



BUREAU OF NATIONAL STATISTICS
AGENCY FOR STRATEGIC PLANNING AND REFORMS
OF THE REPUBLIC OF KAZAKHSTAN

THE CONCEPT

OF THE DEVELOPMENT OF STATE
STATISTICS AND THE NATIONAL
DATA ECOSYSTEM

Astana 2022

TABLE OF CONTENTS

1

PREFACE

page 3

02

GENERAL PROVISIONS

page 4

03

ANALYSIS OF THE CURRENT SITUATION

page 7

04

OVERVIEW OF INTERNATIONAL EXPERIENCE

page 16

05

DATA MANAGEMENT MODEL IN KAZAKHSTAN

page 22

06

BASIC PRINCIPLES

page 27

07

GOALS AND OBJECTIVES

page 29

08

EXPECTED RESULTS

page 35

09

STAGES OF IMPLEMENTATION

page 37

PREFACE



The rapid development of digital technologies has changed many processes familiar to us and opened up a unique opportunity to use the big volume of data accumulated in both the private and public sectors. In these conditions, data becomes an **additional factor of productivity**, stimulating innovation and correct decisions at all levels.

However, along with the **potential benefits**, data also creates **new challenges** to ensure quality, accessibility and confidentiality of data, introduce transparent mechanisms for its monetization, as well as to create processing and storage infrastructure.

Working with data at the Bureau of National Statistics of the Agency for Strategic Planning and Reforms (BNS) on a daily basis from both producer's and user's side of official statistical information, we are aware of the **significant potential of an effective use of data** and its impact on all areas of state development.

In this regard, the BNS has faced a **difficult question** – how to integrate the established national statistical system and the emerging new data ecosystem, while at the same time solving the noted challenges.

Having carefully studied the international experience and best practices of the corporate sector, we came to the conclusion that data management should be defined as a **separate function of the state** and taken out as a priority policy.

In order to implement this policy, it is important to consider data as an asset that has an economic value and requires effective management. In this regard, given the best practices of data management, it is proposed to lay down **5 principles** as the basis of the national data ecosystem: **availability (inclusivity), safety (confidentiality), single entry (the «only once» rule), efficient use and timeliness**.

Based on these principles, there has been prepared the **Concept of the development of State Statistics and the National Data Ecosystem** (the Concept). This document has been developed by the BNS and presented to your attention in order to **share the vision** of where the national statistical system is moving to and what challenges it has to overcome to form an effective data ecosystem.

Many proposals **go beyond the scope** of official statistical activity and are therefore proposed **for discussion** within the framework of the Concept. We invite you to take part in the discussion and contribute to the formation of the foundation of the national data ecosystem.

We look forward to your feedback

Head of BNS ASPIR
Zhandos Shaimardanov



02

GENERAL PROVISIONS

General provisions

Presented for public discussion Concept was developed **with the aim of reforming the national statistical system and unleashing the potential of the data ecosystem.**

In Kazakhstan, the national statistical system is being developed in accordance with the legislation and the fundamental principles of official statistics. At the same time, along with the national statistical system, there emerges a **larger data ecosystem**, which currently has no regulatory environment.

According to McKinsey's^[1] definition, a data ecosystem is a platform that combines data from multiple vendors and creates value through the use of processed data.

The concept of a data ecosystem is just coming into use, and therefore other terms may be used, which, in fact, are equivalent to the definition presented above. So for example, within the framework of the World Bank report «Data for Better Lives»^[2], there was introduced the concept of **an integrated national data system**, which is defined as a platform that promotes fair and secure data exchange between government, individuals, civil society, academia and the private sector.

Within the framework of this Concept, the national data ecosystem is divided into two components:

- 1. State sources** that form statistical (collected by BNS) and administrative data (collected by other state bodies).
- 2. Alternative sources** are other non-governmental data sources, which may include mobile operators, marketplaces, social networks, aerial images of the Earth, etc.

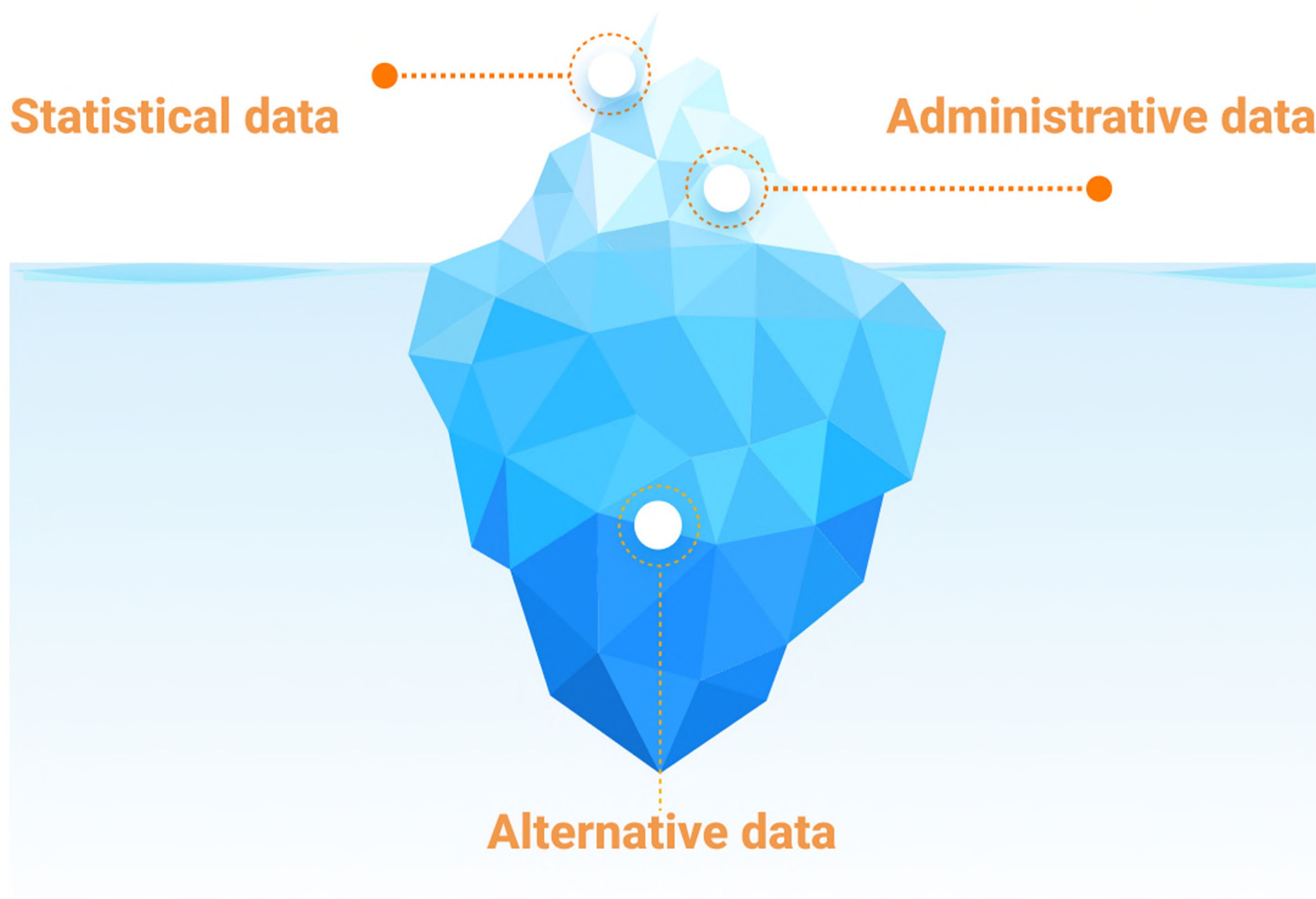
The provisions of the Concept are applied to both **state** and **alternative data** sources when they are provided for inclusion in the information systems of statistical bodies and public sector organizations, taking into account the requirements of the legislation of the Republic of Kazakhstan regarding the observance of the rights and legitimate interests of data owners who are not state sources, as well as compliance with the principles of non-interference in their economic activities.

For clear understanding of approaches and differentiation of roles within the framework of the Concept, the following concepts are also used.

State data is the data contained in the information systems of statistical bodies and public sector organizations, as well as in information systems created for the purpose of exercising the powers of public sector bodies and organizations.

[1] <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/data-ecosystems-made-simple>

[2] <https://wdr2021.worldbank.org/>



National Register is the reference database that meets the criteria of data quality, containing a systematized list of subjects and (or) objects and their characteristics, formed on the basis of existing information systems and (or) data of relevant industries.

Data management is the set of processes for collecting, transmitting, storing, processing, providing, distributing, destroying and ensuring the quality of data, as well as providing access for reuse.

Data management model is the set of data descriptions, organizational and technological rules and standards used for data management purposes, including a description of the relationships between types of data, as well as between the objects defined by them, including for interdepartmental exchange (interaction).

03

ANALYSIS OF THE CURRENT SITUATION

Analysis of the current situation

As a basis for the development of the national data ecosystem, it is proposed to use the national statistical system, which already has a legislative status and accumulates a large volume of structured data that meets the international standards.

At the same time, **the development of the national statistical system and its transformation into a full-fledged data ecosystem is hindered by a number of systemic problems in Kazakhstan**, as well as in overall international practice.



INSUFFICIENT USE OF ADMINISTRATIVE SOURCES OF INFORMATION, DUE TO THEIR POOR QUALITY AS WELL

The BNS is officially entrusted by law with the coordination of statistical activities carried out in the country, however, according to the current legislation, the BNS cannot significantly influence the formation of departmental data of state bodies (administrative sources). The problem is compounded by the fact that often there is no unified data register in the departments, national classifiers and unified reference books are not used.

Due to the lack of authority of the BNS and the closeness of administrative sources, as well as the lack of a single centralized data management, there is a limit to obtaining data from administrative sources in full, and it is almost impossible to verify their quality and compliance with international standards.

Currently, a significant part of administrative sources is concentrated in state databases, which have a **disparate structure and different levels of maturity**. It significantly complicates the integration of all these sources. In this regard, in order to drastically reduce statistical reporting and switch to the use of administrative sources, there is a need for **proactive approach to integration** with existing and new information systems based on the **target data architecture**.

At the same time, state bodies that own information systems are **not always interested in improving the quality** of departmental data and, accordingly, their further dissemination. The information accumulated by state bodies is practically not published on their official websites, which does not allow proper control of the quality of administrative sources. In this regard, it is necessary to extend the standards of official statistics to all administrative sources and give the function of quality control to the relevant data stewards.



THE IMPOSSIBILITY OF USING ALTERNATIVE SOURCES OF INFORMATION AND THEIR FULL INVOLVEMENT IN ECONOMIC TURNOVER

Today, every user of information resources leaves a digital «footprint» that accumulates both in state and alternative sources. These data, combined with others – publicly available or private – open up great opportunities for identifying patterns, producing leading indicators, developing new business models or products, as well as making decisions for the common good of the citizens of Kazakhstan.

At the same time, access to this data is restricted for reasons related to privacy and competition rules, as well as the consequences associated with the costs of obtaining and exchanging information.

As a result, state **misses the opportunity to promptly respond** to certain acute problems **without using alternative sources** such as transactional activity (fiscal authorities and payment systems), data from state monopolists (freight volumes, electricity consumption), data from mobile operators and other real time indicators.

Unlike official statistics, alternative data is generated faster and almost immediately. In the future, after confirming the quality of these sources and ensuring stable access to them, a significant part of statistical reporting may be replaced by alternative sources.

Alternative sources are already being actively used to **monetize data**, thus extracting **economic benefits** from data. At the same time, data monetization transactions are non-transparent, which does not guarantee compliance with the rules of confidentiality and fair competition in this market.

The national data ecosystem will **unleash its full potential**, when the transactions in the data market become **transparent without the intervention of state institutions**.



NO DATA MANAGEMENT POLICY

The lack of a data management policy becomes an increasingly important issue, compounded by a multitude of new legislative requirements, as well as an increase in the volume of data coming from numerous structured and unstructured sources.

For an effective management of the national data ecosystem, it is necessary to take into account the areas of knowledge on **data management according to the definitions of the data management framework (Data Management Association)**. Figure 1 below presents an **expert assessment** of the maturity of Kazakhstan's data ecosystem for each of the 10 functions of the Data Management Framework (**DAMA wheel**).

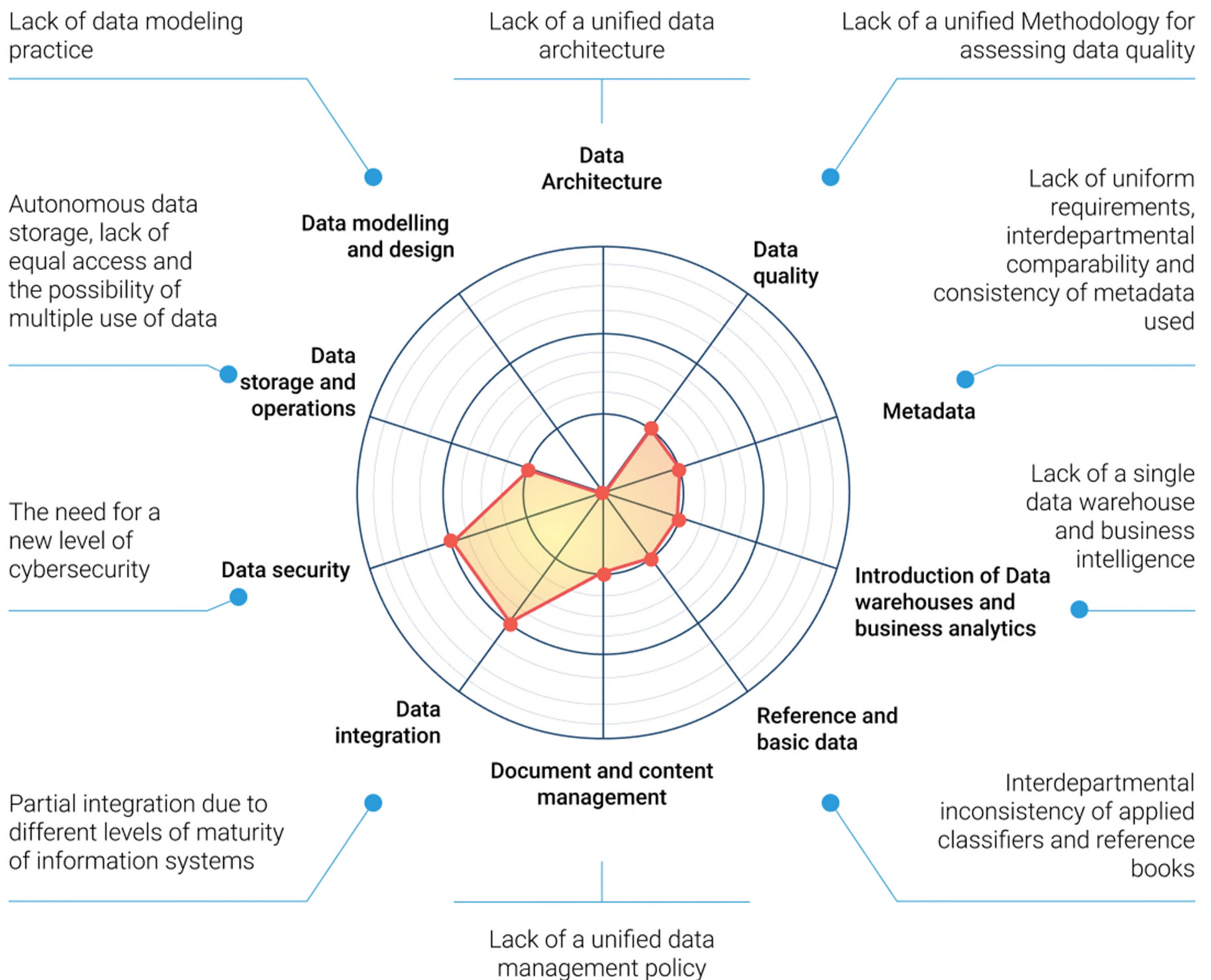


Figure 1. Data Management Framework (DAMA International)

It is difficult to implement national initiatives or ensure interoperability, consistency and data exchange due to the lack of interdepartmental data standards, data management policy requirements. In this regard, the issues of data and information management, data access, data privacy and security protection, data quality, data architecture are **key components of the data management policy**.



OUTDATED COLLECTION METHODS AND MULTI-STAGE DATA PROCESSING SYSTEM

Outdated methods of data collection and processing are a serious problem for the development of the national data management system.

Despite the fact that in **2021 over 80% of statistical** reports were submitted by respondents in **electronic form**, there still remains traditional paper reporting. The preservation of both electronic and paper reporting formats **significantly reduces the quality and reliability, increases the timing** of data collection and processing, and also leads to **significant additional budget expenditures**.

The current traditional method of collecting statistical reports involves filling out reporting forms by respondents in a manual format, which **leads to** errors and does not exclude the possibility of **entering false information**.

In addition, a multi-level system of data collection and processing leads to an increase in the processing time of statistical reports and an increase in the number of errors, and as a result, a decrease in the quality and relevance of the generated statistical information.

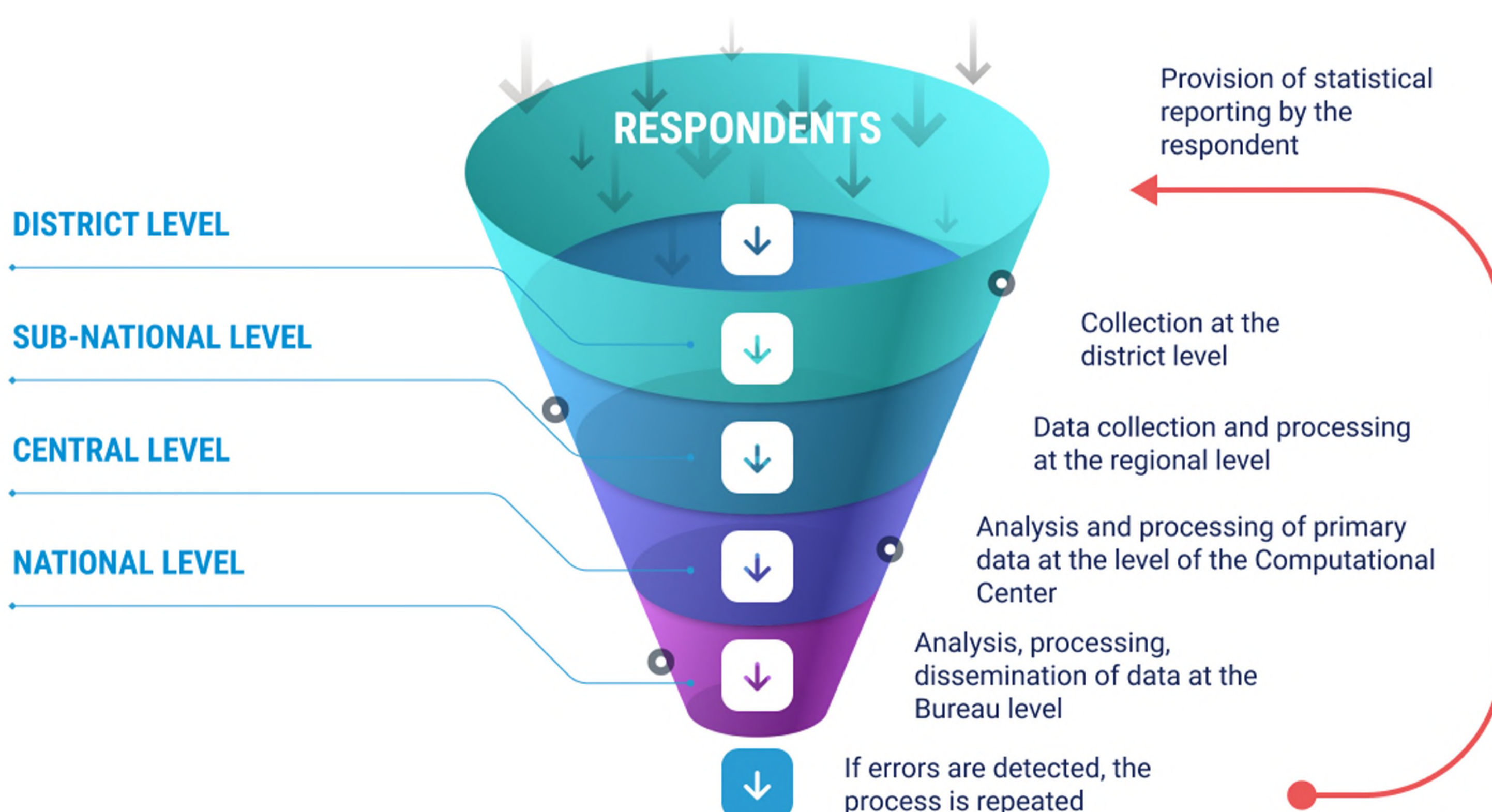


Figure 2. Multi-stage data collection and processing scheme

A similar situation is observed in other state bodies and institutions that form departmental reports for their own purposes. The largest share of paper reporting falls on local executive bodies. This leads to an **excess of stored data, the use of various formats and technologies** in relation to data. As a result, there is a **high cost** (or impossibility) of linking data, and there is also no overview of what information is available.



DUPLICATION OF REPORTING

Along with statistical observations, **state bodies introduce additional departmental forms of reporting**, thereby exercising the right to collect information in order to fulfill the functions assigned to them.

The main reason for the introduction of additional forms of departmental, regional reporting is the legally established prohibition of subjects of official statistical accounting to submit primary statistical data for use for other purposes not related to the formation of official statistical information.

Kazakhstan remains one of the few countries in which the collection of statistical and tax information is **practically not synchronized**. For example, statistical forms on the financial and economic activities of enterprises are almost completely identical in their content to tax declarations that are submitted to the State Revenue Committee of the Ministry of Finance of the Republic of Kazakhstan. At the same time, **the lack of integration** of the BNS to the information systems of tax authorities does not currently allow eliminating duplicate statistical reports.

In this situation, the principle of effective data management «**only once**» does not work. The use of existing data should be managed efficiently, the collection of duplicate data should not take place if there is a government agency that has already collected the necessary data from the respondent

The inconsistency of statistical and tax reporting, as well as the lack of access to administrative sources in full, is an **obstacle to eliminating duplicate reporting and, accordingly, reducing the burden on business**.



LOW LEVEL OF USER TRUST

The absence of an interdepartmental communication policy, complete methodological coherence and compatibility of data from various sources, as well as the existing closeness of administrative sources significantly **reduce the credibility of official statistics**.

Statistical surveys of various types may rely on different approaches, classifications and methodologies, therefore they cannot provide the user with fully linked information. It is not uncommon for users to encounter the fact that two different measurements of the same quantity are published with different values. There is not always sufficient information about metadata and data quality assessment, there is no comparability between subject areas.

In 2020, the population was particularly distrustful of the published data on mortality from Covid-19. When analyzing the data on morbidity and mortality indicators formed by the BNS and the Ministry of Health of the Republic of Kazakhstan, a number of inconsistencies and significant data discrepancies were revealed. The reason for the data discrepancy was the different approach in the formation of statistical and departmental information and the lack of a unified information system between medical organizations.

Thus, the Ministry of Health of the Republic of Kazakhstan considered only those deaths caused by Covid-19 that they diagnosed with a positive PCR test of a patient. If the analysis was not carried out, then another diagnosis was made or the cause of death was uncertain. In addition, cases of pneumonia and covid were unreasonably separated. This was significantly different from the approach in advanced countries.

To ensure the **credibility of statistics** and, in general, of information provided by public authorities, the methodology and policy should be **transparent, documented and published for public access and correct interpretation**.



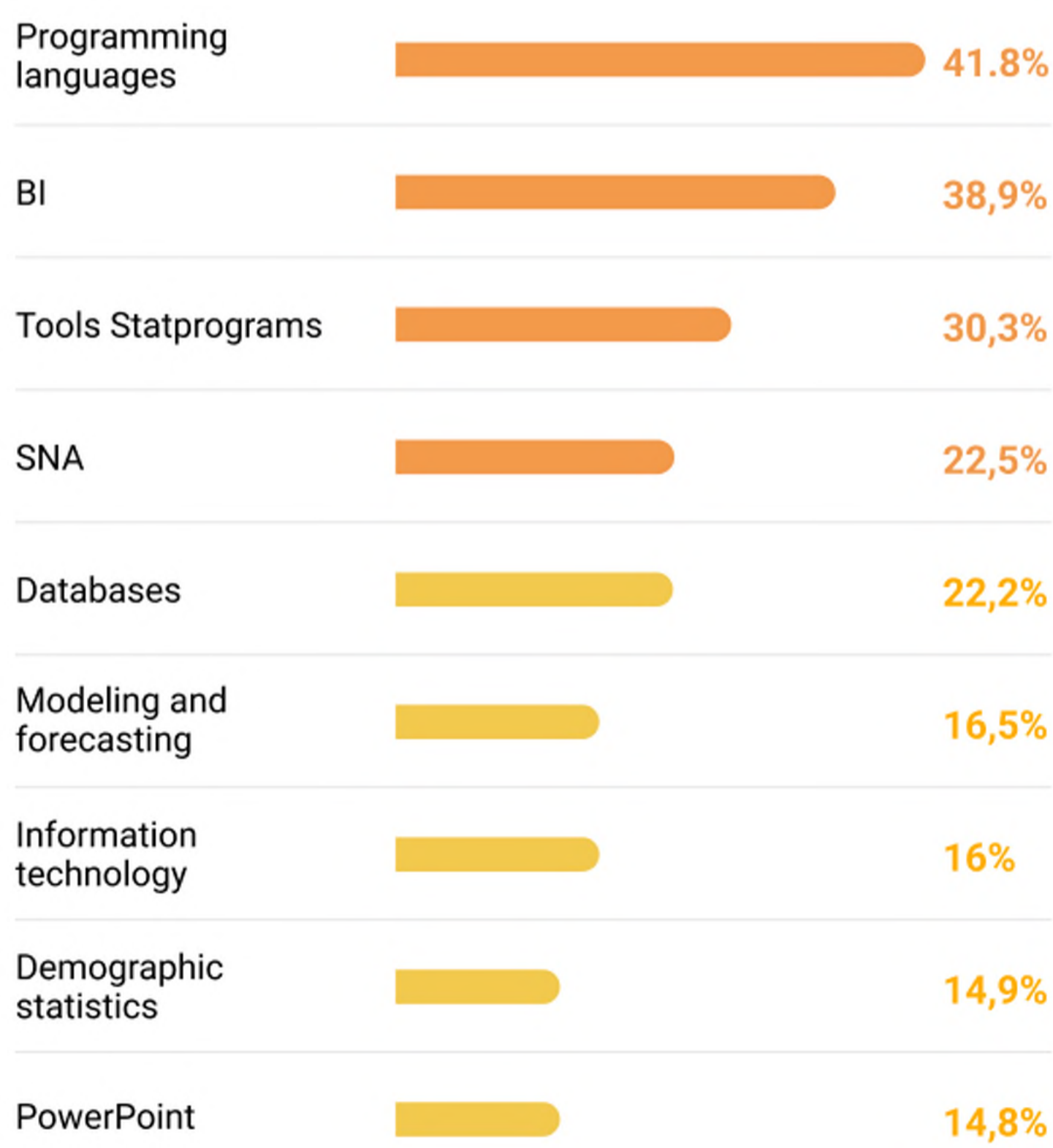
LOW LEVEL OF COMPETENCIES IN DATA ANALYSIS

In order to continuously improve the quality of formed statistics, it is necessary **to develop skills in working with data** throughout the civil service.

Today, the employees of the BNS are **mainly statisticians** with economic education and **limited competencies in business analytics**, there are practically no specialists with competencies in programming and data management. In this regard, the published statistical information does not contain in-depth analysis, cross analytics of interrelated indicators necessary for timely and proactive decisions in real time.

An internal analysis (gap-analysis) conducted in August 2022, based on the self-assessment of the BNS employees, revealed a number of problems with the current level of competence development. **2,221 BNS employees** took part in the survey. It turned out that more than **63%** of the surveyed specialists want to expand their skills in a particular field of knowledge, in particular, to gain skills in working with modern **BI tools, programming**, as well as to acquire the necessary **business analytics skills**.

LOW COMPETENCE HIGH DEMAND



HIGH DEMAND

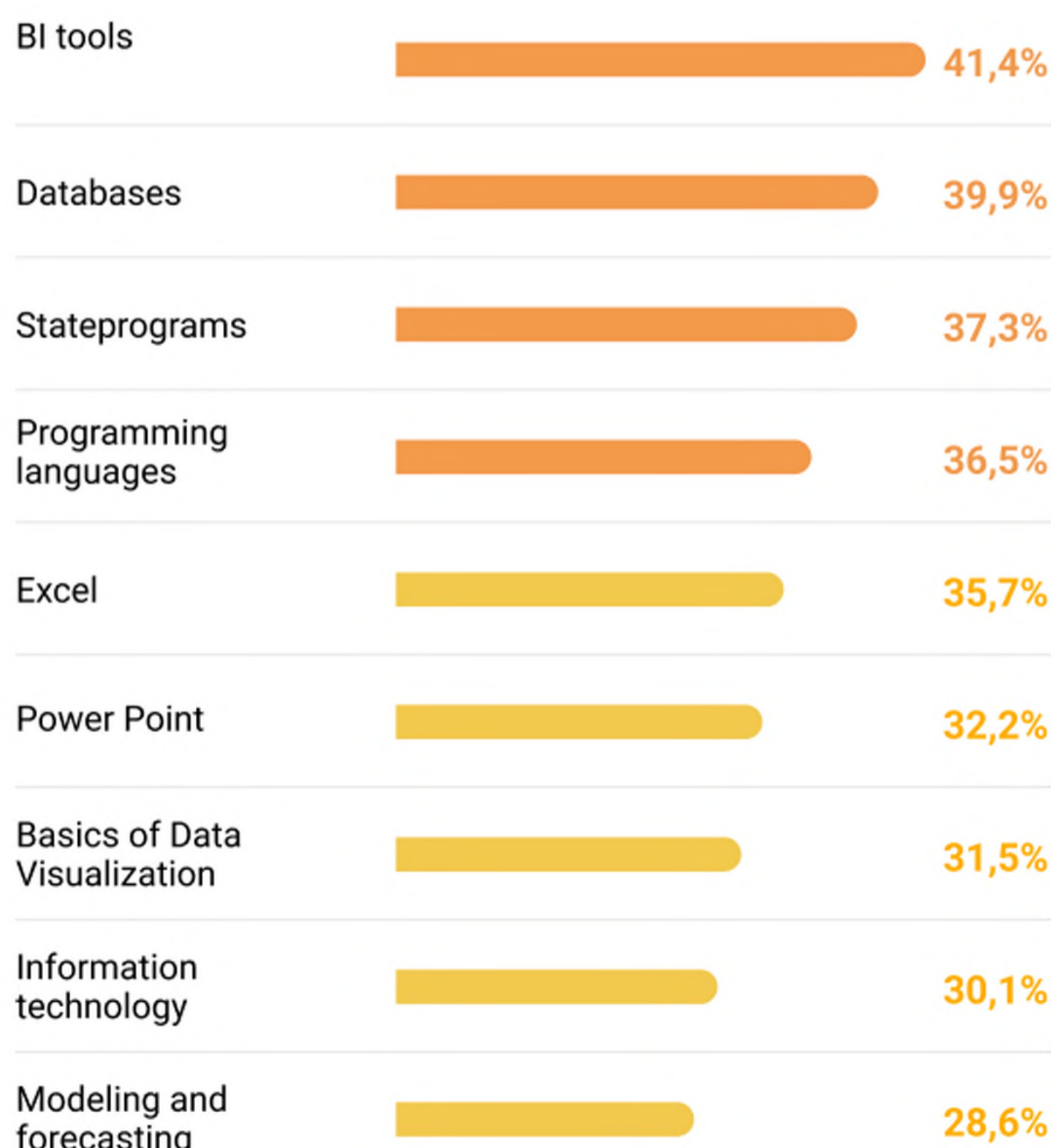
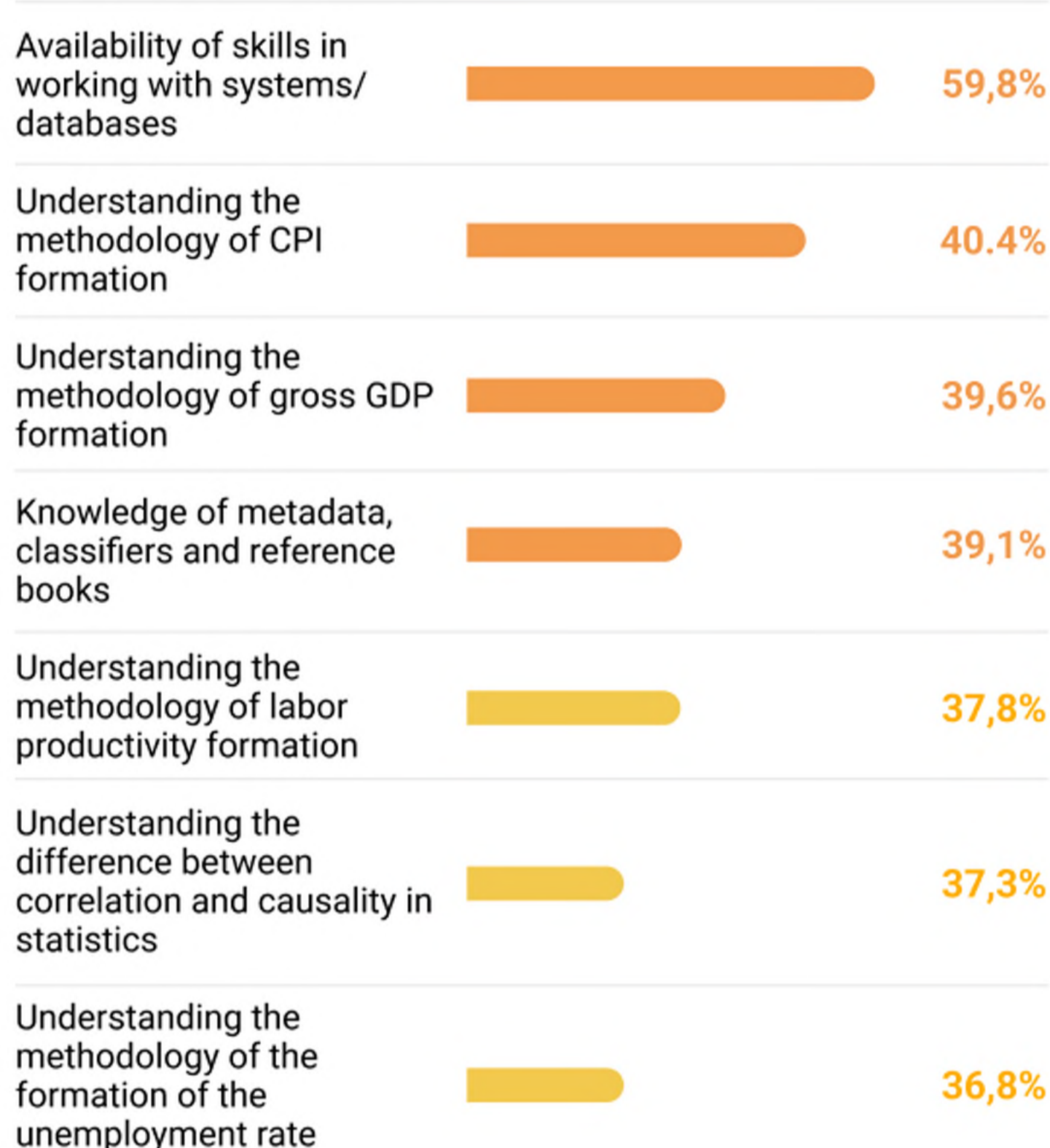


Figure 3. Results of gap-analysis of BNS employees

It should be noted that a **similar situation** takes place in **other state** bodies, there is a significant shortage of specialists with data processing skills. In addition, civil servants face lack of **basic knowledge in the field of statistics** necessary for the correct use of official statistical information in decision-making.

Overall, 30 central state bodies and 20 local municipalities took part in the survey. It turned out that **63.8%** of the surveyed civil servants are interested in training in **statistical literacy and working with data**.

LOW COMPETENCE



HIGH DEMAND

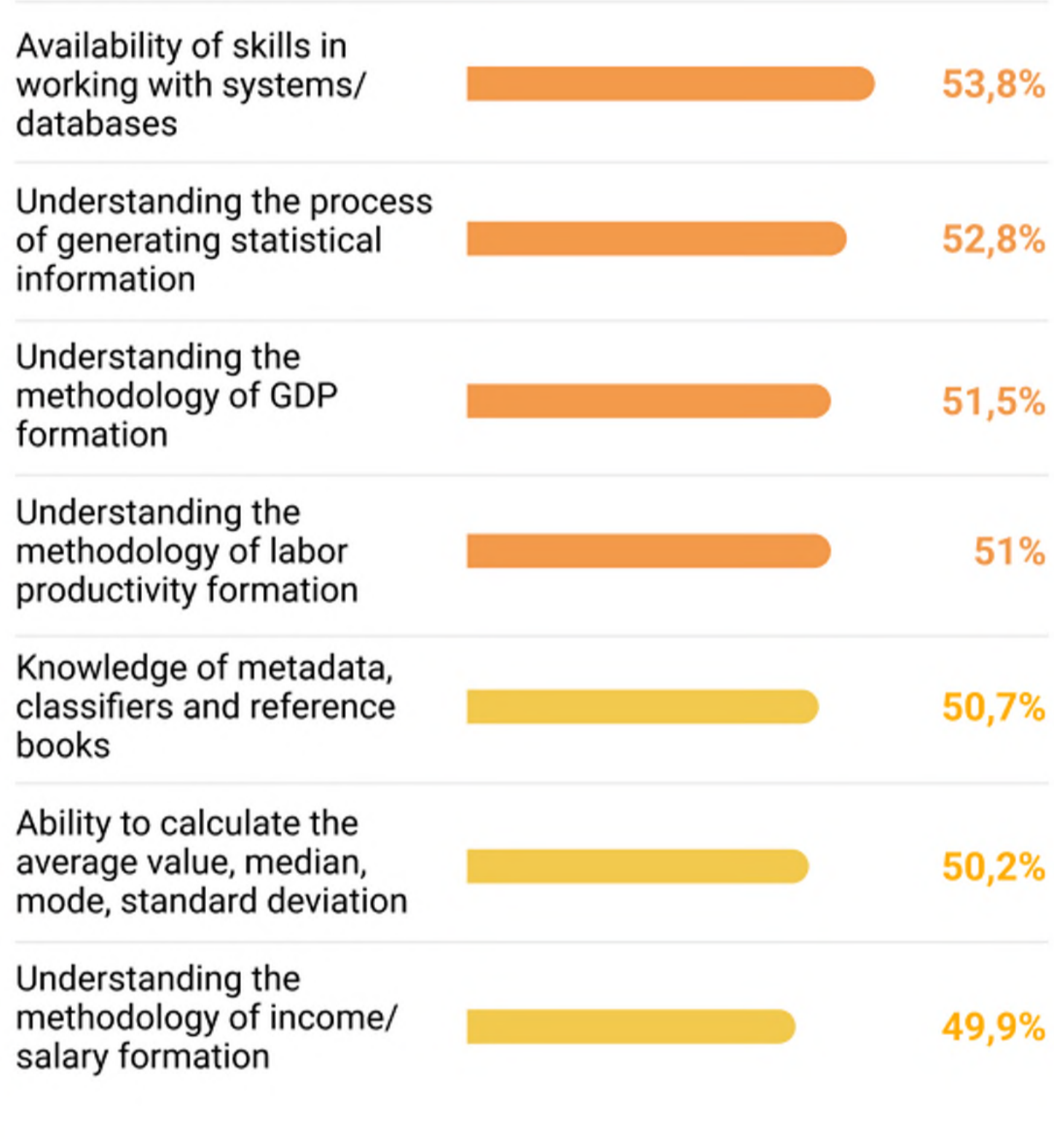


Figure 4. Results of gap-analysis of civil servants

Thus, all of the above problems hinder the **development of the national data ecosystem** and require solutions within the framework of the presented Concept.

Review of international experience

According to the results of 2019 **assessment of the statistical effectiveness of the World Bank[1]**, the national statistical system of Kazakhstan entered the **TOP-50** of the world ranking, scoring **78.6 points out of 100**. As positive aspects of this assessment were noted the legal basis and mandate of official statistics, sufficiently high level of quality and protection of the generated information, and the use of a methodology that meets international standards and requirements.

A significant modernization of the national statistical system of Kazakhstan was conducted in 2012-2017 jointly with the World Bank within the framework of the «**KAZSTAT**» project.

According to the World Bank expert group's opinion, the next important step will be the creation of an **integrated national data system** (data ecosystem), in which **one of the key roles** is assigned to the **National Statistical Office (NSO)**.

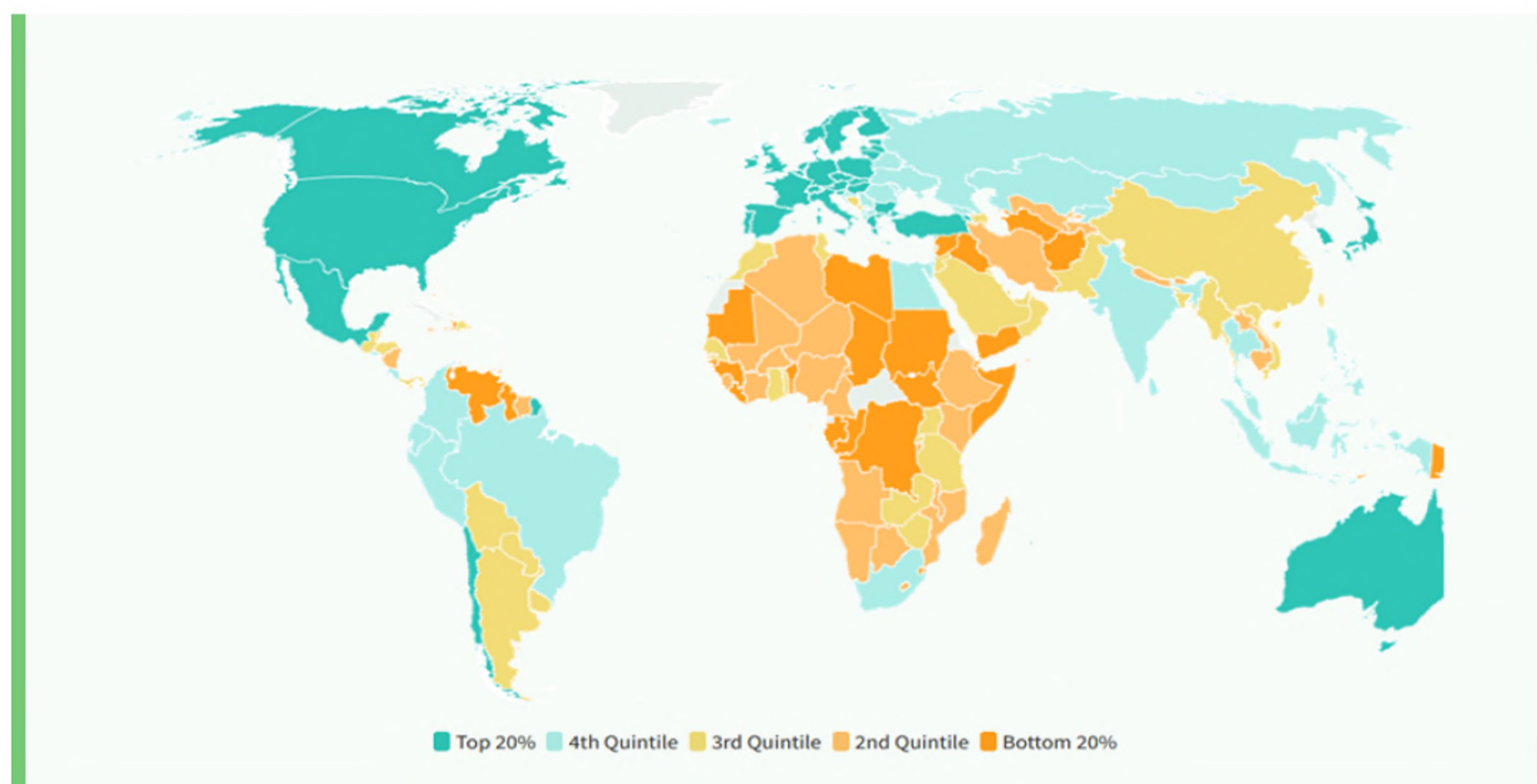


Figure 5. Statistical Performance Indicators (SPI) for 174 countries

The creation of an integrated national data system is an urgent task that meets modern needs and opportunities for the practical implementation of the Data-Driven approach in public administration. In turn, **an effective data management model** will significantly improve the quality of management processes.

Depending on who is assigned the role of a data manager, the International Association DAMA identifies three **types of data management models: centralized, distributed and hybrid models**.

[1] <https://www.worldbank.org/en/programs/statistical-performance-indicators>

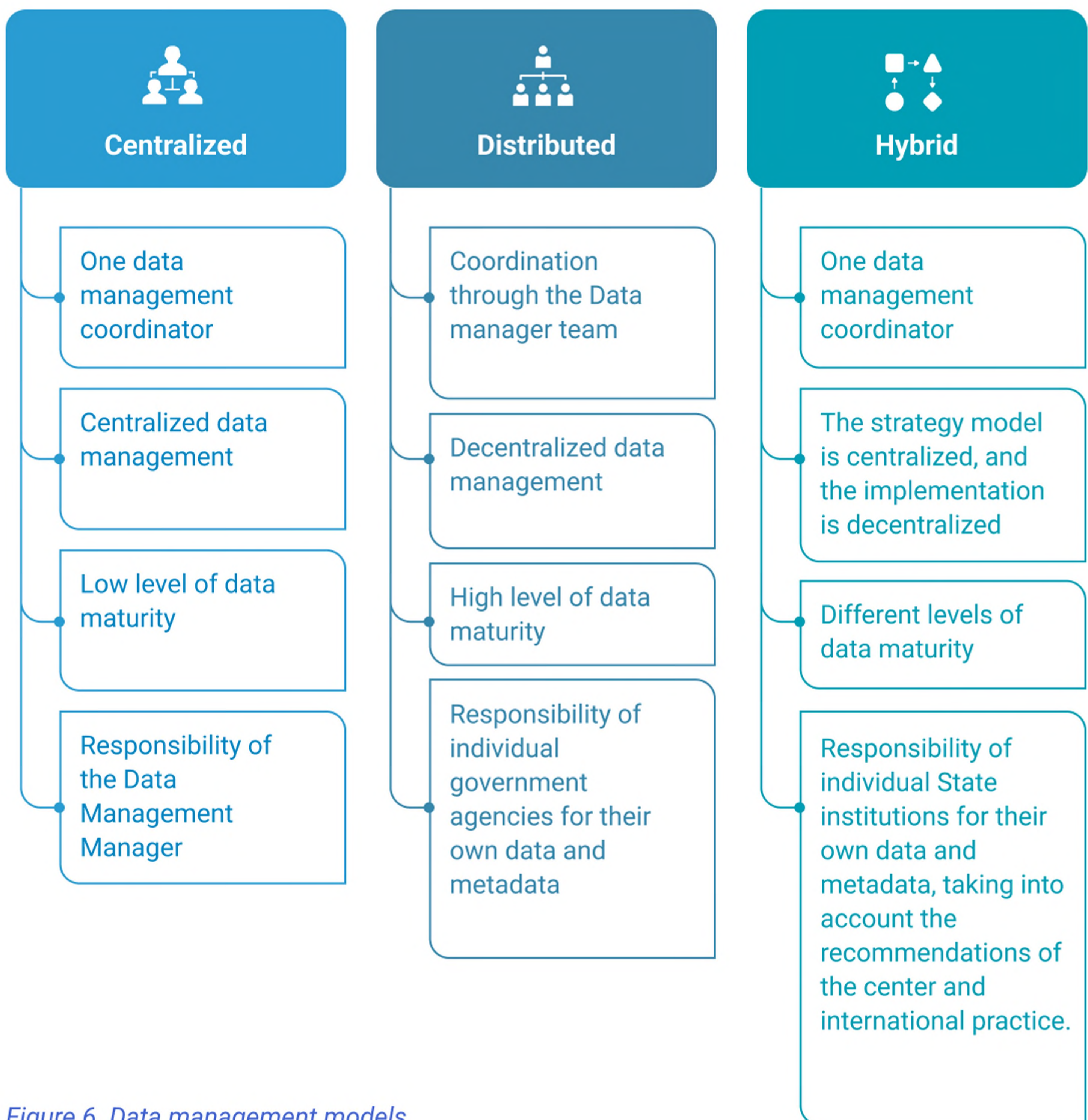


Figure 6. Data management models

According to the **OECD Digital Government Index 2020 Report**[1], «only 12% of countries have a **single special data policy** (or strategy), and **82%** implement data as part of a broader policy (for example, digital government or open data) ».

At the same time, many countries (USA, Canada, UK, Germany, New Zealand) have already **developed national data strategies** in order to bring together various aspects within a single policy instrument, ranging from data access to data sharing and ethics of data use.

[1] <https://www.oecd.org/gov/digital-government-index-4de9f5bb-en.htm>

In international practice, there are several approaches to the implementation of various data management models. The experience of the countries that are most **actively implementing** data management at the **national level** is presented below.









Name of the country	Management model	Data Owner	Coordinator	The role of the NSO
 Lithuania	Centralized	GA	NSO	Leadership role
 Great Britain, New Zealand	Distributed	GA + Information Group	NSO (Chief Government Data Steward)	Leadership role
 Australia	Distributed	GA	Customer Service Department	No data
 Canada	Distributed	GA	NSO Data Director (CDO)	No data
 Russia	Distributed	GA	NSO	Leadership role
 Finland	Hybrid	GA + Information Management Board	Expert knowledge in the field of data management	Membership in the Data Management Board
 Mexico	Hybrid	GA	Expert knowledge in the field of data management	Leadership role
 Estonia	Hybrid	GA	NSO	Leadership role

Figure 7. Comparison of data management models

Lithuania chose a **centralized** data management **model**, all state data is collected in one hand – the National Data Agency (formerly the Department of Statistics). The transition to a centralized model makes it possible to significantly change approaches to data management, create additional tools for collecting and using high-quality data.

The transformation into a National Agency allows the NSO to solve new tasks and become an effective state data management body that brings together data from different state institutions in one hand. This change provides a unified approach to data management, eliminates the duplication of collected data, and the quality control tools used significantly increase user confidence in the published information, and also creates ample opportunities for making the right political decisions.

The amendments to the Law on Official Statistics and State Data Management of Lithuania, adopted in 2022, establish a new approach to the management of state data and the regulation of its processing, reduce the administrative burden on institutions providing data, provide more flexible use of data collected in the state sector, as well as faster data exchange between state data systems. The law also defines the responsibilities of data providers – respondents, the ways of providing statistical data, expanded the scope of data sources due to Big data collected by private companies.

The United Kingdom, New Zealand, and Australia form a separate group and manage data using a distributed management model, which has a number of advantages, but also has some limitations in terms of system approaches to data management.

Mexico, Finland and Estonia have chosen a hybrid data management model for their countries, which has direct control over the production cycle of surveys conducted for the production of census statistical information, as well as statistical data obtained through the use of administrative records and other sources.

In terms of institutional structures and leadership roles, countries such as New Zealand, Finland, Australia, Mexico, Estonia preferred to define specific Chief Data Steward positions for the entire government, while others preferred to create bodies such as Data Councils, in which data leaders at the institutional level discuss priorities and agree on their coordinated implementation.

Each institution that provides information or collects statistical data of national interest has a coordinator who is responsible for implementing the policies and standards defined by the Data Management Commissioner. The coordinator, represented by the NSO, develops a methodology, requirements and standards that other state bodies should follow in order to obtain data when preparing statistical information.

For example, New Zealand created the position of the Government Chief Data Steward (GCDS), recognizing that strong system leadership is required to help New Zealand realize data as an asset. The Public Services Commissioner has appointed the head of Statistics New Zealand (Stats NZ) to this role.

In conjunction with the GCDS role, Stats NZ serves as the lead data agency within the New Zealand Government Data System, facilitating and providing a unified approach to data-related opportunities and challenges. In this capacity, Stats NZ, together with GCDS, strives to support agencies to maximize the potential of their data and ensure its effective use, while maintaining the trust of New Zealanders.

The Digital Government Leadership Group (DGLG) is led by the Chief Government Data Administrator and the Chief Government Digital Officer.

In September 2022, the «Law on Data and Statistics» came into force, prepared on the basis of a review of legislation in the field of statistics and public consultations on high-level proposals for new data.

In Mexico, **the National Institute of Statistics and Geography (INEGI) is an autonomous body**, completely independent of the federal government. INEGI plays a dual role in the national information system: as a direct producer of official information and as a coordinator setting norms and standards for the production of official statistics in various state institutions.

The Canadian Government gives priority to innovation and increasing «horizontality». The country has created the position of Director of Data Management, who is responsible for providing overall management of data and information management, developing administrative frameworks and management and quality standards.

A new ecosystem of state data and a well-built data management model allows to quickly collect, combine and visualize data collected in various registers and information systems, in accordance with the task being solved.

Thus, in almost all countries, the key role of the coordinator of data management policy is assigned to NSOs who have legislative access to personal data, appropriate powers and skills in working with data.

In the context of Kazakhstan, the experience of statistical activity and the existing powers of the authorized body in the field of state statistics can identify the BNS as a key player in the field of data management policy.

The background of the top half of the page is a dark blue field filled with a complex network of light blue and white lines. These lines radiate from the right side towards the left, creating a sense of depth and connectivity. Interspersed among these lines are various numbers in white and light blue, some appearing as small data points and others as larger, more prominent values. The overall effect is that of a digital data landscape or a network graph.

05

DATA MANAGEMENT MODEL IN KAZAKHSTAN

Data management model in Kazakhstan

Based on the presented international experience of implementing three data management models, as well as taking into account the current state of the national statistical system and the development of state information systems, a **hybrid data management model** is proposed for Kazakhstan.

The hybrid model is a type of data management model with **centralized control and distributed management**. In other words, in the hybrid model, the management strategy is centralized, and the implementation is decentralized, allowing to consolidate the advantages of centralized and distributed models, while significantly reducing their individual disadvantages.

For an ecosystem with a **very different level of maturity** of information systems, a hybrid model is the **most suitable option** for a nationwide approach.

Decisions regarding standards and shared data are made centrally. Processes, standards, guidelines and systems are usually set by the core team, task force or board, and individual participants are responsible for their application at the local level.

The model may include the adoption of a common data model by various organizations, and may also adopt a common approach for the entire ecosystem (inter-agency) to serve as an answer to a data problem in a particular area.

The main advantages of the hybrid model:

- State bodies are responsible for their own data.
- It provides autonomy for the development of standards, policies and procedures at the level of individual participants, but requires that they comply with the nationwide data exchange model to ensure interoperability.
- The solution of problems is carried out on the principle of «bottom-up» (there should be a clear responsibility and process for solving the issue).

The principle of the hybrid model requires the establishment of norms (laws, administrative regulations) necessary for the reuse of data. The adoption of a hybrid model may require **legislative changes** to avoid multiplying negotiations to obtain resources in each case. This includes creating institutional mechanisms and clarifying the roles needed for this model to work.

If this model is adopted, there will be need a **data steward** at the national level.

Data management is defined as the activity for the implementation of governing and control powers (planning, monitoring and enforcement) in relation to the management of information assets.

The experience of statistical activity and the existing powers of the authorized body in the field of state statistics can determine the **BNS as a key player in the field of data management policy**.

Today, the BNS **is the only state body** with significant experience in the formation of official statistical information, which has expert knowledge in the field of data ethics, high maturity in relation to data, legislative access to personal data and resources. BNS has modern methods of quality assurance and data protection.

Thus, given the extensive experience in the production of statistical data and indicators, data management, access to them and the practice of sharing them in the public sector, the national statistical system represented by the BNS should become the manager in the data ecosystem.

The role of the data management coordinator is key and consists in defining and coordinating the data management policy. Treating state statistics as a public good, the BNS will reconsider approaches to data management policy.

The Data Management Coordinator has complete information about where the data is located, where it comes from, what it means, who uses it and how it can be provided.

The main task of the data management coordinator is to build interagency coordination that ensures the use of a unified methodology, requirements and standards to provide the quality of state data.

Data demand monitoring is executed by the data management coordinator, by conducting a needs analysis.

The role of the IT management coordinator is to define and coordinate the IT management policy, as well information security part.

The IT management coordinator is responsible for processes related to digitalization. The focus of his attention is on updating processes through the implementation of digital solutions, including working with real data and a strategic view on the areas of knowledge in data management.

The main task of the IT management coordinator is to build processes and implement software solutions in such a way as to provide quick and secure access of the organization's employees to the necessary data.

IT Management Coordinator is an authorized body in the field of informatization in the Republic of Kazakhstan.

The role of the security coordinator is to prevent the leakage of data, access to which is restricted by regulatory legal acts, including information that constitutes official information of limited distribution or state secrets.

The Security Coordinator is an authorized body in the field of security in the Republic of Kazakhstan.

Data owners are owners of state or alternative sources of information who make decisions about data in the subject area and are responsible for their safety, while at the same time ensuring the quality of data.

Data management roles can be defined both at the functional and individual levels. At the same time, the names of the roles, their relative importance and necessity may vary.

When developing a national data ecosystem at the state level, **coordinated interaction of three key roles is necessary.**

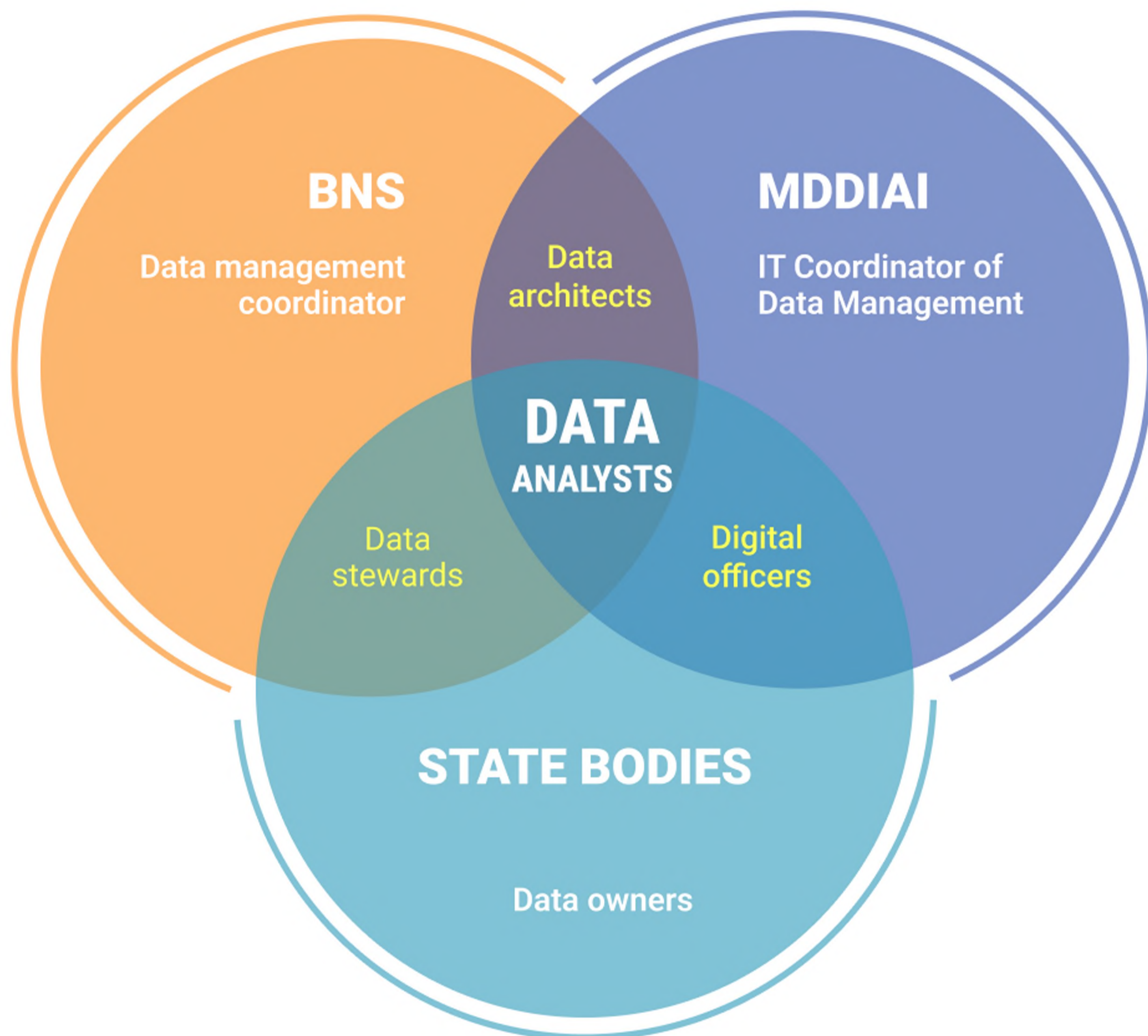


Figure 8. Architecture of interaction between participants of the state data ecosystem

Data Steward is a data management expert who coordinates the data management system of the national register^[1] and has access to information systems determined by the Government of the Republic of Kazakhstan within the framework of the list of national registers. In the area of responsibility of the data steward is the supervision and management of data, as well as ensuring its quality.

[1] Data stewards will be assigned to national registries, and not to government bodies, which are regularly reorganized. More details about national registers are set out in the section "Goals and Objectives"

Data Architect is an employee who develops the architecture of data collection, storage and unloading, determines the data source, sets requirements for data management policy.

Digital Officer is a manager responsible for digital transformation, defining the digitalization strategy, ensuring coordination with government digital platforms and projects.

Digital Officer is a manager responsible for digital transformation, defining the digitalization strategy, ensuring coordination with government digital platforms and projects.

[1] Data stewards will be assigned to national registries, and not to government bodies, which are regularly reorganized. More details about national registers are set out in the section "Goals and Objectives"

Basic principles

As for now, the national statistical system of Kazakhstan is based on **the fundamental principles of official statistics** adopted by the **United Nations[1]**. At the same time, in order to develop a national data ecosystem, should also be taken into account **the DMBOK[2] data management principles**.

Based on these two systems that do not contradict each other, the following 5 principles have been formed that recognize **data as an asset**: availability (inclusivity), safety (confidentiality), single entry (the «only once» rule), efficient use and timeliness.

A

Availability

The data ecosystem should ensure equal access to data on an impartial basis in the interests of realizing the right of citizens to publicly available information. The use of data should, in turn, bring a fair share of value to ecosystem participants.

S

Safety

Personal data collected by state bodies, regardless of whether they relate to individuals or legal entities, must be strictly confidential and used exclusively for the purposes established by law.

S

Single entry

The use of existing data should be managed effectively, the collection of additional data should not take place if there is a state body that has already collected the necessary data. This principle provides an incentive to establish common rules and standards that promote interoperability.

E

Efficient use

The demand for data is a qualitative assessment of the degree of their usefulness to users and, accordingly, the value that is extracted from their use. Given the possibility of reuse of data, their value is directly proportional to the amount of their use. The more users a particular data set has, the greater its value.

T

Timeliness

The relevance (timeliness) of information is the degree to which information corresponds to the current moment in time. The relevance of information is determined by how timely and important this information is for a person or society, whether it can be used in a specific situation to solve problems or make decisions.

Figure 9. Basic principles of data management

[1] <https://unstats.un.org/fpos/>

[2] <https://www.dama.org/cpages/body-of-knowledge>



07

GOALS AND OBJECTIVES

Goals and objectives

The main goal of this Concept is to **unleash the potential of the national data ecosystem** based on the principles listed in the previous section.

The goal will be achieved by implementing the following tasks:

TASK 1

IMPROVEMENT OF THE LEGAL AND METHODOLOGICAL FRAMEWORK

To ensure access to data, there is a need for a legal framework in the field of state data management, legislative regulation of requirements and access, the order of interaction of information systems, rules for creating a model of state data, the definition of reference data, bringing state data into compliance with uniform data management requirements. These initiatives will be reflected in the Draft Law «On State Statistics and Data».

In addition, in order to ensure the quality and reliability of data there is a need for a **methodological «reboot»** of the national statistical system and all state bodies.

As for now, to ensure the quality of methodological support for statistical activities there was established an Expert Council and the Department of Methodological Coordination and Analysis within the structure of the BNS. Particular attention is paid to the issues of unification of methodological approaches and overcoming methodological inconsistency between agencies.

An important element in the implementation of this task will be the **Global Assessment** of Kazakhstan's Statistical System by the **UNECE** in **early 2023**. Additionally, an **in-depth review of the OECD** will also be conducted over 3 years (2023-2025) with further obtaining the status of a «member of the OECD Committee on Statistics and Statistical Policy».

The «reboot» of the methodology will be the first stage of the internal transformation of the BNS and will contribute to increasing confidence in official statistics.

TASK 2.

ACTIVE USE OF ADMINISTRATIVE AND ALTERNATIVE SOURCES BASED ON NATIONAL REGISTERS

This task will require the modernization of information systems of state bodies, as well as the expansion of computing capacities of the BNS, automation of all internal processes of data collection and processing, a change in the IT architecture of the state apparatus and approaches to data management in accordance with international data management standards (DAMA).

In order to realize the synergy effect between different sources and the full use of administrative and alternative data, there will be required a **unified data management policy and new targeted data architecture**. They will allow to develop a common approach to data management and unify the methods of their processing and analysis.

At the same time, in order to improve the quality of these administrative sources, it is necessary to create **centralized national registers** for all spheres of the socio-economic development of the country.

National registers will be reference databases that store only clean data that has passed format-logical and desk controls, thus solving the existing problem of data quality and reliability.

National registers will become a base for public services, as well as an interface for interaction between government agencies, business structures and individuals as the only source of reliable data. Thus, the issue of duplication of data in state databases of different departments will be resolved, multiple administrative procedures of interaction with state bodies will be excluded. It will become the basis of debureaucratization of the state apparatus.

National registers will allow the formation of **digital statistical registers** by collecting reliable data in a pre-agreed format with certain regularity. This will significantly **reduce the burden on respondents and increase the accuracy of statistical production**.

Currently, the BNS together with the Ministry of Digital Development, Innovations and Aerospace Industry (MDDIAI) have started the project **«Transition to digital statistical registers»**. It involves the revision of all existing statistical indicators, the analysis of existing databases of government agencies and reporting forms, as well as the formation of fully digital statistical registers, which will be filled on the basis of administrative data. The project will also make it possible to compare existing data in official statistics and in existing state databases and, together with state bodies, improve the quality of data.

TASK 3.

STRENGTHENING THE ANALYTICAL CAPACITY OF STATE BODIES

In order **to strengthen analytical potential** and conduct deep business analytics, it is necessary to form an analytical culture, which includes the formation of a library of indicators and standardization of visualization, the formation of empowered to optimize business processes, the creation of a **competence center to improve data literacy**.

In addition, there is needed a state policy with a long-term focus on strategy, culture, ethics, roles and conditions, as well as the capabilities of people who can maintain an effective data ecosystem.

It is necessary to assess how effective the current recruitment system is to meet the special needs of the national statistical system and the development of the data ecosystem, on the basis of which further measures should be taken to adapt the recruitment process to the needs.

In this context, the BNS has launched the **Data Literacy** initiative, which aims **improving the literacy** of civil servants in working with data. It is not enough just to equip employees with analytical tools (applications), it is necessary to teach them how to work with data, analyze information and draw conclusions from it.

The paradigm shift towards **Data-driven decision making** (from the Decision-driven data making that often takes place in practice) requires training users to be critical of the results obtained, as well as to justify decisions based on data.

For mass coverage of the Data Literacy program, the BNS forms a **series of online training** lessons with control tasks that will be posted on the **training platform** and in the future will become **mandatory for all state employees**.

In addition, by the end of 2022 in partnership with the **Asian Development Bank** there will begin training of BNS staff in more advanced business intelligence tools, including BI tools and immersion in Python-based Data science, within the framework of the **Knowledge and Experience Exchange Program**.

In 2023, it is planned to train key personnel of the BNS to DAMA standards and pass the **Certified Data Management Professional CDMP certification**.

Thus, a **competence center** will be formed on the basis of the BNS, which on a renewable basis will continue further training of employees of state bodies on its own.

The noted training programs will also cover **data stewards** to improve their skills and form a unified vision for the development of the data ecosystem. Trained data stewards will subsequently apply the acquired skills when working with data from national registers.

The analytical potential of the BNS staff will also be developed through **research activity** together with universities. Joint research will be conducted on topics of interest to both Bureaus and universities. The Bureau presents on the website a **list of research topics** aimed at solving current research needs. The Bureau will also be actively involved in research work with the teaching staff of universities on topics of interest to them.

Along with the provision of databases in a de-identified form, methodological assistance with possible co-authorship of publications will be provided within the framework of the mentioned areas.

In addition, **the mechanism of interaction** between the BNS and universities in terms of providing depersonalized microdata will be **systematized** and an **information service for researchers** will be launched, which will reflect the list of data recipients and links to research results. This will increase the transparency, coordination and effectiveness of research activities based on BNS data.

TASK 4.

ENSURING ACCESS TO DATA AND CREATING A TRANSPARENT COMPETITIVE DATA MARKET

An important direction in the development of the data ecosystem is to ensure **equal access to data**. Mass use of data will increase their value, as well as fully implement the practice of making decisions based on data.

For these purposes, the BNS forms services for simplified access to data. In particular, **interactive dashboards** are being developed for all major statistical indicators. In the future, it is planned to create a **researcher's office** and a **virtual data laboratory** for the comprehensive use of microdata by accredited users.

As part of this direction, users' access to microdata will be **expanded by law**. Access to depersonalized microdata makes it possible to analyze complex problems taking into account the diversity of the population and business, to develop **micro-modeling tools to assess the potential impact** of changes in government policy and macroeconomic shocks.

One of the key partners in the dissemination of data and their mass application are scientific organizations and higher educational institutions that have legislative access to depersonalized microdata. To promote the culture of working with data among students, are being created joint projects, competitions and research using microdata.

Along with statistical data, it is important to provide users with access to administrative data available in central state and local executive bodies. Often administrative data is not available to users and is applied only once, which reduces the potential effect of the collected data.

Disclosure of data from alternative sources generated by the private sector has significant potential. Their active use will allow us to get a more complete picture of the socio-economic situation.

To ensure access to these combined data sources, the **role of a data market regulator** is important, which will ensure coordination and **clear rules** of the game in order to simultaneously preserve the **inclusiveness and confidentiality** of data. In this regard, it is necessary to create a regulatory field for the **balanced development of the data ecosystem**.

An important driver of the development of the data ecosystem will be the ability to **monetize data**. In this regard, as a tool for commercializing the services of a data platform, there will be created environment for setting up a universal **data exchange**. It will allow various market participants to make transactions with data and data products. Technologies for data exchange and analysis will be available at the site, while the site will offer ready-made data monetization scenarios. The exchange should become the core of the data ecosystem, which will subsequently allow the development of a national data market.

In the long term, the implementation of these tasks will create a **full-fledged data market**. Data market participants will be able to **mutually enrich** and monetize data, thereby increasing profitability, and the state will be able to make more effective decisions based on **combined data sources**.






08

**EXPECTED
RESULTS**

Expected results

The transformation of the national statistical system by the end of **2025** will result in the creation of a full-fledged **national data ecosystem**, the benefits of which will benefit **all users of the data market**, including the public sector, citizens and businesses.

EXPECTED EFFECTS FOR STAKEHOLDERS

	 State	 Citizens	 Business
Collection of reports and provision of services	Optimization of statistical observations by 2 times (2022 – 146 forms)	Proactive provision of public services	Load reduction by eliminating duplication of reporting
Disclosure of data	Increasing the value of data by disclosing and reusing data	Compliance with the requirements of personal data security and information confidentiality	Opening a niche for effective business by eliminating information asymmetries
Data usage	Increase in planning accuracy, speed and quality of decisions made	Lots of new data-driven services and products	Introduction of new data products and legitimate data monetization opportunities



09

STAGES OF IMPLEMENTATION

Stages of implementation

The tasks proposed in the framework of the Concept will be implemented in **two stages within 3 years**. Below are the activities planned for implementation in this period (2023-2025).

Nº	Events	2023	2024-2025
1	Task 1. Improvement of the normative legal and methodological framework		
1.1	Development of the Law «On State Statistics and Data» and relevant NLA	■	
1.2	Improvement of methodological requirements for data management audit	■	
1.3	Development of Methodology for data quality assurance and control	■	
1.4	Conducting an audit of information systems of state bodies	■	■
1.5	Global assessment of the Statistical System of Kazakhstan by the UNECE	■	
1.6	Review of the OECD with further obtaining the status of a «member» of the OECD Committee	■	■
1.7	Implementation of the OECD and UNECE recommendations	■	■
2	Task 2. Active use of administrative and alternative sources based on national registers		
2.1	Development and approval of the target data architecture for the country	■	
2.2	Development of a data catalog and standardization of metadata	■	
2.3	Creation of national registers based on state databases and improvement of their quality	■	
2.4	«Reboot» and training of data stewards	■	
2.5	Use of alternative sources of information	■	■
2.6	Digitalization of internal processes of statistical activity	■	
2.7	Updating the information system «e-Statistics»	■	■
2.8	Expansion and redistribution of production capacities intended for data storage and processing	■	■
3	Task 3. Strengthening the analytical capacity of state bodies		
3.1	Training and development of competencies of civil servants	■	
3.2	Formation of research competencies and technological practices	■	
3.3	Development of specialized professional development programs	■	■
3.4	Conducting joint research with universities	■	■
3.5	Creating a new mechanism for providing depersonalized data and launching an information service for researchers	■	■
4	Task 4. Ensuring access to data and creating a transparent competitive data market		
4.1	Providing access to microdata for research and implementation of analytical solutions	■	■
4.2	Creating a Data Exchange		■
4.3	Preparation for the transition from traditional to register-based census		■

Abbreviations



IMF	The International Monetary Fund
DAMA	Data Management Association
BNS	The Bureau of National Statistics of the Agency for Strategic planning and reforms of the Republic of Kazakhstan
MDDIAI	The Ministry of Digital Development, Innovations and Aerospace Industry of the Republic of Kazakhstan
CSB	Central State Bodies
LEB	Local executive bodies
NSO	National Statistical Office
IS	Information systems
UIP	Unified information platform
IS	Information Security
CPI	Consumer Price Index
GDP	Gross domestic product
OECD	Organisation for Economic Co-operation and Development
GCDS	Government Chief Data Steward
Stats NZ	Statistics New Zealand
DGLG	Digital Government Leadership Group
INEGI	National Institute of Statistics and Geography
DMBOK	Data Management Body of Knowledge
UNECE	United Nations Economic Commission for Europe
RLA	Regulatory legal act