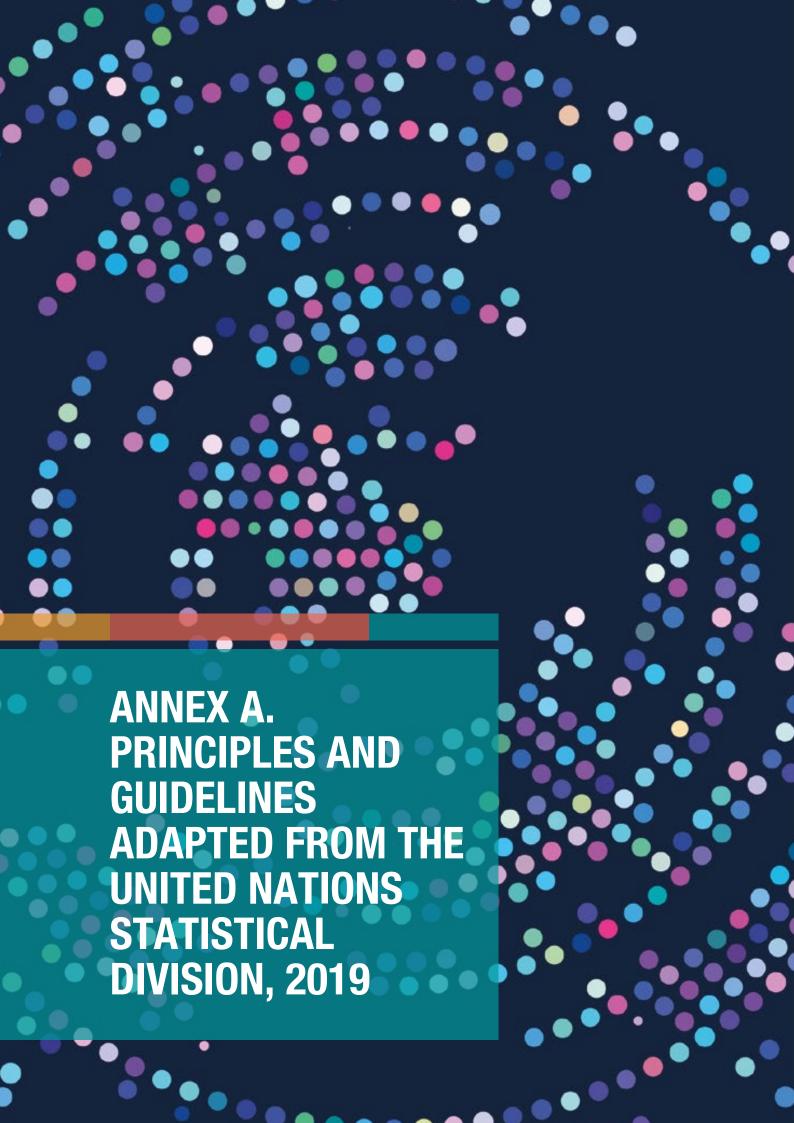
Dahbi, K.Y., H. Lamharhar and D. Chiadmi (2019), "Toward a user-centered approach to enhance data discoverability on open government data portals", Third International Conference on Intelligent Computing in Data Sciences (ICDS) 1-5

Lněnička, M. (2015), "An in-depth analysis of open data portals as an emerging public e-service", *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, Vol. 9/2, pp. 589-599

Lourenço, R.P. (2015), "An analysis of open government portals: A perspective of transparency for accountability", *Government Information Quarterly*, Vol. 32/3, pp. 323-332, https://doi.org/10.1016/j.giq.2015.05.006
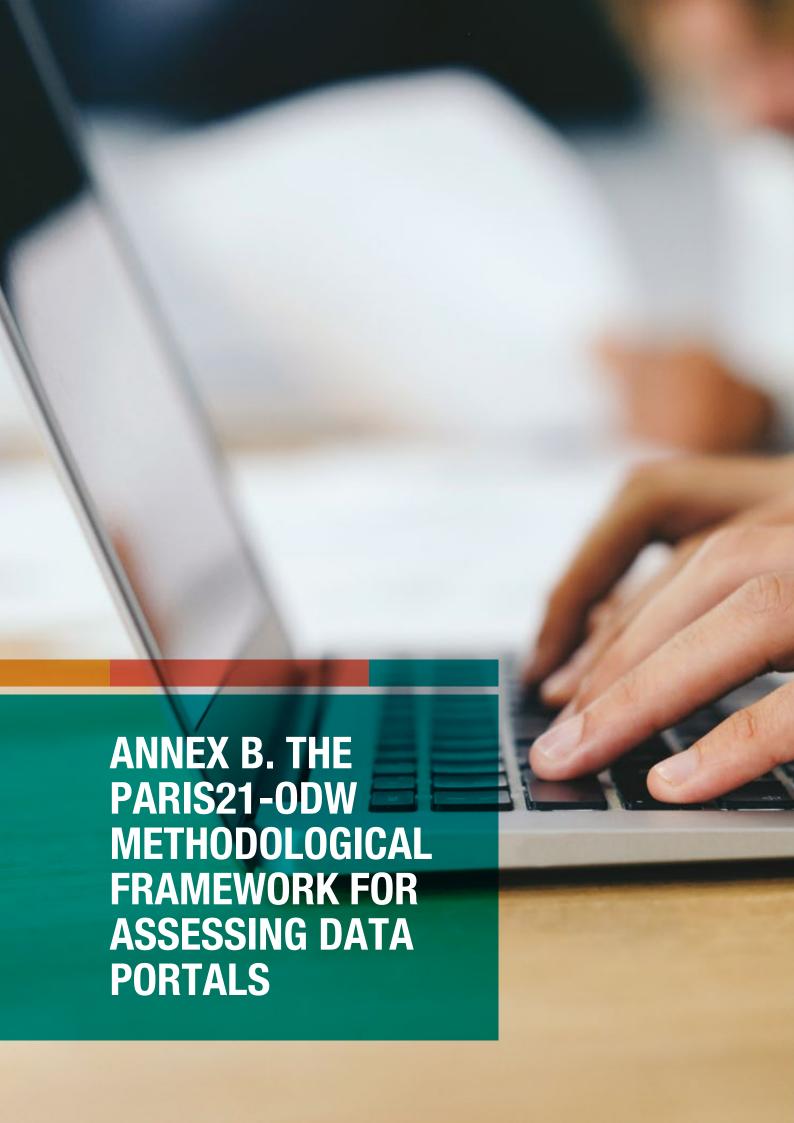
Lourenço, R.P. (2013), "Open government portals assessment: A transparency for accountability perspective", *Lecture Notes in Computer Science*, pp. 62-74, https://link.springer.com/content/pdf/10.1007%2F978-3-642-40358-3_6.pdf

Neeta, V. and M.P. Gupta (2013), "Open government data: Beyond policy & portal, a study in Indian context", *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance*, pp. 338 341, https://doi.org/10.1145/2591888.2591949

ANNEX A. PRINCIPLES AND GUIDELINES ADAPTED FROM THE UNITED NATIONS STATISTICAL DIVISION, 2019

ANNEX A. PRINCIPLES AND GUIDELINES ADAPTED FROM THE UNITED NATIONS STATISTICAL DIVISION, 2019

Guidelines	Principles			
	1	2	3	4
	Clear institutional arrangements and management	Fit for purpose	Sustainability	Interoperability and statistical standards
National ownership	X	X	X	
Collaboration	X	X	X	
Multilingualism and accessibility		X		
User-centred design		X	X	
Data communication	X	X	X	
Data disaggregation		X		
Modularity and extensibility		X	X	X
Standardised interfaces			X	X
Scalability		X	X	
Metadata		X		
Open data		X	X	
Linked data		X	X	

ANNEX B. THE PARIS21-ODW METHODOLOGICAL FRAMEWORK FOR ASSESSING DATA PORTALS

ANNEX B. THE PARIS21-ODW METHODOLOGICAL FRAMEWORK FOR ASSESSING DATA PORTALS

Assessment modality	Assessment questions
Guideline 1: National ownership National statistical offices should have the ability to maintain, adapt, transform, and customise their data portals to address their own and their users' needs, such as the management of subnational administrative boundaries, country-specific ethnic and language groups, and additional indicator definitions related to national development priorities.	
Desk research	Does the "About us" or another page provide information on who hosts, manages or maintains the site? If so, is the national statistical office (NSO) listed as the managing organisation?
Desk research	If there is an "About us" page, is the NSO listed as the managing organisation?
Desk research	Does the portal take you to another international organisation's portal?
Interview/survey	Who is the main data provider for this portal? (multiple options allowed) NSO Line ministries, departments or other government agencies International organisation or development agency Other (please specify)
Interview/survey	Does the data portal allow data entry by multiple users/organisations? No Yes, it allows for multiple users but from single location Yes, it allows for multiple users and from different locations I don't know Other (please explain)
Interview/survey	How many distinct organisations (NSO, line ministries, government, or international agencies, etc.) enter data into this portal directly? 1 1-5 >5 I don't know

Assessment modality	Assessment questions
Interview/survey	**What is the name of the main organisation(s) responsible for the overall input and maintenance of data in this portal?** NSO Other (please specify)
Interview/survey	**What is the name the main organisation(s) responsible for hosting and managing the data portal itself?** NSO Other (please specify)

Guideline 2: Collaboration

Data portals should be designed, developed, improved and maintained on the basis of a collaborative approach that leverages learning between various stakeholders of the national statistical system as well as technology developers, donors, policy makers, subject-matter experts, business partners, advocacy groups, and both institutional and grassroots users.

Assessment modality	Assessment questions
Interview/survey	**Who were the key collaborators in the following stages of the data portal development? Please indicate them under different modes of engagement from the drop-down list.** 1) design 2) software development 3) software quality assurance 4) hosting 4) maintenance and updates **Modes of engagement** Budget support/funding Technical assistance/knowledge transfer **Collaborators** Line ministries/other government agencies Policy makers/planners Donors/development agencies Independent consultants/experts Technology developers/businesses Advocacy groups Civil society organisations Other

Guideline 3: Multilingualism and accessibility

In order to leave no one behind, ensure national ownership and promote the use and impact of data for policy and decision making at the local level, data portals should support national languages and implement national and international best practices in terms of accessibility to persons with disabilities, as well as full access across the range of browsers and devices, including mobile devices.

Assessment modality	Assessment questions
Desk research	List all the language translations that the data portal has available
Desk research	What is the Google Lighthouse accessibility score for the site?
Desk research	List the failing elements on the site from the Google Lighthouse score?

Assessment modality	Assessment questions
Guideline 4: User-centred design Data portals should be designed for and with users (including both operational and end users, such as data consumers or NSO officers), and project owners should engage them in all phases of development. This includes, inter alia, the analysis of user-platform interaction and the establishment of a permanent feedback loop that will result in an iterative process of continuous improvement in response to user demand. Regular collection and analysis of usage data and online user feedback should further assist in providing guidance to future modifications and enhancements. To facilitate the central role of the user, data platforms should be developed following the Agile principles and using strategies such as design thinking.	
Desk research	(for end user/data consumer) Is there a facility (chat, email, telephone or other) to give the agency feedback on its data portal?
Interview/survey	**Are the data on the portal entered:** Manually (including bulk upload) Automatically from an underlying database Via APIs (application programming interface) Via other means (please specify)
Interview/survey	**How easy or difficult is the portal to maintain and update?** Very easy Easy Difficult Very difficult
Interview/survey	What are the biggest challenges you face in maintaining and updating the portal?
Interview/survey	Does the portal allow for scheduling data releases? (embargo feature)
Interview/survey	**Do you use a web analytics tool on the data portal to record or analyse website traffic?** Yes No I don't know **If you answered yes to the above question, please name the tool(s):** Google Analytics Adobe Analytics Yandex metrica HotJar In-house tools Other (please specify) **If you answered yes above, how often are the findings from the web analytics used to change the content or design of the data portal?** Quite often Sometimes Rarely Not at all I don't know

Assessment modality	Assessment questions
Interview/survey	**How many dataset downloads does your web portal record in a month?** We don't have that information Less than 250 downloads Between 250 and 500 Between 500 and 1 000 Between 1 000 and 5 000 Between 5 000 and 10 000 More than 10 000
Interview/survey	**Who are the most frequent users of your data portal?** Government employees Researchers/academics NGOs/civil society organisation employees Private businesses Citizens Other (please specify)
Interview/survey	**How do you gather feedback from users? (mark all that apply)** Comments and feedback from the website's contact us page User surveys Focus groups and user interviews Other tools or methods (please specify)

Guideline 5: Data communication

Data portals should implement innovative strategies to improve the presentation, communication and use of data for sustainable development. They should support multiple ways to explore, represent and communicate data on statistical indicators, and address the needs and priorities of diverse groups of users, including policy makers, legislators, civil society, the private sector, the media, the public and academia. This includes innovative data visualisation and data storytelling capabilities.

Assessment modality	Assessment questions
Desk research	Are there options to filter search results by country, year or other variables?
Desk research	Please list all the visualisation options available on the data portal: maps, charts, graphs, scatter plot, tables, other
Desk research	Is shareability on social media an integrated feature of the platform?
Desk research	Does the portal allow for subscriptions?
Desk research	Does the system support user-created (end data consumer) web-based charts, tables, and maps?

Guideline 6: Data disaggregation

Data portals should support improved access to, and use of, disaggregated data to focus on all segments of the population, including the most vulnerable. In particular, data portals should allow the management and dissemination of data disaggregated by subnational geographic areas, sex, age group, residence, wealth and income group, disability, ethnicity, migrant status, and other important characteristics relevant to national context.

Assessment modality	Assessment questions
Desk research	Can the user view/filter/select/download data disaggregated by sex, administrative unit and other relevant characteristics?

Guideline 7: Modularity and extensibility

Data portals should be modular, composed of modules (sub-systems) and components that interoperate to service the different phases of the data life cycle. The data that these modules and components consume as inputs and produce as outputs should as much as possible be based on open standards and protocols such as Statistical Data and Metadata Exchange (SDMX) and Common Statistical Production Architecture. The system should support extensibility through the addition of modules or components, upstream or downstream.

Desk research	Does the system support major open standards for data documentation and exchange (for example, SDMX [aggregate data standard], DDI [microdata documentation])?
Desk research	Is there an SDMX registry on the NSO data portal for dissemination?
Interview/survey	**What interoperable software components does your data portal have? (multiple options allowed)** SDMX APIs I don't know Other standards (please specify)

Guideline 8: Standardised interfaces

Data portals should provide standardised APIs in accordance with the industry's best practices such as the OpenAPI Specification. This facilitates creating and sharing data across global, regional, national and subnational data communities.

Desk research	Does the data portal support APIs?
Desk research	If the data portal supports APIs, is there documentation for the API?

Guideline 9: Scalability

Data portals should have an architecture that enables a statistical office to start with a limited scale implementation and iteratively progress towards a full-scale system. Tasks such as adding indicators or breakdowns should be performed directly by operational user at the statistical office and should not require additional software development. By taking into account the needs and resources of different NSO departments and other national agencies, the design of data platforms should facilitate their adoption across the national statistical system.

Interview/survey	**Does the data portal contain data pertaining only to a specific sector (for example, agriculture)?** No (multiple sectors) I don't know Yes (only one sector), please specify

Assessment modality	Assessment questions
	If you answered yes above, can the data portal also host data pertaining to other sectors? Yes No I don't know
Interview/survey	Does the data portal contain data pertaining only to a specific monitoring framework (for example, Sustainable Development Goals or country's national development plan)? No (multiple monitoring frameworks) I don't know Yes (only one monitoring framework), please specify **If you answered yes above, can the data portal also host data pertaining to other monitoring frameworks?** Yes No I don't know
Interview/survey	Is the data portal meant for disseminating indicators from: A single source like a survey or a census Multiple sources I don't know
Interview/survey	How adequate are the resources allocated for maintaining and updating the data portal? All that are needed Most of what is needed Less than needed None at all Any comments?
Interview/survey	Please rate the skill sets available among staff managing the data portal (levels: minimal, basic, intermediate, advanced) Database management system (DBMS) Data modelling SDMX API Other (please specify)
Interview/survey	What kinds of additional competencies would you need to better manage and update the data portal? Please provide examples. Technical competencies Managerial competencies Other (please elaborate)

Assessment modality	Assessment questions
Guideline 10: Metadata Data portals should support statistical metadata at the appropriate level of granularity. This includes structural metadata such as codes and their descriptions; reference metadata such as methodology and quality aspects of published indicators; and other relevant information, such as the date of last update.	
Desk research	Are all the indicators accompanied by the minimum metadata?
Desk research	What elements of metadata are missing?
Guideline 11: Open data Data portals should be consistent with open data best practices, summarised as "Open data and content can be freely used, modified, and shared by anyone for any purpose". Data portals should include and follow a data license consistent with the Open Data principles, such as Creative Commons Attribution (4.0) or the Open Database License. Published datasets should be clearly attributed to the originating organisation.	
Desk research	Is there a terms of use available on the portal?
Desk research	If there is a terms of use on the portal, does it conform to a CC0, CC-BY, CC BY-SA or other open data license?
Desk research	Is a bulk download option available for the data?
Desk research	What are all the available download options for datasets? Please list all.
Guideline 12: Linked data Responding to the increasing demand for open data to leverage global and national investments in data for evidence-based policy and decision making, data portals should comply with a minimum of Level 3 of the 5-star Linked Open Data Principles: 1. Make data available on the web (in any format), under an open license. 2. Make data available as structured data (for example, Excel instead of image scan of a table). 3. Make data available in a non-proprietary open format (for example, csv instead of Excel). 4. Use URIs to denote data items, so they can be referenced. 5. Link own data to other data on the web, to provide context.	
Desk research	Does the portal use linked data formats?
Desk research	If there are linked data formats available, can they be found in Google Dataset Search?

Note: URIs: Universal Resource Identifier

GLOSSARY

Application programming interface (API): A software tool that transfers information between computer systems. An API can be used to retrieve data from one computer or web-based application and transfer it to another application. (IBM, 2020[31])

CC0, CC-BY, CC BY-SA: Creative Commons licenses. CC0 is a public domain dedication; CC-BY permits free use and reuse if attribution to the source is provided; and CC BY-SA imposes a further restriction that the same license be applied to subsequent uses of the data. (Creative Commons[32])

Data portal: A web-based, interactive data and metadata platform with databases modelled for specific data types such as microdata, macrodata, or geospatial data or topical domains such as agriculture, health and nutrition, or the Sustainable Development Goals (UNSD, 2021[19]).

IDA-eligible countries: Countries listed by the World Bank as eligible for concessional lending or grants from the International Development Association (IDA), including "blend" countries that are also eligible for non concessional lending under terms of the International Bank for Reconstruction and Development. This study uses the list of IDA-eligible countries from 1 July 2020 (World Bank, 2021[33]).

Indicator data: Aggregate data compiled as indicators, sometimes called macrodata or statistical data. Indicators may take many forms: sums, shares or rates of change. Distinguished from microdata, which are the unit record data of individuals (or other entities) collected through census, survey, or administrative records. Also distinguished from geospatial data that include co-ordinates of geographic points or areas.

Open data: There is general agreement on the core meaning of open data. As summarised in the Open Definition, Version 2.1, "Knowledge is open if anyone is free to access, use, modify, and share it – subject, at most, to measures that preserve provenance and openness." The definition states four requirements for open data:

• 1.1 Open license or status: The work must be in the public domain or provided under an open license.

• 1.2 Access: The work must be provided as a whole and at no more than a reasonable one-time reproduction cost and should be downloadable via the Internet without charge.

• 1.3 Machine readability: The work must be provided in a form readily processable by a computer and where the individual elements of the work can be easily accessed and modified.

• 1.4 Open format: The work must be provided in an open format. An open format is one which places no restrictions, monetary or otherwise, upon its use and can be fully processed with at least one free/libre/open-source software tool (Open Knowledge Foundation[34]).

This definition has been operationalised in the International Open Data Charter and the Open Data Inventory.

Open government data: The two main elements of Open Government Data are normally defined as follows: 1) Government data: is any data and information produced or commissioned by public bodies. 2) Open data: are data that can be freely used, re-used and distributed by anyone, only subject to (at the most) the requirement that users attribute the data and that they make their work available to be shared as well (Ubaldi, 2013[35]).

Open-source software: Software developed with source code that anyone can inspect, modify, and enhance (opensource.com, 2021[36]).

Principles and guidelines: The Principles of SDG Indicator Reporting and Dissemination Platforms and Guidelines for their application were proposed by the United Nations Statistics Division in 2019 and are a set of guidelines for member states to consider when developing an SDG national reporting and dissemination platform (UNSD, 2019[1]).

RSS: Really Simple Syndication is a format used to provide subscribers with new content from frequently updated websites (Encyclopedia Britannica, 2021[37]).

SDMX: Statistical Data and Metadata eXchange: SDMX is an international standard for the exchange of statistical data and metadata among international organisations and their member countries. (SDMX, 2020[38])

URI: Uniform Resource Identifier: A unique sequence of characters that identifies a logical or physical resource used by web technologies. URIs are used to identify anything described using the Resource Description Framework (RDF) (W3, 2004[39]).

Available at: www.paris21.org/dataportals

Contact the PARIS21 Secretariat at contact@paris21.org

Visit www.PARIS21.org

Follow PARIS21

🐦 @contactparis21

in /company/paris21/