

PARiS

Partnership in statistics for development in the 21st century

FLO 11 11 SIS. EWOR

Guidelines for Analysing Data Flows in National Statistical Offices

Acknowledgements:

Rajiv Ranjan, Innovation Team Lead at the Partnership in Statistics for Development in the 21st Century (PARIS21) and Jonathan Challener, Partnership and Community Manager, Statistics and Data Directorate at the Organisation for Economic Co-operation and Development (OECD) authored this guideline.

The guidelines gained significant insights from the joint work undertaken by PARIS21, OECD, the United Nations Children's Fund (UNICEF), and the United Nations Statistics Division (UNSD) during 2018 and 2019 in many countries including but not limited to: Cambodia; Ghana; Thailand; and Tunisia. Personnel in all National Statistical Offices (NSOs) in these countries provided rapid response to inquiries and were available to assist in obtaining the information authors requested.

The authors would also like to thank Anastassia Samsonova (OECD), Yves Jaques (UNICEF), and Abdulla Gozalov (UNSD) for their valuable contributions and insights. The guidelines benefited from strong support and guidance from the members of the Expert Review Team: Fredrik Emanuelsson and Lars-Johan Söderberg (Statistics Sweden).

Special thanks go to PARIS21 colleagues: François Fonteneau, and Philippe Gafishi, for sharing their experiences, knowledge and feedback, allowing us to turn these guidelines into a useful tool for the broader statistical community. Thanks go to Sasha Ramirez-Hughes for support with editing as well all communication and dissemination activities. The report was designed by Karamelaki Creatives.

Available at: http://paris21.org/dfaf

Contact the PARIS21 Secretariat at contact@paris21.org Visit www.PARIS21.org Follow PARIS21

- 🥑 @contactparis21
- in /company/paris21/

Visit www.OECD.org Follow OECD @OECD_Stat in /company/oecd/

FOREWORD

The COVID-19 pandemic has underscored the importance of efficient, timely and accurate data production. Government agencies, nonstate institutions and millions of people worldwide, for example, relied on online dashboards during the pandemic to understand, track and respond to the virus's manifestations.

Decision-making in the modern policy landscape requires up-to-date data infrastructure. In this context, information technology (IT) plays a central role in broadening access to statistical products, facilitating efficient data collection, analysis, packaging and dissemination. The importance of data in digital form for each of the processes is obvious, as it engenders efficiency by making use of IT to streamline operations seamlessly.



Yet, data portals (an IT solution) in many national statistical offices (NSOs) still rely on hybrid systems (both paper and digital). The online platforms hosts either digitalised paper publications (in scanned PDF format for example) or data copied from paper publications. This arrangement is limiting for apparent reasons: machines cannot read the data and is not scalable. In effect, they provide little benefit beyond providing a central repository for the information.

To move beyond paper-based dissemination systems, NSOs need data portals that include data in the digital format, as soon as they are available. Therefore, the Data Flow Analysis Framework (DFAF) provides an essential guideline for NSOs to undertake the background work required to establish a workflow-based integrated dissemination solution for indicators.

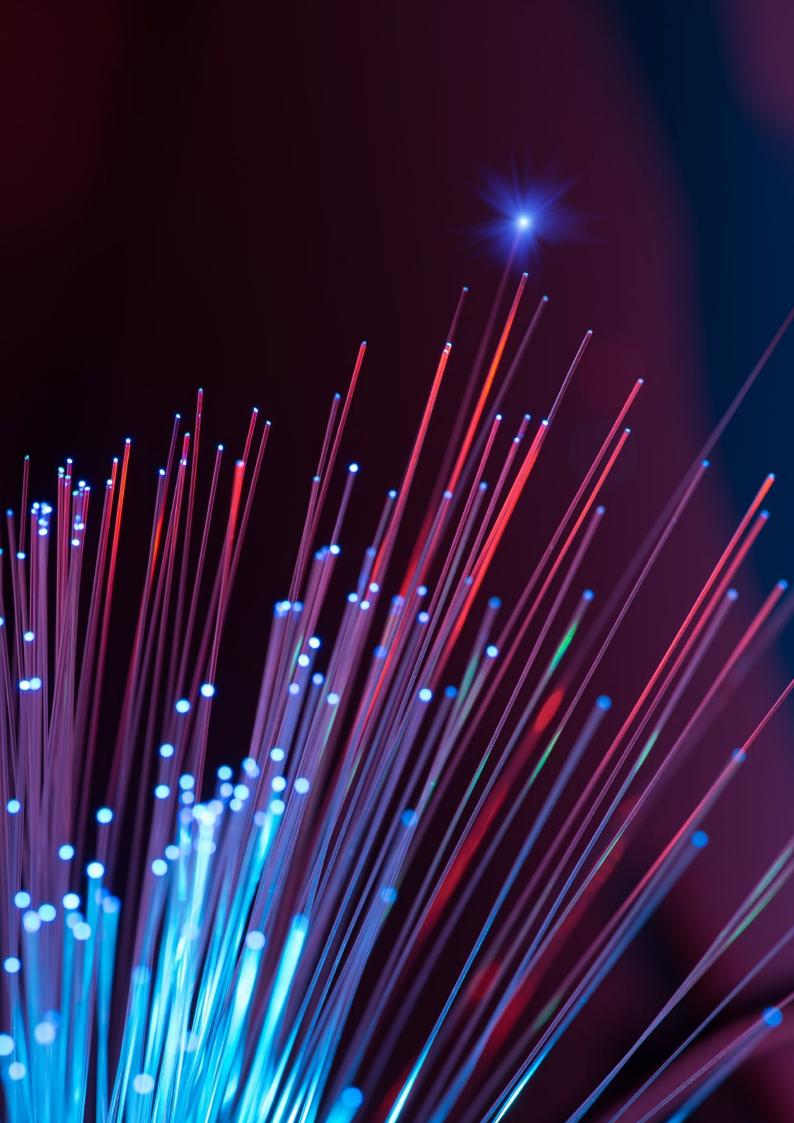
The Framework is straightforward and comprehensive. NSOs can map relevant existing processes using a data-centric view and identify recommendations for achieving the desired changes. The document is organised into a series of logical steps, from outlining the rationale for a proper data flow analysis to describing future processes for indicator dissemination.

Data-flows from new sources, including the private sector, the internet of things, and civil society, will only increase in the future. NSOs need broad digitalisation across the entire data lifecycle. This approach will support managing the deluge of data in the modern data ecosystem, support digital-government, develop standardisation and improve data availability and access.

Let me finish by saying that I am delighted to see this initiative come about and contribute to its creation by being one of the first organisations to undertake the analysis with the support of PARIS21 and the OECD.

I trust that this Framework will provide an invaluable resource to any national statistical office embarking on this important exercise.

Samuel Kobina Annim (Ph.D.) National Statistician of Ghana



CONTENTS

1. INTRODUCTION
2. WHY THE GUIDELINES?
3. HOW TO USE THIS GUIDELINE
4. ANALYSIS METHODOLOGY
4.1. Availability and organisation of indicators
4.2. Indicator dissemination on online channels
4.3. Data flow in indicator dissemination processes
4.3.1. Process identification
4.3.2. Entry criteria and inputs
4.3.3. People identification
4.3.4. Activity descriptions
4.3.5. Exit criteria and outputs23
4.3.6. Process diagram
4.4. Data modelling in practice
4.5. Information Technology infrastructure
5. WAY FORWARD
5.1. General recommendations
5.2. Proposed solution architecture
5.3. Proposed high-level action plan31
ANNEXES
Annex 1: Analysis segments and example questions
Annex 2: Data flow analysis report template

INTRODUCTION

1. INTRODUCTION

Rationale

National Statistical Offices (NSOs), particularly in low and middle-income countries, have experienced significant changes in recent years. Many have benefited from noteworthy improvements in their statistical processes. Examples of advances include faster data collection using handheld devices and flawless data processing with sophisticated software tools.

In these transformations, Information Technology (IT) has played a critical role in helping NSOs to save time, improve accuracy, and enhance access to their products. Yet, the improvements have not been consistent in all areas of statistical processes.

Worth noting is the case of **indicator dissemination**, which still lags in many NSOs. Although many use **data portals**¹ for indicator dissemination, such IT platforms have failed to make the desired impact. Often, NSOs data portals do not have the indicators people want and need. PARIS21 highlighted this in a study on national data portals², published in 2016.

One key reason for the sub-optimal performance of many data portals used by NSOs is that they host indicators originating from paper-based publications. The issue with paper-based publications is that they have inherent limitations, particularly in accommodating large volumes of data. Because of this, the inclusion of indicators in paper-based publications are often subject to prioritisation. As a result, there is still a gap between processed and published indicators in many NSOs. The difference propagates to data portals as well when NSOs transfer the indicators from paper-based publications to data portals.

In principle, though, data portals can dispense unlimited amounts of data. Currently, though, many do not, but this need not be the case. Since many NSOs already produce indicators as digital records, there should be no need to re-digitise the data from paper-based publications for data portals. The situation provides a strong business case for solutions, which can help NSOs quickly disseminate all indicators produced.

NSOs need data portals to be designed to capture the validated indicators as soon as they are available in digital format for the first time in the dissemination processes. However, to reach this state, design and development of data portals could make effective use of a proper **data flow analysis**, informing their functionalities, especially considering the distinctions of handling data and associated metadata in the digital realm. The analysis will help identify areas requiring specific attention.

Purpose

The purpose of this guideline is to provide detailed information regarding the analysis of indicator dissemination processes in National Statistical Offices (NSOs). This guideline lays out both basic and advanced work required in NSOs to set up a workflow-based, integrated dissemination solution for indicators.

Scope

The primary scope of this guideline is to map relevant processes using a data-centric view to capture the current state ("as-is") and to determine recommendations for the enhanced state ("to-be"). This guideline also gives an overview of methods, standards, and tools for guiding the

¹ A data portal is a web-based, interactive data platform that provides access to one or more databases containing statistical indicators.

² https://paris21.org/sites/default/files/Paper_on_Data_Portals%20wcover_WEB.pdf

design, development, and implementation of relevant solutions, but avoids recommending specific solutions.

As process mapping is a vastly complex field, the guideline is limited to a few selected aspects. Detailed information regarding the process mapping referenced in this guideline is available at https://www.uml.org/.

Who should read this guideline?

The intended audience of this guideline includes (but is not limited to):

- Primarily, heads of NSOs, project managers, and IT experts who need to establish dissemination systems in the digital age;
- Subject-matter experts or statisticians who define and specify requirements for a digital dissemination environment; and
- Development cooperation organisations and tool developers who support data management and dissemination solutions.

Document outline

Despite the use of data portals for indicator dissemination by many NSOs, such IT platforms have failed to make the desired impact.

- Section 1 outlines the rationale for a proper data flow analysis, which can inform the functionalities
 of effective data portals.
- Section 2 presents the business case, providing the contexts in which this guideline would be relevant. These include a strong wish by NSOs to gain operational efficiency, improve data availability and access for users, strengthen data sharing, and play a significant role in the expanding data ecosystem.
- Section 3 provides an overview on how to use this guideline, i.e., what steps to follow to capture the current state ("as-is") and devise recommendations for the enhanced state ("to-be") of relevant processes, using a data-centric view.
- Section 4 details the methodology of the assessment including the segments where information would require capturing: availability and organisation of indicators; indicator dissemination on online channels; data flow in indicator dissemination processes; data modelling in practice; and information technology infrastructure.
- Section 5 describes the 'way forward' after the assessment. This section focuses on articulating the "to-be" state of the indicator dissemination processes - the future processes for indicator dissemination.
- Annexe 1 presents the guiding questions for the analysis, separated by various segments of the assessment.
- Annexe 2 describes the template for the report, produced after the analysis, which guides further work in establishing a dissemination environment in the digital age.



WHY THE GUIDELINES?

2. WHY THE GUIDELINES?

From the extensive work undertaken in many countries, including but not limited to Cambodia, Ghana, Thailand, and Tunisia, which led to the development of the DFAF, the need for data flow analysis is inevitable.

Although NSOs differ significantly from one another, they all share similarities in requirements, processes, and often constraints, when it comes to indicator dissemination. These include a strong wish to gain operational efficiency, improve data availability and access for users, strengthen data sharing, and play a significant role in the expanding data-ecosystem.

Realise the benefits of using digital technology

For a long time, NSOs have used Information Technology (IT) to improve the efficiency of statistical processes. However, such innovations have remained confined only to specialised areas, creating islands where IT tools are less likely to interoperate, share data, and work together as a cohesive system. This fragmented environment of individual and disparate systems created what many researchers called silos, within NSOs.

Silos restrict the clarity of holistic vision, meaning that NSOs have struggled to have the significant impact desired with regards to using digital technologies. These reasons include a mismatch between IT systems and actual processes, and a lack of seamless flow of data across its lifecycle.

A data flow analysis offers active measures that NSOs can take to break the silos by harnessing data in digital format to realise the assured gains of using digital technology.

Improve data availability and access

There is an ever-increasing demand for data on NSOs. To cope with this growing call for more data, particularly since the adoption of the Sustainable Development Goals (SDGs), many NSOs have struggled to implement the necessary and sustainable statistical infrastructure.

This situation has resulted in fewer data being made available than produced by NSOs, affecting access and, hence, data use in policy and decision making.

Without first better understanding the data-flows, NSOs will not be able to support the development of a more robust data backbone, nor will they be able to facilitate for data to flow according to a decentralised, yet orchestrated model.

An analysis of data-flows provides constructive steps that NSOs may take to identify areas for improvement and develop tools to help increase the availability of data by crunching time, reducing errors, and making new data accessible.

Develop standardisation and comparability of data

Creating a model for the data in a database is a critical step in the data management process in a digital set-up. The modelling describes the data by relationships, semantics, and constraints, and helps with the enforcement of rules and regulatory compliances on the data. A standard data model can facilitate data interoperability across systems.

The use of the Statistical Data and Metadata Exchange (SDMX) standard by the official statistics community is an example. The data model underlying SDMX provides a standardised structure and vocabulary to express data and metadata.



Monitoring and reporting on global development frameworks such as SDGs using common standards such as SDMX would be a rational progression. Based on data-driven cooperation, it would also foster more collaborative, broader data ecosystems.

A data flow analysis makes a case for NSOs to adopt data modelling standards to maintain consistent data structures across different databases, including for dissemination.

Support digital government

Many governments aspire to digitalise their service delivery. However, the objective for digital transformation, i.e., integrating technology and data into all areas of functions, and fundamentally changing how governments operate and deliver value to their stakeholders, is far from having been achieved.

The use of technology by the public institutions has existed for some time now and has become popular. However, the availability of data in digital format throughout its lifecycle is a relatively new development. Governments are struggling with integrating data in digital format into their processes. Hence, the changes expected due to the use of digital and information technology in improving operational efficiency are lacking in many governments.

NSOs could play a pivotal role by bringing in expertise in digital data management across the government, thereby reinforcing a data-driven culture in the public sector. They could also be valuable by increasing confidence in government services by addressing issues such as privacy and confidentiality. Similarly, by removing obstacles to data interoperability, NSOs may develop standard agreements on how to represent data from different sources in a consistent way, across the broader data ecosystem.

The data flow analysis enables NSOs to address issues of data management, including digital security and privacy issues throughout the lifecycle of data, from collection to dissemination and use.

HOW TO USE THIS GUIDELINE

3. HOW TO USE THIS GUIDELINE

This document outlines steps for a successful analysis of **indicator dissemination processes** in NSOs using the Data Flow Analysis Framework (DFAF). The DFAF provides a consistent approach to capturing the current state ("as-is") and recommendations for the enhanced state ("to-be") of relevant processes using a data-centric view. The steps of the analysis fall under the following five themes.

- 1. Availability and organisation of indicators;
- 2. Indicator dissemination on online channels;
- 3. Data flow in indicator dissemination processes;
- 4. Data modelling in practice; and
- **5.** Information technology infrastructure.

The analysis phase combining all five steps is the first and most crucial stage for identifying and defining the target product vision for a data management and dissemination solution (e.g. a data portal), the level of capacity development required for data preparation, and the enabling environment for hosting the solution.

ANALYSIS METHODOLOGY

4. ANALYSIS METHODOLOGY

The analysis of indicator dissemination processes is the examination of activities in dissemination processes, their interconnections, and data flow. More detailed considerations would include identifying the sequence of activities and decision points; people who are making decisions; inputs and outputs of each activity; entry and exit conditions; interdependence; and legal requirements as well as the time and cost needed for completing the processes. The data flow would be the movement of data through various activities, potentially via IT systems, though they are not necessarily connected or automated.

The recommended approach to undertake the analysis is a combination of face-to-face interviews, workshops, surveys, and reviewing relevant documents.

In practice, the organogram of the NSO (National Statistical Offices) must first be considered in order to understand its structure. Understanding the roles and responsibilities of personnel within various organisational units will be useful. It will give insights into the inner workings of processes, including bottlenecks, to help create the basis for workflow (i.e., the sequence of steps through which a piece of work passes from initiation to completion) improvements. The analysis should not necessarily follow a linear path as one can learn that, in many NSOs, individuals perform multiple roles spread across the processes. Hence, focusing on roles instead of positions while conducting the analysis would be useful.

Since the focus of the analysis is on indicator dissemination, it is essential to study the usual channels of indicator dissemination: paper, digital, and online. Within each channel, there could be multiple sub-channels. For example, in "paper," survey reports and press releases; in "digital," CDs and UBS Flash Drives; and in "online," organisational websites, various social media platforms and data portals. Also, noting the sequence of channels in use would be insightful, i.e., establishing if there are any dependencies.

Delving deeper into each of the identified indicator dissemination channels and sub-channels for details will also be valuable. Information on the number of indicators they host, the basis of the organising indicator (e.g., if it is by source or sector), data modelling standards, indicator updating processes, and frequency of updates will be critical. These would be significant in designing dissemination scenarios for the "to-be" state.

The coverage of the analysis is to be limited to the processes related to data within the NSOs. However, it should also aim to capture the format, frequency, and structure of external data at the point of entry to the NSO.

To support the systematic and comprehensive coverage of all the topics in the analysis, a detailed list of suggested questions is in Annexe 1.

As an output of the data flow analysis, a report produced for further collective consultations would be useful. A guiding template for the report is in Annexe 2.

The scope of analysis should remain within the boundaries of the following five themes. During the analysis phase, activities should not be undertaken outside of this framework, such as actual data modelling or deployment of solutions.

It is important to note that the analysis is not a benchmarking exercise to compare one NSO with another, or to carry out a comparison of departments within the NSO. Its focus is on identifying best practices already in action and opportunities for enhancing these. It also serves as a tool for the NSO to help form future projects, either internally or with external partners.

4.1. AVAILABILITY AND ORGANISATION OF INDICATORS

Acquiring an overview of the dissemination of indicators, which the NSO produces, is essential. Even the approximate numbers of both (produced and disseminated indicators) can provide a good idea about the scale of work. It can help estimate, data volumes, and database sizes for future steps.

Note that quite often, the number of indicators produced and disseminated may differ (up to 50% by some estimates) due to specific internal reasons. Such facts should be noted for consideration when benchmarking the "to-be" state.

The analysis of the practice of grouping the indicators in each NSO is necessary for a sound understanding of the indicator dissemination practice. Many NSOs group indicators by publications (e.g., statistical yearbook or country snapshots), some by surveys (e.g., population census) or compilations (e.g., national accounts), and a few by topics or subjects (e.g., agriculture and fisheries or environment and energy). In many NSOs, a mix of these groupings often exists.

The idea of organising indicators by topic essentially refers to the grouping of statistical outputs, by subjects with common characteristics concerning concepts and methodologies. For example, the indicators could be grouped by "educational attainment" and "enrolment ratio," under the topic "education."

The analysis should consider if the NSO consistently maintains a classification for indicators that it produces and disseminates, and the basis of those classifications. A lack of indicators organised by topics or subject matter domains, often indicates weakness in data management practices.

Currently used dissemination channels may help to understand the classification of the indicators. When exploring data in various dissemination channels, paying attention to how users can find the data could prove useful, by establishing if it is grouped logically, and if the same classification exists in each channel.

Note that one indicator may be part of several groupings. For example, "Under-five mortality rate" could belong to both "Health" and "SDGs." Here, both Health and SDGs are group types. The critical aspect is that when accessing the same indicator from distinct groups, the observation value displayed will be the same.

4.2. INDICATOR DISSEMINATION ON ONLINE CHANNELS

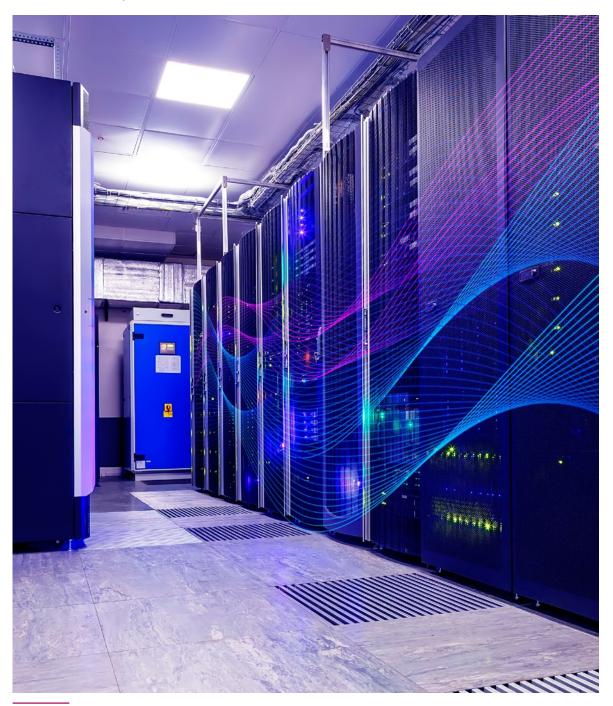
Many NSOs have already invested in IT infrastructure in the hope of realising the benefits it promises - speed and accuracy with hyper-connectivity of digital data. However, though the transition from analogue to digital has happened successfully in most NSOs, leveraging digitalised data to reduce the reporting burden for NSOs has remained largely elusive. In other words, taking advantage of digitalisation to create entirely new processes that lower the ever-growing burden on NSOs, has yet to happen for most.

Assessing the state of indicator dissemination on online channels, particularly data portals, would enable an understanding of the degree by which the NSO is benefitting from the use of digital technology and IT platforms.

This part should look closely at the online channels of indicator dissemination currently in use. Examining their key features (visualisation using tables, charts, maps; search engine; data download; availability of public web-services, multilingual support; and data analysis tools), including advantages and disadvantages, would be useful. Understanding the historical background of the previous decisions and introduction of the dissemination systems may be helpful in the development of a future strategy.

It is also imperative to understand particularly how the online channels are currently interacting with other tools and systems (e.g., using APIs to send or receive data, or providing public web services to other organisations). Having a single source of the truth (SSOT) for indicators is the ultimate goal, enabling the practice of aggregating data from many systems within the NSO into a single data warehouse. The data formats (MS Excel, Comma-Separated Values (CSV)), data structures (SDMX-ML)³, and technologies used for such interaction would be useful.

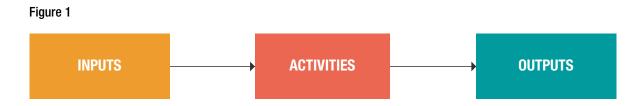
The analysis should consider if the NSO is satisfied with the user's appreciation of its indicator dissemination via the online channels. It should also assess how content the NSO is with the workload for keeping the online channels updated, to ensure dissemination of comprehensive, correct, and timely data.



³ https://sdmx.org/

4.3. DATA FLOW IN INDICATOR DISSEMINATION PROCESSES

A process is a set of interrelated or interacting activities that transforms inputs into outputs. This idea is illustrated in the diagram in figure 1.

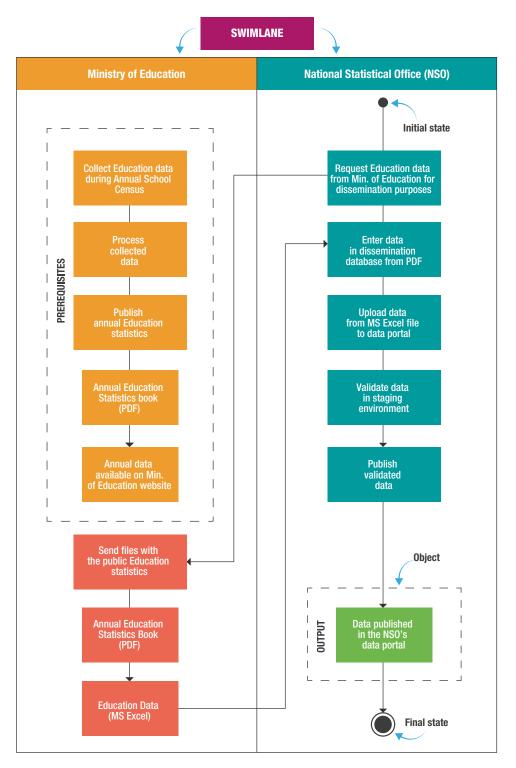


The current state, or "as-is" processes, should define the current situation of the related activities undertaken in the NSO for indicator dissemination. For example, in an NSO studied the process of updating indicators from the education sector looked like the diagram in figure 2. This activity diagram is an elaboration of the process. It portrays a sequence of activities and information flows from one responsible party to another. It informs its audience not only who is doing what and in which order, but also documents inputs that serve as prerequisites to activities and outputs that can be obtained upon completion of activities.

The scope may include upstream data collection as well as data management processes. The goal is to clarify precisely how the processes work at present, with a data-centric view. It is vital to capture the data flow between various stages of the processes. The following six steps may serve as guidance.



Figure 2



4.3.1. Process identification

The first step is to determine which processes need improvement. To identify a process, consider the start and end state of each, with specific inputs and outputs.

For example, within an NSO, a process in consideration for indicator dissemination could begin when the produced indicators are available in digital format for the first time (potentially after the data processing step) and end when those indicators are available on a data portal.

Note that there could be different processes for different sets of indicators, as well as for different channels of indicator dissemination. The complexity can vary significantly based on the size of the NSO and the organisation of indicators at the point of dissemination.

4.3.2. Entry criteria and inputs

Entry criteria are the criteria or requirements, which must be met before a specific process begins. It is a predefined set of conditions that must exist before activities can commence. In the example of the process of updating indicators from the education sector (figure 2), the availability of the annual education statistics book in PDF and annual data on the Ministry of Education's web site are the entry criteria. The statistics book and data are inputs. An input is usually an output from a preceding process.

This step will identify what completed actions must be performed before the process in consideration can begin. If there are preliminary processes that require completion before the main process can initiate, a brief description of them should also be provided.

The entry criteria and inputs for the process should be listed and the deliverables that the person responsible for initiating the process receives, in either physical or digital format, should be described.

This step should capture and describe existing formats used as input files (e.g., MS Excel, CSV, PDF) and whether any standardisation of templates is in use, as well as the frequency of the process execution (e.g., monthly, quarterly, and annually).

4.3.3. People identification

The individuals (and roles) involved in conducting the activities in the process must be identified; for example, individuals from different departments and positions within the NSO, and external organisations (line ministries and government institutions such as the central bank). Then, the primary responsibilities of each entity should be described, as well as their roles in the process.



When putting together an "as-is" process, access to people who perform the business process is critical. Secondarily, access to people who understand the process (such as a manager or subject matter experts) can be helpful, even if these individuals do not routinely perform the business process.

At least one person who represents each role in the processes should be part of the analysis. Once it is determined who needs to be involved, information about the current state ("as-is") can be obtained and captured, using a variety of different methods. Interviews and observations tend to be the best elicitation techniques to understand the current state process. Although, if both time and engagement from the NSO allow for it, workshops with each department (e.g., Economic Statistics) can significantly enhance the analysis insights gathered by generating discussion and engagement between the staff involved in different processes.

4.3.4. Activity descriptions

At this point, the process can be followed step by step, considering what an actor (an individual) in the process does at each stage. The important activities that different people in the NSO need to perform to make data ready for dissemination, can be listed at this point. This should include intra-departmental and inter-departmental collaborations and dependencies on the successful completion of the previous step. It is important to be clear about what the handoff is and who is doing what, by using different actor or role names.

Variations and constraints (e.g. embargo, data confidentiality, internal and external dissemination) should be included, remembering that process = input + set of activities + output.

4.3.5. Exit criteria and outputs

Exit criteria are the criteria or requirements, which must be met before considering a specific process complete. The outputs are tangible artefacts produced at the end of the process. In the example of the process of updating indicators from the education sector (figure 2), uploading the validated data is the exit criteria and data published in the NSO's data portal is the output.

In this step, identifying how the NSO and related organisations further use the outputs is also critical.

This step should also capture and describe existing formats used as input files (e.g. MS Excel, CSV, and PDF) and whether any standardisation of templates is in use, as well as the frequency of the process execution (e.g. monthly, quarterly, and annually).

4.3.6. Process diagram

A process diagram is a visual model derived by drawing a series of steps that occur between the start and the end of the process.

The various roles involved in different activities or tasks in the process should be identified and illustrated, marking the start and end with the interrelation of activities in order to make the diagram comprehensive.

Every step must integrate into the diagram, with decision boxes to capture variant flows. Using a pen and paper or a whiteboard for initial drafts is preferred. UML⁴ graphical notations could provide the elements of the representation. It is also appropriate to draw from existing diagrams the NSO may have created.

⁴ ttps://www.uml.org/

4.4. DATA MODELLING IN PRACTICE

Preparing data for data portals is different from preparing it for paper-based publications. Data portals are typically database-driven websites. In paper-based publications, a tabular representation of indicators is a static display of data, but in data portals, it can be flexibly dynamic. Since many NSOs use data portals for indicator dissemination, it is important, therefore, to understand and exploit the functionalities it could offer. For that, though, it is critical to understand data modelling.

Data modelling is a process of creating a template for storing data in a database. For indicators, the modelling will be an organisation of different data elements (e.g., concepts, dimensions, observation values) and their relationships with each other. Data modelling enables diverse representation, including tabulated data, while maintaining the consistency in naming, grouping, and observation values.

The idea of the data modelling practice is to produce common structures for the data (and accompanying metadata) to be organised, such that it is possible to describe similar data from one or several statistical domains in a similar way each time. For example, when disseminating indicators, the regional breakdown would always stay the same (official country-level classification). When there is a breakdown by sex, the NSO would always label it the same way (Male/ Female and not Boys/Girls or Masculine/Feminine). Similarly, when there is an internationally agreed classification available (e.g., ISIC4, ISCED, COICOP), it should be reused in the same way so that the data can be comparable, not only on a regional but on a global scale. This adherence is essential, particularly in the context of internationally monitored indicators like SDGs.

As a reference, the SDMX Content-Oriented Guidelines⁵ could be used to assess if standardised modelling practices may already be in use, and if they are aligned with international standards. Cross-domain concepts and code lists, available as part of the Content-Oriented Guidelines and at the SDMX Global Registry⁶, can be particularly useful in simplifying data modelling and facilitating interoperability.

The examination of the practice of data modelling in NSOs provides a reference to represent, clarify, define, and relate important statistical concepts, before producing indicators in the form of presentation tables.

⁵ The guidelines focus on harmonising specific concepts and terminology that are common to a large number of statistical domains: https://sdmx.org/?page_id=4345

⁶ http://registry.sdmx.org

4.5. INFORMATION TECHNOLOGY INFRASTRUCTURE

Information Technology (IT) has touched many aspects of statistical processes in most NSOs. When it comes to indicator dissemination on online channels, reference to websites, databases, and associated IT infrastructure used to collect, store, secure, and ultimately disseminate data (and related metadata as well as documentation) is inevitable.

While studying the indicator dissemination processes, assessing the underlying technologies is often the first step to making long-lasting, critical changes to the environments it has enabled. Therefore, the NSO must be clear, comfortable, and confident that this analysis will lay the foundation for the future 'to be' the solution the NSO desires.

The process begins with external or end user-facing interfaces (e.g., online dissemination channels), then delves deep into downstream internal networks, hardware, and software.

Factors taken for granted in many organisations, such as internet access, shared network drives, and functioning computers connected to the local area network, are essential, as well as the ability for an NSO to host a given application with adequate infrastructure and human resources capacity.

Some NSOs can face the additional limitation of an outsourced IT support and infrastructure hosting by another government agency, which manages a multitude of servers and applications with little knowledge or desire to learn the underlying software. The analysis should consider all these factors.

Although technology is a small part of the overall equation when looking at strengthening data management and dissemination practices, it can significantly contribute to how an NSO undertakes the related steps in the entire process.



WAY Forward

5. WAY FORWARD

The way forward is articulating the "to-be" state of the indicator dissemination processes - the future processes for indicator dissemination.

Designing future processes would rely on the opportunities for improvements identified during the analysis phase. Usually, the areas to focus on are the following:

- Possible ways to improve the quality of inputs (e.g., data and metadata) received for dissemination;
- Indicator dissemination process optimisation (e.g., online workflow);
- Human resource optimisation (e.g., necessary training, capacity development exercises, responsibility review);
- Ways in which to better align IT/digital strategy with organisational strategy;
- Required software developments aiding existing tools, or possible alternatives to the tools used by the NSO that would better support the needs; and
- Strategies to assure quality throughout the data flow.

Recommendations for each NSO will largely depend on the local context and should respond to the key issues identified during the analysis, with mitigation measures to minimise risks. Also, recommendations should consider obligations around legislation and compliance.

5.1. GENERAL RECOMMENDATIONS

To propose recommendations that help eliminate the identified bottlenecks and inefficiencies in the examined processes, a comprehensive view should be taken on possible ways to enhance NSO processes of indicator dissemination.

The recommendations could follow the typical themes of the analysis.

Availability and organisation of indicators: Developing an updated inventory of available indicators would be useful. Although it is likely challenging to identify an exact list of what is available, this should not become a blocking factor of the analysis. While adequacy is important, the comprehensive dissemination of indicators should also be stressed.

In addition, there would be added value in emphasising the importance of maintaining the consistency of indicator classifications across dissemination channels.

Indicator dissemination on online channels: Proposing necessary and short-term improvements to existing online dissemination channels (requests for new features, fixing existing bugs, configurations, switching solutions) would be practical. The analysis would indicate how to achieve a positive effect with different possible means.

Data flow in indicator dissemination processes: It would be useful to portray data flow through "to-be" processes, sketched in the form of diagrams. The illustrations should include textual descriptions that address the issues identified during the "as-is" process analysis. The "to-be" process diagram should map to the Generic Statistical Business Process Model (GSBPM) steps.

Data modelling in practice: This provides a timely opportunity to list the required changes to the existing modelling practice, describe the efforts needed for the data re-structuration, consider standards adoption (e.g., SDMX) that supports data interoperability, and national and international data exchange, including in the context of the Sustainable Development Goals (SDGs).

Information technology infrastructure: It would be critical to describe necessary changes to the existing information technology (IT) infrastructure to improve performance and assure compliance with the technical requirements of the "to-be" solution (servers, computers, network configuration, firewalls, internet connections, IT services, and other technical resources).

An additional overall recommendation could be to list the necessary training that would increase the skill level of NSO staff who work on data dissemination, or who support other essential related processes. It would be worth exploring possibilities to describe needed capacity-development exercises, as well as proposing possible changes to the dissemination team structure and role distribution to achieve improvement in the team-efficiency, including better use of existing skill sets.

Also, a more general set of recommendations could be on the organisational set-up and structure, including the alignment of IT or digital strategy with the National Strategies for the Development of Statistics (NSDS) or data plans.

5.2. PROPOSED SOLUTION ARCHITECTURE

A solution architecture would help to understand the overall structure of a future solution to ensure that it meets the needs of the users. It would be a vision for the tools and background processes and standards that would need to be designed and developed by the NSO.

Articulation of the proposed solution, which would eliminate or minimise the gap between "as-is" and "to-be" activities, could be detailed in the form of a diagram with notations clearly describing the role and purpose of various components within the overall solution architecture.

While designing the solution, close attention should be given to the availability of off-the-shelf tools that can be used to automate the processes. A large number of such tools are available now, many as open-source, that rely on underlying standards including SDMX, CSPA, and others. On the one hand, off-the-shelf tools can greatly reduce the cost and complexity of implementation by eliminating or reducing the need for custom software design and development. On the other hand, as many of the tools are maintained by well-established communities or organisations, they can significantly improve the sustainability of proposed solution.

The architecture diagram could contain the following components:

- · Main software items and their interconnections;
- Digitised data formats that are transmitted and processed during the different data lifecycle stages;
- Technological stack to achieve the efficient operation of the solution; and
- Main hardware items.

A viable solution could be elevated by meeting the following criteria:

- Interoperability between different components and software maintained by other organisations (by reusing commonly agreed standards like CSPA and SDMX);
- Provision of machine-readable access to data and metadata through an Application Programming Interface (API);
- Use of open source components;
- Support of the whole data lifecycle (GSBPM phases);
- Agility to technological change; and
- Support for localisation and multi-language interface.

5.3. PROPOSED HIGH-LEVEL ACTION PLAN

A high-level action plan would focus on establishing essential requirements and deliverables. It might be helpful to split the broad picture into smaller chunks so that planning can concentrate on each. At this stage, delineating the scope is critical.

Recommendations could be far-reaching and may cover a wide range of areas. That is why the analysis must provide a prioritised list of action items, aligned with the proposed solution architecture, which the NSO can define, in order to form a clear plan in actionable steps. The analysis report should not aim to create a project plan but lay the groundwork for the NSO to define the project scope, milestones, costs, organisation, dependencies, and risks for a likely Minimum Viable Product (MVP).

When prioritising, it would be useful to focus on areas that bring the most value with lower risks, build on existing best practices, motivate NSO staff, align with strategic objectives, and present the potential for reuse.



ANNEXES

ANNEXES

ANNEX 1: ANALYSIS SEGMENTS AND EXAMPLE QUESTIONS

Segment 1: Organisation (NSO)

Organisational set-up and structure

- a. Is the NSO governed by, and its work guided by a statistical law/bill? (If so, please provide the law and relevant decrees)
- b. Is the NSO part of another government agency (e.g. Ministry of Finance)?
- c. What is the organisational structure of the NSO (e.g. federated, centralised, or other)?
- d. What is the NSO's current role and level of engagement within the NSS?
- e. What is the size of the NSO (number of staff and their breakdown)?
- f. How is the senior management team organised (e.g. deputies, responsible areas, etc.)?
- g. Does the NSO have any advisors in place (e.g. embedded, supported by external partners)?
- **h.** Does the NSO have a specific unit responsible for the SDGs, or is it integrated?iDoes the NSO have an IT/Digital strategy which is documented and financed?
- i. Has the NSO received support on international IT cooperation such as technical assistance or opportunities to participate in international working groups/events on IT or collaborated with academia on the matter?

Segment 2: Indicator dissemination

General

a. Is a dissemination policy publicly available?

Indicator dissemination channel types

- a. Which channels is the NSO using for indicator dissemination (e.g. paper, digital, online, or others)?
- **b.** Do they support all the official languages, including for reporting purposes (e.g., SDGs)?

Indicator dissemination sub-channels under each channel

- **a.** What are the types of outputs (e.g., regularly printed reports, press releases, CDs, various social media outlets, different data portals, or other types)?
- b. How many different data portals (disseminating indicators) are in place?

Sequence and linkages

- a. Does the data in data portals come after and from paper publications?
- b. Who does the data entry?
- c. What are the processes that lead to data entry?
- d. Is there a single central data dissemination warehouse in place and used by all?
- e. If there is a central data dissemination warehouse, does it make the data available for machineto-machine communication, i.e. does it implement an Application Programming Interface?
- f. If the data dissemination warehouse does provide an API, is it based on a standard such as SDMX?

Segment 3: Data portals (for each identified data portal)

Ownership & collaboration

- a. Who is the primary data provider for this data portal?
- b. Does the data portal allow data entry by multiple users/organisations?
- **c.** How many distinct organisations (NSOs, line ministries, governments, or international agencies, etc.) enter data into this portal directly?
- d. Which lead organisation(s) are responsible for the overall input and maintenance of data in this portal?
- e. What is the name of the primary organisation(s) responsible for hosting and managing the data portal itself?
- f. Who were the critical collaborators in the various roles and stages of data portal development?

User-centricity

- a. How does the data entry work in the data portal (e.g., manually (including bulk upload), automatically from an underlying database, via APIs (Application Programming Interfaces), or other means)?
- b. How easy or difficult is the portal to maintain and update?
- c. What are the biggest challenges you face in maintaining and updating the portal?
- d. Does the portal allow for scheduling data releases (i.e., Embargo feature)?
- e. Do you use a web analytics tool on the data portal to record or analyse website traffic?
- f. How often are the findings from the web analytics used to change the content or design of the data portal?
- g. How many dataset downloads does your web portal record in a month?
- h. Who are the most frequent users of your data portal?
- i. How do you gather feedback from users?

Interoperability & scalability

- a. What interoperable software components does the data portal have?
- **b.** Does the data portal contain data pertaining only to a specific sector (e.g. agriculture)?
- **c.** Does the data portal contain data pertaining only to a specific monitoring framework (e.g. SDGs or the country's national development plan)?
- **d.** Is the data portal only for disseminating indicators from singular (one survey or a census) sources?
- e. How adequate are the resources allocated for maintaining and updating the data portal?
- f. What are the levels of skill sets in areas such as Database Management Systems (DBMS), data modelling, SDMX, API (Application Programming Interface), that are available among staff managing the data portal?
- **g.** What kinds of additional competencies would you need to manage better and update the data portal?

h. Are any data entry processes duplicated among the data portals in place, i.e. does any data have to be entered separately for each portal?

The number of indicators (dissemination)

- a. How many indicators does the data portal disseminate?
- **b.** Of the indicators produced by the NSO, what percentage does the data portal disseminate?
- c. Of the indicators available in the data portal, what percentage is from administrative sources?

Basis of indicator organisation

- a. Are the indicators organised by sectors or subjects, e.g., agriculture or health, etc.?
- b. Do the indicators pertain only to a specific sector or subject?
- c. Do the indicators pertain only to a specific framework such as SDGs?
- d. Do the indicators pertain only to a specific data collection exercise such as a census?

Data modelling

- a. Is there a specific data model in use?
- b. In what formats are indicators available?
- c. What standards (e.g., for classification) are in use in the data portal to organise the indicators?

Updates

- a. When was the data portal last updated?
- b. What is the frequency of updates?
- c. What are the update steps, and who is responsible for them?

Segment 4: Indicator production (by source/sector)

The number of indicators (production)

- a. Has the NSO assessed the number of indicators it produces in response to demanded indicators?
- b. If assessed, how many indicators does the NSO produce, and how many does it disseminate (approximate numbers)?
- c. Is there an exhaustive list of indicators produced by the NSO?
- **d.** Of the total number of indicators produced, how many are from external sources (e.g., line ministries)?
- e. What is the approximate percentage breakdown between indicators derived from administrative sources versus surveys?

Basis of indicator organisation

- a. How does the NSO organise the indicators produced (e.g., in different departments)?
- b. Are the indicators organised by sectors or subjects (e.g., agriculture or health, etc.)?

Data modelling

- a. Are there standards used in production (e.g., standard classification, codes, etc.)?
- b. Does the NSO have a list of standard classifications?

- c. What training takes place on the aspects of improving production? Who are the training partners?
- **d.** Which tools does NSO use to produce statistics? This question is applicable in all phases, including data entry, data analysis, and data production.

Segment 5: Data flow (for each identified process)

Process identification

- a. Which step in the process contains outputs of produced indicators in the digital format for the first time in the flow?
- b. Which step in the process contains outputs in the digital format of produced indicators available for dissemination?

Entry criteria (pre-requisites) and inputs

- a. Before the process begins, what criteria or requirements must be met and inputs (usually outputs from preceding processes) ready?
- b. Before the process starts, in which formats are the products available (e.g., CSV, PDF, etc.)?
- c. How often is the process executed (e.g. monthly, quarterly, annually, etc.)?

People identification

a. Who all are involved in the process and in what capacity (roles)?

Activities descriptions

- a. What activities constitute the process?
- **b.** What are the decisions and what is the mechanism for the approval of a step before moving to the next one?
- c. What are the inputs and outputs of each step?
- d. What are the bottlenecks?

Exit criteria and outputs

- a. What requirements must be valid to consider an activity complete?
- b. After the process ends, in what formats the products are available (e.g., CSV, PDF, etc.)?

Process flow diagram

a. What is the purpose of diagramming the process (e.g., understanding the process)?

Segment 6: IT infrastructure

General

a. What is the IT budget of the NSO for software, hardware, and services, including dissemination tools, etc.?

Internet

- a. Does the NSO subscribe to an Internet Service Provider with dedicated internet service?
- b. Are multiple internet connections in place (ensuring no redundancy)?

- c. What connection types are in use and used most (e.g., WIFI or wired)?
- d. Does the NSO use secure connections, or is it open?

Internal network (such as Local Area Network or LAN)

- a. Who manages the internal network?
- b. Do internal staff have an individual network drive available?
- c. Are there shared drives to manage file sharing across the NSO?
- d. What are the back-up procedures (e.g. frequency, disaster recovery process)?

Hardware (including databases)

- a. Are the NSO servers exclusively on-premises, externally hosted, or a mix of both?
- **b.** If externally hosted, who manages this on behalf of the NSO (e.g. cloud provider, other government agency)?
- c. If the NSO is using the cloud, for what purpose is it doing so?
- d. Which databases are in use (e.g., SQL Server, Oracle, etc.)?

Software

- a. Is the IT department responsible for all software used by the NSO or is there a decentralised model in place?
- **b.** Is the IT department responsible for the organisational website and CMS? Is there a CMS in place? If so, which kind (e.g. custom build, commercial, and open source, etc.)?
- c. What are the criteria for introducing new software?
- d. What policies are in place around software management and installation?
- e. What are the software packages from commercial vendors, development partners, and opensource?
- f. Does the NSO promote the use of open-source software (employing any policy directive)?



ANNEX 2: DATA FLOW ANALYSIS REPORT TEMPLATE

Report on the Analysis of Data Flows in [Insert the name of the NSO here]

Author(s):

Date of Report:

1 EXECUTIVE SUMMARY

This section will be a maximum of two pages with the following elements:

1.1 The overall context of the missions and workshops

To be extracted from the "introduction" section of the report

1.2 The main findings of the analysis exercise

To be extracted from the "analysis findings" section of the report

1.3 The key recommendations, solution features, and project set-up

To be extracted from the "way forward" section of the report

2 ABBREVIATION AND ACRONYMS

This section may include abbreviations and acronyms that require explanation. This list could be in a simple tabular format. However, if there are few, these explanations could be included as footnotes only.

Abbreviation / acronym	Explanation
NSO	National Statistics Office

3 INTRODUCTION

The introduction of the analysis report is to include the following sections:

3.1 Details of the analysis missions undertaken, and workshops conducted

This section will provide an overview of the analysis mission including dates, objectives, important outcomes, a mention of key resource persons and attendees met, responsible NSO staff at technical and political levels, and a note of gratitude to the head of the NSO and the people met during the missions. An annex could detail the list of people met.

3.2 The motivation for the analysis

This section will accommodate a brief statement to convey why the NSO requested and saw the need for a data flow analysis. It would be useful to include the reference to an official decision to conduct the analysis, potentially granted by a high-level authority.

3.3 The objective of the analysis

This section will provide the main reasons and purpose for conducting the analysis. The NSO might have presented a business case or received a specific demand from within the NSS.

It should also capture (at a high level) the current state ("as-is") processes around data production and indicator dissemination in the NSO, highlighting the required adjustments in relevant areas.

4 ANALYSIS METHODOLOGY

This section will provide a brief description of the methodology used for the analysis describing the following areas, including general observations:

- Availability and organisation of indicators;
- Indicator dissemination on online channels;
- Data flow in indicator dissemination processes;
- Data modelling in practice; and
- Information technology infrastructure.

5 ANALYSIS FINDINGS

This section will document the findings of current state ("as-is") analysis that identifies and evaluates NSOs' current processes which lead up to data dissemination on online channels, and critical dependencies.

5.1 Availability and organisation of indicators

This section will provide an overview of the current indicators produced and disseminated by the NSO, and will provide an overview of how the NSO organises indicators, i.e., classification by surveys, or domains etc., with reasoning. Further examples of the current classification(s) could be included as annexes.

5.2 Indicator dissemination on online channels

This section will document the online channels of indicator dissemination currently in use, their key features, including advantages and disadvantages, technologies, and data structures supported, e.g., SDMX-ML, CSV, etc. The focus here could be on data portals.

5.3 Data flow in indicator dissemination processes

For each process identified and assessed, this section details how the process works at present with a data-centric view. The following sub-sections may contain the specifics:

5.3.1 Process identification

This section will list the identified processes concerning indicator dissemination.

5.3.2 Entry criteria and inputs

This section will list the identified actions to be completed, and the inputs required before each process can begin.

5.3.3 People identification

This section will list the identified main parties with the focus being on the roles involved in each process.

5.3.4 Activity descriptions

This section will document the activities under each process.

5.3.5 Exit criteria and outputs

This section will list the exit criteria that must be met, and the outputs produced after each process is complete.

5.3.6 Process diagram

This section will provide visual models derived by outlining the series of steps that occur between the start and end point of each process, with a description of the activities in narrative form.

Here, the exceptions are listed and described in narrative form, adding them to the process flow diagram as necessary, or produce another workflow diagram to illustrate an alternative scenario.

The criteria for measuring the effectiveness of the process is described through a list of the critical painpoints in the current process, including through identification of the most time-consuming activities.

5.4 Data modelling in practice

This section will provide the current state of data modelling practices in the NSO. The details may include the following aspects:

5.4.1 Awareness and current practice

This section will document the general practice of data modelling in the NSO with an analysis of the knowledge, skills, attitudes, and beliefs in this regard.

5.4.2 Use of standards

This section will document the current use or adoption of common standards for data management and dissemination, and outline whether there is any use or adoption of common standards for classifications in the different subject matter.

5.4.3 Human resources

This section will document who the people involved in undertaking data modelling activities are, and an estimate of the approximate need for resources (people and time) to undertake the data modelling exercise.

5.5 Information Technology infrastructure

This section will document the status of the information technology infrastructure in the NSO, including internal databases, network, hardware and software issues, and human resources situations, under the following groupings:

5.5.1 Network & hardware

This section will document the current network and hardware capabilities, including connectivity, and how internet access works at the NSO. This section also serves to note if computers in the NSO work in online or offline mode, and if there are any shared network drives, etc.

5.5.2 Software

This section will document what software and hardware (e.g., Operating System for servers, databases for data storage) are in use across the distinct phases of the data flow.

5.5.3 Human resources

This section will list the main user types of tools identified in the previous section, as well as the person(s) involved in the overall management of IT at the NSO. Note that the systems and infrastructure may be managed outside of the NSO, or even decentralised within the NSO.

6 WAY FORWARD

This section will contain the recommendations, based on the future processes ("to-be") that are to happen, for indicators produced, and should respond to the main findings as identified in the current state of data flow ("as-is") section of this report. The aim of this section is to act as an initial high-level blueprint for action.

6.1 General recommendations

This section will provide general recommendations that should follow the usual themes of the analysis. It should summarise key finding in the form of a SWOT analysis (Strengths, Weaknesses, Opportunities, Threats). The table below illustrates the potential presentation. The SWOT analysis is a strategic planning technique used to help NSOs identify strengths, weaknesses, opportunities, and threats related to future project planning.

Theme	Findings	Strenght	Weakness	Opportunity	Threat
Availability and organisation of indicators					
Indicator dissemination on online channels					
Data flow in indicator dissemination processes					
Data modelling in practice					
Information Technology infrastructure					
Organisation set-up and structure					

6.2 Proposed solution architecture

This section will provide a proposed high-level architecture vision for a solution. Along with text, the solution (e.g., a data portal) should be detailed, with a sketch showing distinct blocks, with each block marked to identify its purpose.

6.3 Proposed high-level action plan

This section will provide a high-level action plan, which will list the action items in the order of priority. The aim of this section is to provide a starting point whereby the NSO can define a clear plan, in actionable steps, of what a future project might involve. Some extracted information from previous sections will appear here, including reference to the proposed solution architecture.

7 ANNEXES

7.1 Annex 1: List of people met

A list (table format) of the people met during the mission, including their roles and responsibilities.

7.2 Annex 2: List of documents collected during the analysis

A list of the documents (with links) collected before, during, and after the analysis.



Available at: http://paris21.org/dfaf

Contact the PARIS21 Secretariat at contact@paris21.org Visit www.PARIS21.org Follow PARIS21 @ contactparis21 in /company/paris21/

Visit www.OECD.org Follow OECD @OECD_Stat in /company/oecd/

