УДК 339.732:004.6
DOI: https://doi.org/10.30838/EP.198.53-62

Korytska Olha
PhD. in Economic Sc.
Lviv Polytechnic National University
Корицька О.І.
кандидат економічних наук
Національний університет «Львівська політехніка»
https://orcid.org/0000-0002-4852-188X

Kursanova Anna
Lviv Polytechnic National University
Курсанова А.Г.
Національний університет «Львівська політехніка»
https://orcid.org/0009-0008-9936-693X

WORLD BANK OPEN DATA: A COMPARATIVE ANALYSIS OF MULTI-DATABASE COVERAGE

This study analyzes data coverage in five World Bank databases (World Development Indicators, Doing Business, Gender Statistics, Statistical Capacity Indicators, and Millennium Development Goals) across seven countries: Australia, Burundi, Haiti, Spain, Mexico, Turkmenistan, and Ukraine. The study employs statistical variation metrics to evaluate data completeness and identify information gaps.

The analysis reveals significant data coverage heterogeneity, particularly in databases with over 200 indicators spanning more than 20 years. The findings indicate that the presence of indicators does not guarantee data availability, with some metrics having minimal or single-year entries. The World Development Indicators and Gender Statistics databases show coverage of up to 32–35 units out of 63 possible units.

The study concludes that although the World Bank platform provides extensive resources, data completeness remains a significant challenge. The authors attribute these gaps to factors including economic development levels, industry structures, and historical circumstances. They recommend using standard indicators for research while consulting national statistical offices for specialized data needs.

Keywords: *open science, World bank database, data completeness, socio-economic indicators, comparative analysis.*

JEL classification: *C80, O19.*

ВІДКРИТІ ДАНІ СВІТОВОГО БАНКУ: ПОРІВНЯЛЬНИЙ АНАЛІЗ НАПОВНЕНОСТІ БАЗ ДАНИХ

У статті представлено комплексний статистичний аналіз наповненості відкритих баз даних Світового банку для різних країн. Дослідження охоплює п'ять ключових баз даних Світового банку: «Індикатори світового розвитку», «Ведення бізнесу», «Гендерна статистика», «Статистичні показники ефективності» та «Цілі розвитку тисячоліття». У роботі аналізується повнота та якість даних для семи обраних країн: Австралії, Бурунді, Гаїті, Іспанії, Мексики, Туркменістану та України.

Обчисливши статистичні показники варіації, оцінено закономірності наповненості даними та виявлено прогалини в інформаційному забезпеченні. Методологія дослідження полягала в обчисленні та порівнянні статистичних показників, зокрема середніх значень, розмаху варіації, моди, медіани, середньоквадратичного відхилення та коефіцієнта варіації для кожної бази даних та країни.

Результати дослідження вказують на значну неоднорідність у наповненні даних, залежно від бази даних та країни. За результатами статистичного аналізу та візуалізації кумуляти виявлено, що найбільш нерівномірний розподіл даних демонструють великі бази даних, що містять понад 200 показників та охоплюють періоди понад 20 років. Встановлено, що наявність показників у базах даних не гарантує доступність даних, деякі показники мають мінімальні або однорічні записи. Бази даних «Індикатори світового розвитку» та «Гендерна статистика» демонструють наповненість даними до 32–35 одиниць із 63 можливих для обраних країн.

Цифрова платформа Світового банку пропонує широкі інформаційні ресурси, але при цьому повна інформація за бажаними показниками є не завжди наповнена на 100%. Ці прогалини можуть бути спричинені різними факторами, включаючи економічний розвиток країн, специфіку галузевої структури, деталізацію гендерної

© Корицька О.І., Курсанова А.Г., 2025

статистики, історичні обставини та відмінності в методології збору даних. У дослідженні рекомендується використовувати загальноприйняті показники для наукового пошуку, звертатися до національних статистичних служб для отримання спеціалізованих або регіонально-специфічних показників.

Ключові слова: *відкрита наука, база даних Світового банку, повнота даних, соціально-економічні показники, порівняльний аналіз.*

Problem Statement. In the era of Industry 4.0, global international organizations actively develop and maintain their own digital online repositories of open data. This trend underscores the need for a thorough examination of the completeness and quality of such datasets. Amidst global digitalization, researchers, analysts, government officials, educators, and other stakeholders increasingly rely on open databases to obtain relevant indicators based on specific search criteria. However, empirical observations reveal certain gaps and inconsistencies in the systematic provision of information for selected indicators, particularly in the case of the World Bank databases [1].

The World Bank's open databases contain approximately 10,000 indicators and are regarded as one of the most authoritative and comprehensive online resources for conducting research across various socio-economic fields. Nevertheless, users frequently encounter challenges related to missing data for specific indicators in particular countries or the insufficient completeness of time series data.

Considering these factors, a comprehensive analysis of the completeness of the World Bank's open databases across countries with varying economic development levels and geographic locations is both valuable and timely. The findings of such an analysis hold significant practical relevance for the academic community and participants in the research and educational processes, facilitating more efficient data retrieval and collection for empirical studies.

Review of recent studies and publications. This article continues the authors' research on the digital capabilities of statistical databases, particularly those of the Main Statistical Office in the Lviv Region [2] and Eurostat [3]. The research landscape on World Bank databases remains relatively specialized, with a limited number of academic publications dedicated to this subject. Most existing studies focus either on specific thematic indicators or sectors or on technical aspects related to the functionality of such databases.

For instance, John R. Hahn et al. [4] analyzed the Water and Sanitation database of the World Bank, evaluating its structure, indicators, and shortcomings. Their findings revealed issues related to data quality and completeness, which hinder statistical processing, and they proposed recommendations for improving the database's structure. Similarly, Galbraith et al. [5] compared several global inequality datasets, including those of the World Bank, and found that the latter has limited suitability for cross-country comparative analysis due to inconsistencies in data standardization.

Moreover, Q. Ran and Jie Zhang [6] investigated networked databases, comparing various approaches to storing and processing large-scale data. Another study [7] examined the use of the World Bank API for accessing open data, detailing how users can integrate these data sources into their own analytical applications. Additionally, Yin Lin, Yifan Guan, Abolfazl Asudeh, and H. Jagadish [8] explored data coverage issues in multidimensional databases, which is highly relevant to the World Bank's open data. They proposed algorithmic methods for assessing data sufficiency in complex multi-database environments.

However, studies on the completeness and coverage of the World Bank's open databases remain limited, especially regarding cross-country comparisons based on economic development levels and geographic distribution. This gap highlights the need for a more systematic and comprehensive evaluation of these databases.

Objective of the study – the objective of this study is to conduct a statistical analysis of the data coverage within the World Bank database, focusing on five datasets and seven countries, utilizing digital analytical tools. Additionally, the study aims to evaluate the digital capabilities of the World Bank's online platform.

A comparative analysis of the World Bank's database completeness will help assess both the organization's capacity for data collection and the systematic nature of data gathering within specific countries. For this research, the selected countries are Australia, Burundi, Haiti, Spain, Mexico, Ukraine, and Turkmenistan. Another key objective of the study is to determine whether the extent of data availability is influenced by a country's geographical location and economic development level.

Presentation of key research findings. The World Bank's data repository consists of 86 databases; however, the majority of these databases cover only a single continent or a specific region. Therefore, for the purposes of this empirical study, it is most appropriate to select more comprehensive databases, of which there are no more than 10 on the World Bank's online platform. For this research, five databases out of the 86 available have been selected: World Development Indicators (WDI), Doing Business (DB), Gender Statistics (GS), Statistical Capacity Indicators (SCI), Millennium Development Goals (MDGs).

Analysis of data coverage and key observations.

The study focuses on analyzing the completeness of these databases for the selected countries. The results of this analysis are presented in table 1.

The table provides both absolute and relative values of data completeness across selected indicators. Based on the analysis, we observe the following:

1. The WDI database contains the highest number of indicators, ranging from 1 to 64 per country. The country with the most recorded indicators is Mexico (1,416 indicators), while Turkmenistan has the fewest (932 indicators).

2. Unlike other databases, the Doing Business database includes data from 191 countries, whereas other databases cover over 200 countries: World Development Indicators – 266 countries, Gender Statistics – 265 countries, Statistical Capacity Indicators – 217 countries, Millennium Development Goals – 263 countries.

3. In the Gender Statistics database, Ukraine has the highest number of recorded indicators (935), while Turkmenistan has the lowest (558).

4. The Statistical Capacity Indicators database contains the fewest indicators (72 total) compared to the other selected databases. However, it exhibits the highest level of data completeness across countries in comparison with the other datasets.

Table 1

Comparison of database coverage by country

Database	Australia	Burundi	Haiti	Spain	Mexico	Turkmenistan	Ukraine	Total
Indicators in the database, units (% of the declared number of indicators in the database)								
WDI	1122 (74.90)	1320 (88.12)	1252 (83.58)	1189 (79.37)	1416 (94.53)	932 (62.22)	1405 (93.79)	1498 (100)
DB	194 (100)	191 (98.45)	187 (96.39)	194 (100)	194 (100)	0	194 (100)	194 (100)
GS	702 (60.88)	875 (75.89)	864 (74.93)	732 (63.49)	833 (72.25)	558 (48.40)	935 (81.09)	1153 (100)
SCI	72 (100)	72 (100)	70 (97.22)	72 (100)	72 (100)	70 (97.22)	72 (100)	72 (100)
MDGs	90 (68.18)	107 (81.06)	102 (77.27)	81 (61.36)	95 (71.97)	81 (61.36)	98 (74.24)	132 (100)

Source: calculated by the authors based on [1].

To fully understand why these databases were chosen, it is essential to examine their content in greater detail–specifically, the extent of their indicator coverage and the scientific relevance of these indicators.

Overview of the World Development Indicators database.

The WDI represent the primary collection of development indicators compiled by the World Bank from officially recognized international sources. This dataset provides the most recent and accurate available data on global development, as well as national, regional, and global estimates [1].

For this study, data from the period 1960 to 2023 have been utilized, with the maximum number of available indicators totaling 1,498. Below, table 2 presents the distributional analysis of the WDI dataset across the selected countries.

Table 2

Results of calculating variation indicators by country according to the World Development Indicators database

Indicator	Australia	Burundi	Haiti	Spain	Mexico	Turkmenistan	Ukraine
Number of observations	1122	1320	1252	1189	1416	923	1405
Mean	35.12	27.86	27.42	34.23	33.09	22.06	23.73
Range of variation	63	63	63	63	63	63	63
Mode	63	63	63	63	63	1	31
Median	33	24	24	32	32	23	24
Standard deviation	20.38	20.45	20.49	19.05	20.95	18.06	15.38
Coefficient of variation	0.58	0.73	0.75	0.56	0.63	0.82	0.56

Source: calculated by the authors based on [1].

Based on the calculations, we observe the following key findings:

1. Range of Variation – The range of variation is identical across all selected countries, equaling 63 units. This indicates a non-uniform data distribution, where the highest and lowest values are significantly distant from each other. This was expected from the beginning of the study, as an initial review of the dataset (generated in Excel) revealed numerous missing cells and inconsistencies in data availability. Many indicators contain gaps, with some having data only for a single year rather than a continuous time series. Furthermore, this pattern is consistent across identical indicators, suggesting systemic gaps in data collection.

2. Mode (Most Frequent Value) – Australia, Burundi, Haiti, Spain, and Mexico exhibit a modal value of 63 units, indicating that these countries predominantly have maximum data coverage across indicators. Ukraine's mode is 24, which is a relatively strong result, considering that the country gained independence only in 1991 and thus has a shorter historical record in global databases. Turkmenistan has the lowest mode (1 unit), indicating that for

most indicators, data are available for only a single year. Given that Turkmenistan also has the fewest total indicators (923) among the selected countries, this low mode suggests a negative trend in the completeness of World Bank data for this country.

3. Median Value – The calculated median values reveal the following distribution: Australia: 33 units (highest), followed by Spain and Mexico: 32 units. Ukraine, Burundi, and Haiti: 24 units and Turkmenistan: 23 units (lowest). These results are relatively strong, considering each country's economic structure, political system, and historical context.

4. Standard Deviation and Mean Comparison – There is noticeable dispersion in the data, indicating that values are widely spread around the mean. Specifically, for Australia, Spain, and Mexico, the difference between the standard deviation and mean is approximately 12–15 units. In contrast, Burundi, Haiti, Turkmenistan, and Ukraine exhibit smaller deviations of 4–8 units, suggesting more clustered data points around the mean.

5. Coefficient of Variation – The variation coefficient values indicate that data completeness is highly uneven across all selected countries, as the coefficient exceeds 0.33, confirming significant variability in data availability.

Figure 1 provides a cumulative visualization of data completeness across countries. The distribution shows that for Ukraine, Turkmenistan, Burundi, and Haiti, most indicators contain 32 or fewer data points. In contrast, Spain, Mexico, and Australia exhibit a more evenly distributed pattern of indicator coverage, resulting in smoother cumulative curves on the graph.

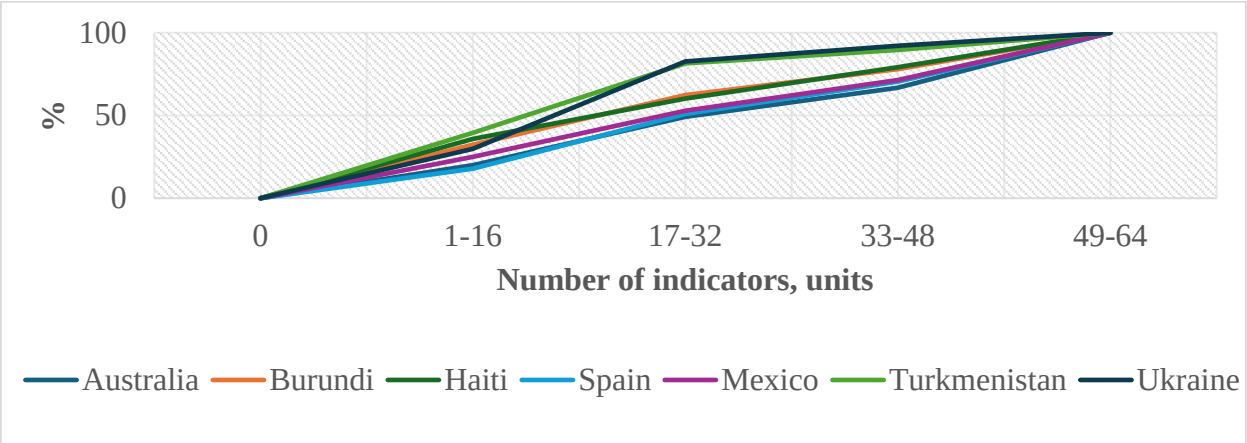

Fig. 1. Cumulative data coverage of indicators by country according to the World Development Indicators database
Source: calculated and visualized by the authors according to [1].

Analysis of the Doing Business database.

The Doing Business database contains data on objective regulatory measures affecting business operations and their enforcement across economies and selected cities at subnational and regional levels [1]. However, as of September 16, 2021, data collection for the Doing Business database was officially discontinued [9].

At the time of this study, the database covered the period from 2003 to 2019 and included a total of 194 indicators. Notably, among all countries and territorial entities available for selection on the World Bank website, Turkmenistan is absent from this dataset. Table 3 presents the results of the distributional analysis of the Doing Business database for the selected countries.

Table 3

Results of calculating variation indicators by country according to the Doing Business database

Indicator	Australia	Burundi	Haiti	Spain	Mexico	Ukraine
Number of observations	194	191	187	194	194	194
Mean	9.56	9.55	9.57	9.56	6.06	6.06
Range of variation	16	16	16	16	16	16
Mode	6	6	6	6	7	7
Median	9	9	9	9	6	6
Standard deviation	4.92	4.93	4.88	4.92	3.37	3.37
Coefficient of variation	0.51	0.51	0.51	0.51	0.56	0.56

Source: calculated by the authors based on [1].

The statistical analysis of the Doing Business dataset reveals the following key observations:

1. Range of Variation – The range is identical for all selected countries (16 units), indicating a non-uniform distribution of data. The highest and lowest values are significantly distant from each other. However, unlike the World Development Indicators (WDI) database, the gaps in data coverage occur specifically between 2003 and 2018, with the exception of 25–35 missing indicators.

2. Mode – Australia, Burundi, Haiti, and Spain have

a mode of 6 units, while Mexico and Ukraine have a mode of 7 units. This suggests that the majority of indicators for these countries contain less than half of the available data points in this dataset.

3. Median Value –The highest median values were observed in Australia, Haiti, Spain, and Burundi (9 units). The lowest median values were found in Ukraine and Mexico (6 units). These values are relatively high, considering that the Doing Business database covers the shortest time period among the analyzed datasets.

4. Standard Deviation and Mean Comparison – The data exhibits significant dispersion, with values deviating from the mean. In Australia, Spain, Burundi, Haiti, and Mexico, the difference is approximately 4–5 units, while in Turkmenistan and Ukraine, the difference is around 3 units.

5. Coefficient of Variation – The coefficient of variation suggests that the distribution of data coverage is inconsistent across all selected countries, confirming a high degree of variability in data availability.

Figure 2 visualizes the cumulative distribution of indicator completeness across the selected countries. The results indicate that for Ukraine, Burundi, Australia, Spain, and Haiti, most indicators contain up to 12 data points, and their cumulative distribution curves closely overlap, suggesting similar data availability patterns.

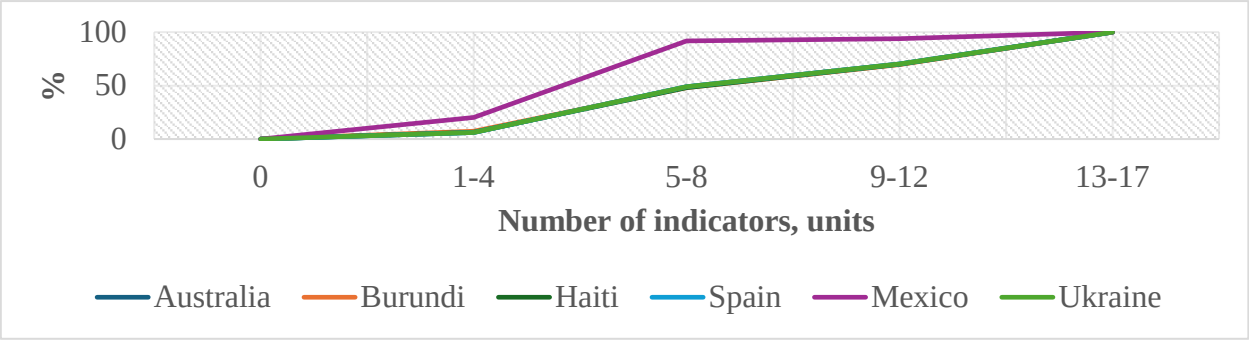

Fig. 2. Cumulative data coverage of indicators by country according to the Doing Business database
Source: calculated and visualized by the authors according to [1].

However, Mexico deviates significantly from this trend. The analysis shows that most indicators for Mexico are limited to 8 data points, a noticeably lower value compared to other countries. Moreover, while other countries exhibit a relatively uniform data distribution, Mexico's data coverage appears clustered around 5–8 units, which aligns with its mode value of 7 units.

Analysis of the Gender Statistics database.

The Gender Statistics database provides data on key gender-related topics. It covers multiple domains, including demographics, education, healthcare, labor force participation, and political engagement [1]. This database ranks among the three largest datasets on the World Bank's platform in terms of the number of available indicators: Education Statistics – 8,450 indicators, WDI – 1,498 indicators, GS – 1,153 indicators.

At the time of this study, the Gender Statistics database contained data spanning from 1960 to 2023, with a total of 1,153 indicators. Table 4 presents a detailed distributional analysis of data completeness within this dataset.

Table 4

Results of calculating variation indicators by country according to the Gender Statistics database

Indicator	Australia	Burundi	Haiti	Spain	Mexico	Turkmenistan	Ukraine
Number of observations	702	875	864	732	833	558	935
Mean	22.50	17.59	15.85	23.71	20.55	15.92	15.65
Range of variation	62	62	62	62	62	62	62
Mode	1	2	4	1	1	1	1
Median	12	4	4	18	8	3	4
Standard deviation	22.93	21.87	21.59	22.66	21.98	22.20	20.62
Coefficient of variation	1.02	1.25	1.36	0.96	1.07	1.39	1.32

Source: calculated by the authors based on [1].

Based on the obtained calculations, we can conclude the following:

1. Range of Variation – The range is identical for all selected countries (62 units), indicating a non-uniform data distribution. The highest and lowest values are significantly distant from each other, mirroring the pattern observed in the WDI database.

2. Mode – Australia, Spain, Mexico, Turkmenistan, and Ukraine have a mode of 1, meaning that most indicators for these countries contain data for only a single year. In Burundi, the mode is 2, while Haiti has the highest mode (4 units), indicating slightly better data availability. This suggests that across all selected countries, most indicators exhibit minimal data coverage.

3. Median Value – The highest median values were recorded for Spain (18 units), Australia (12 units), and Mexico (8 units). The lowest median values were observed in Ukraine, Burundi, and Haiti (4 units) and Turkmenistan (3 units). These values are relatively low, given the total number of indicators in the Gender Statistics database. However, this may be explained by the categorization of indicators, which are often split based on age groups or legal status categories. For example, the indicator "Women who own a home, either individually or jointly (% of women aged 15–49 years)" is further divided into: Q1 (lowest quintile), Q2, Q3, Q4, Q5 (highest quintile).

4. Standard Deviation and Mean Comparison – The data exhibits variation around the mean, indicating some degree of dispersion. However, for Australia, Spain, and Mexico, the mean and standard deviation are relatively close, suggesting a more stable data distribution compared to other countries.

5. Coefficient of Variation –The coefficient values indicate that data completeness remains uneven across the selected countries.

Figure 3 presents a cumulative visualization of data availability across countries.

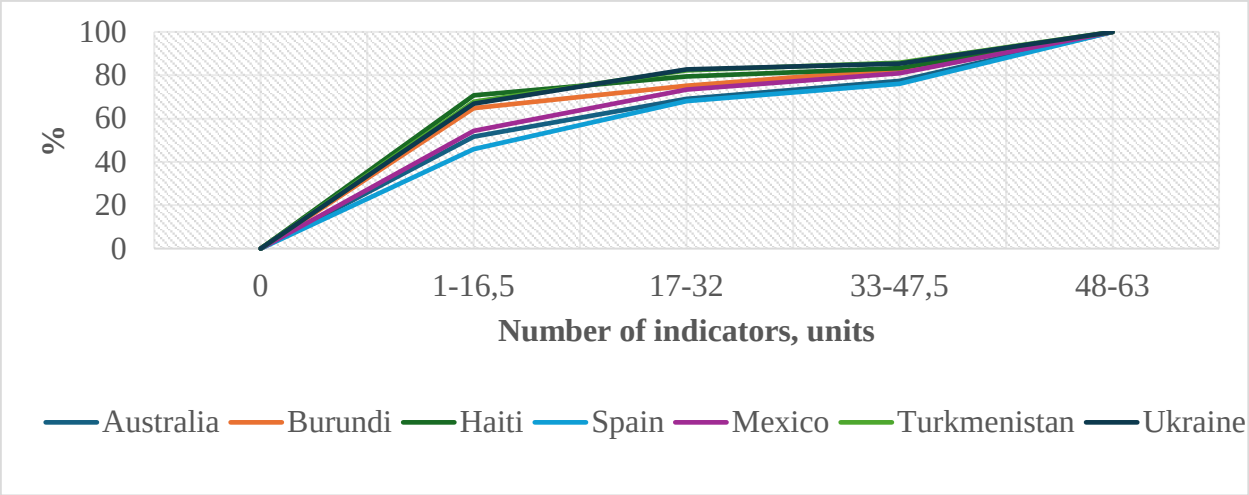

Fig. 3. Cumulative data coverage of indicators by country according to the Gender Statistics database
Source: calculated and visualized by the authors according to [1].

For all selected countries, the graph lines differ, yet they exhibit a similar trend: in Ukraine, Turkmenistan, Burundi, and Haiti, most indicators contain data for up to 16 units. In Spain, Mexico, and Australia, the cumulative data availability extends up to 32 units. This level of data completeness is relatively low. However, considering that many indicators are distributed based on age, gender, legal status, and, in some cases, are tied to specific continents, this result is logical and expected.

Analysis of the Statistical Capacity Indicators database.

The Statistical Capacity Indicators database is classified as publicly accessible according to the World Bank's Access to Information Classification Policy. This means that users both within and outside the World Bank can access this dataset freely [1]. At the time of this study, the dataset covered the period from 2004 to 2022 and included 72 indicators.

The SCI framework serves as a benchmark for measuring progress in the development of statistical capacity and related investments. It evaluates five key dimensions: data use, data services, data products, data sources, data infrastructure.

The World Bank team makes ongoing efforts to ensure the accuracy of the data presented in the SPI indicators. However, it is acknowledged that some sources used to assign indicator values may be outdated or inaccurate [1]. Table 5 presents a detailed distributional analysis of data completeness within this dataset.

Table 5

Results of calculation of variation indicators by country according to the Statistical Capacity Indicators database

Indicator	Australia	Burundi	Haiti	Spain	Mexico	Turkmenistan	Ukraine
Number of observations	72	72	70	72	72	70	72
Mean	14.67	13.21	13.47	14.67	13.21	13.47	13.21
Range of variation	12	15	12	12	15	12	15
Mode	19	19	19	19	19	19	19
Median	19	18	18	19	18	18	18
Standard deviation	5.68	5.91	5.78	5.68	5.91	5.78	5.91
Coefficient of variation	0.39	0.45	0.43	0.39	0.45	0.43	0.45

Source: calculated by the authors based on [1].

The calculations from the dataset reveal the following key insights:

1. Range of Variation –The range for Australia, Haiti, Spain, and Turkmenistan is 12 units, while for Burundi, Mexico, and Ukraine, it is 15 units. This indicates non-uniform data distribution, with the highest and lowest values significantly distant from each other. However, compared to the Doing Business database, which covered a shorter data collection period, this result is more favorable: the minimum value is no longer 1 but instead ranges between 4 and 7, the maximum value reaches 19.

2. Mode – For all selected countries, the mode is 19 units, which is a strong result, as it indicates that the most frequently occurring value corresponds to the maximum available data coverage.

3. Median Value – The median is 19 units for Australia and Spain, while for all other countries, it is 18 units.

4. Standard Deviation and Mean Comparison – There is a noticeable dispersion in data values, indicating significant deviations from the mean across the dataset.

5. Coefficient of Variation – The values suggest that the distribution of indicator completeness remains uneven across all selected countries.

Figures 4 and 5 illustrate the cumulative distribution of indicator completeness across countries.

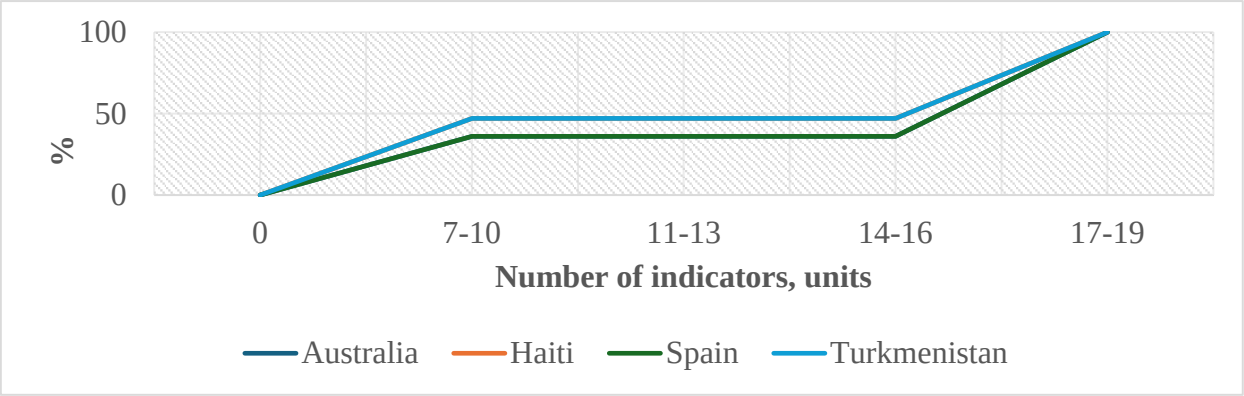

Fig. 4. Cumulative data coverage of indicators for Australia, Haiti, Spain, Turkmenistan according to the Statistical Capacity Indicators database
Source: calculated and visualized by the authors according to [1].

For Ukraine, Burundi, and Mexico, most indicators contain data ranging from 16 to 19 units, with their cumulative distribution lines overlapping on the graph. Similarly, for Spain, Haiti, and Australia, the majority of indicators have values between 17 and 19 units, and their lines also closely overlap. For Turkmenistan, a similar pattern is observed—most indicators contain 17 to 19 data points. However, its cumulative distribution line is positioned slightly above those of the previously mentioned countries.

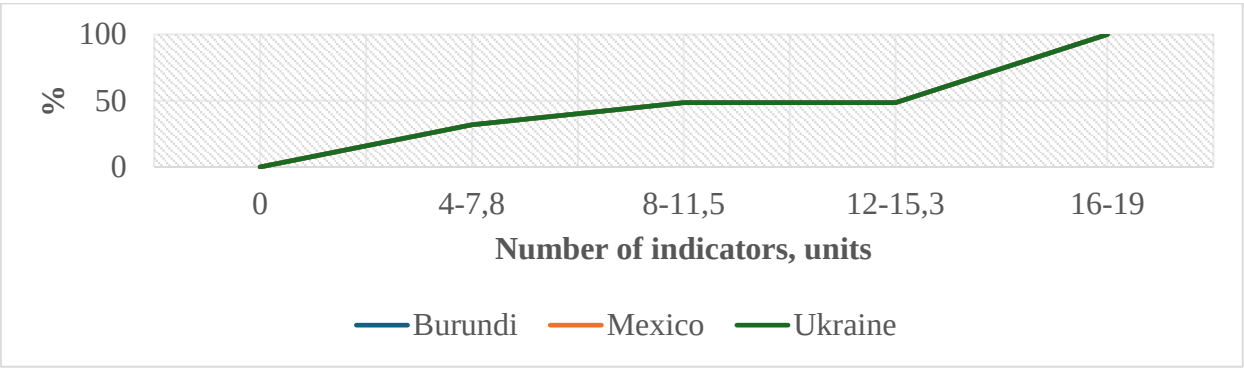

Fig. 5. Cumulative data coverage of indicators for Burundi, Mexico, and Ukraine according to the Statistical Capacity Indicators database
Source: calculated and visualized by the authors according to [1].

This result aligns with the variation coefficient calculations, but the cumulative distribution analysis provided a more precise and visually interpretable representation of the data completeness across countries.

Analysis of the Millennium Development Goals database.

The Millennium Development Goals database includes indicators related to poverty, gender equality, education, environment, climate change, social development, urban development, economic policy, and external debt [1]. This dataset covers the period from 1990 to 2015, reflecting the United Nations (UN) Millennium Declaration, which established the Millennium Development Goals – a set of eight international development objectives formulated

during the Millennium Summit and adopted as part of the UN Millennium Declaration [10].

The MDGs were based on the OECD DAC International Development Goals, which were agreed upon by development ministers in the «Shaping the 21st Century Strategy» framework. In 2016, the MDGs were replaced by the Sustainable Development Goals (SDGs), which continue the global development agenda with an expanded and updated set of objectives [11]. Table 6 presents the statistical calculations of data completeness within the MDGs database.

Table 6

Results of calculating variation indicators by country according to the Millennium Development Goals database

Indicator	Australia	Burundi	Haiti	Spain	Mexico	Turkmenistan	Ukraine
Number of observations	90	107	102	81	95	81	98
Mean	21.12	15.92	15.21	15.46	20.74	15.46	17.77
Range of variation	25	25	25	25	25	25	25
Mode	26	26	26	25	26	25	26
Median	25	19	19	22	25	22	22
Standard deviation	7.14	10.58	10.62	10.66	7.80	10.66	9.09
Coefficient of variation	0.34	0.66	0.70	0.69	0.38	0.69	0.51

Source: calculated by the authors based on [1].

The results of the statistical calculations indicate the following:

1. Range of Variation – The range is identical for all selected countries, equaling 25 units. This suggests a non-uniform data distribution, where the highest and lowest values are significantly distant from each other.

2. Mode – Australia, Burundi, Mexico, Haiti, and Ukraine have a mode of 26 units, meaning that most indicators for these countries contain the maximum possible amount of data. In contrast, Spain and Turkmenistan have a mode of 25 units, indicating high database completeness for these countries as well.

3. Median Value – The highest median values were recorded for Australia and Mexico (25 units). The median for Spain, Turkmenistan, and Ukraine is 22 units. The lowest median values were observed in Burundi and Haiti (19 units). These results indicate that more than half of the indicators for the selected countries have near-maximum data completeness.

4. Standard Deviation and Mean Comparison –There is a noticeable variation in data, meaning that values are widely dispersed around the mean.

5. Coefficient of Variation – The data distribution remains uneven across all selected countries, confirming inconsistencies in data completeness.

Figure 6 presents the cumulative distribution of indicator completeness across the selected countries. For all selected countries, the graph lines differ, yet they exhibit a similar trend. For all countries except Haiti, most indicators contain 21 to 26 data points. In contrast, Haiti has most indicators filled with data up to 20 units. Additionally, the percentage of indicators with 21 to 26 data points is higher for Spain, Mexico, and Australia compared to Ukraine, Burundi, and Turkmenistan.

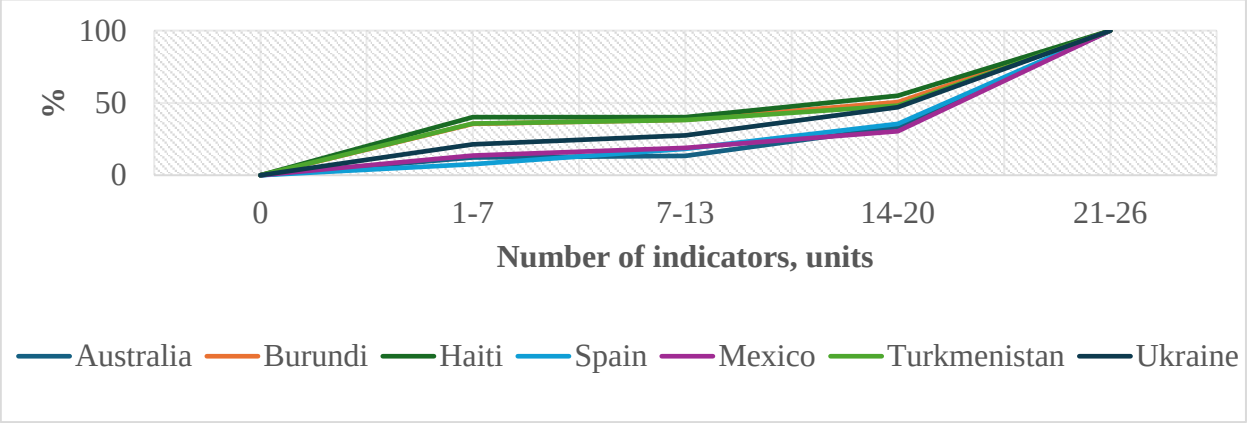

Fig. 6. Cumulative data coverage of indicators by country according to the Millennium Development Goals database
Source: calculated and visualized by the authors according to [1].

This result, when considered alongside the previous variation coefficient calculations, is significant and provides valuable insights into data completeness trends within the World Bank's databases.

Conclusions. Summarizing the results of the statistical analysis, we conclude:

1. The distribution of indicators within each selected database is uneven and inconsistent. This pattern is

particularly evident in databases containing more than 200 indicators and covering periods longer than 20 years.

2. A high number of indicators in a database does not guarantee comprehensive data availability. Some indicators lack data entirely or contain only 1–5 data points, meaning that data is available for just 1–5 years within the dataset.

3. Cumulative distribution results indicate varying levels of data completeness across databases: Gender Statistics and World Development Indicators contain data for most selected countries, with values ranging between 32 and 35 data points out of 63 possible. Doing Business has data completeness of 8–12 units out of 17, but this is justified by irregular data collection and the fact that the database was discontinued in September 2021. Additionally, confidentiality concerns and ethical considerations related to data collection may have contributed to gaps in this dataset. Mode values provide insight into the most common data completeness levels across indicators.

4. The lowest mode values are observed in the Gender Statistics database. The highest mode values appear in Millennium Development Goals, Statistical Performance Indicators, and partially in World Development Indicators.

The World Bank's online data platform offers a vast repository of datasets; however, it does not provide complete information for most indicators. The lack of statistical data for specific indicators can be attributed to several factors, including: economic development levels of different countries; structural characteristics of national economies; limited availability of gender-disaggregated data, often due to religious or socio-cultural constraints. Historical factors, such as gaining independence, government changes, and geopolitical crises, which have resulted in missing data, particularly between 1960 and 1990. Variations in national data collection methodologies, data quality issues, and challenges in data digitization and archiving. Despite these limitations, the World Bank remains an official, reliable, and publicly accessible online data platform. For scientific research, it is recommended to prioritize widely accepted and commonly used indicators. When seeking specialized or region-specific data, researchers should consult primary sources, such as official national statistical agencies. This approach to utilizing statistical resources ensures an optimal balance between data reliability and completeness, ultimately enhancing the validity of scientific research outcomes.

References:

1. DataBank (2025). The World Bank Group. Retrieved from: https://databank.worldbank.org/home. [in English].
2. Korytska, O., & Lytvyn, Yu. (2024). Statystychnyi analiz zovnishnoi torhivli Ukrainy: tsyfrovi instrumenty doslidzhennia (na prykladi Lvivskoi oblasti) [Analysis of foreign trade of Ukraine: on the example of Lviv region]. Ekonomika ta suspilstvo, No. 61. DOI: https://doi.org/10.32782/2524-0072/2024-61-13. [in Ukrainian].
3. Korytska, O., & Murynets, Yu.–I. (2024). Analizuvannia rozvytku ahrarnoho sektoru krain YeS: tsyfrovi mozhlyvosti Eurostat [Analysis of the productivity of the agricultural sector of the EU countries: digital opportunities of Eurostat]. Ekonomika ta suspilstvo, No. 62. DOI: https://doi.org/10.32782/2524-0072/2024-62-16. [in Ukrainian].
4. Hahn, J., Jordan, J., Van Pernis, P., Whelan, E., & Williamson, E. (2015). Characterization and use of the World Bank water and sanitation database. 2015 Systems and Information Engineering Design Symposium, 306–311. DOI: https://doi.org/10.1109/SIEDS.2015.7116995. [in English].
5. Galbraith, J., Choi, J., Halbach, B., Malinowska, A., & Zhang, W. (2016). A Comparison of Major World Inequality Data Sets: LIS, OECD, EU-SILC, WDI, and EHII. No. 44. Pp. 1–48. DOI: https://doi.org/10.1108/S0147-912120160000044008. [in English].
6. Ran, Q., & Zhang, J. (2012). Comparative Study on Currently Popular Network Databases. Applied Mechanics and Materials, No. 380–384. Pp. 2629–2632. DOI: https://doi.org/10.4028/www.scientific.net/AMM.380-384.2629. [in English].
7. Mishra, A. (2011). Accessing the World Bank open data programmatically. XRDS, No. 18. Pp. 44–45. DOI: https://doi.org/10.1145/2043236.2043253. [in English].
8. Lin, Y., Guan, Y., Asudeh, A., & Jagadish, H. (2020). Identifying insufficient data coverage in databases with multiple relations. Proceedings of the VLDB Endowment, No. 13 Pp. 2229–2242. Retrieved from: https://dl.acm.org/doi/10.14778/3407790.3407821. [in English].
9. World Bank Group to Discontinue Doing Business Report (2021). World Bank Group. Retrieved from: https://surl.li/jxeeuv. [in English].
10. UN Documentation: Development. Introduction, 2000–2015. United Nations. Retrieved from: https://research.un.org/en/docs/dev/2000-2015. [in English].
11. What are the Sustainable Development Goals? (2016). United Nations Development Programme. UNDP. Retrieved from: https://www.undp.org/ukraine/sustainable-development-goals. [in English].

Список використаних джерел:

1. DataBank. The World Bank Group. Retrieved from: https://databank.worldbank.org/home.
2. Корицька О., Литвин Ю. (2024). Статистичний аналіз зовнішньої торгівлі України: цифрові інструменти дослідження (на прикладі Львівської області). Економіка та суспільство, Вип. 61. DOI: https://doi.org/10.32782/2524-0072/2024-61-13.

3. Корицька О., Муринець Ю.І. (2024). Аналізування розвитку аграрного сектору країн єс: цифрові можливості Eurostat. Економіка та суспільство, Вип. 62. DOI: https://doi.org/10.32782/2524-0072/2024-62-16.

4. Hahn, J., Jordan, J., Van Pernis, P., Whelan, E., & Williamson, E. (2015). Characterization and use of the World Bank water and sanitation database. 2015 Systems and Information Engineering Design Symposium, 306–311. DOI: https://doi.org/10.1109/SIEDS.2015.7116995.

5. Galbraith, J., Choi, J., Halbach, B., Malinowska, A., & Zhang, W. (2016). A Comparison of Major World Inequality Data Sets: LIS, OECD, EU-SILC, WDI, and EHII. No. 44. Pp. 1–48. DOI: https://doi.org/10.1108/S0147-912120160000044008.

6. Ran, Q., & Zhang, J. (2012). Comparative Study on Currently Popular Network Databases. Applied Mechanics and Materials, No. 380–384. Pp. 2629–2632. DOI: https://doi.org/10.4028/www.scientific.net/AMM.380-384.2629.

7. Mishra, A. (2011). Accessing the World Bank open data programmatically. XRDS, No. 18. Pp. 44–45. DOI: https://doi.org/10.1145/2043236.2043253.

8. Lin, Y., Guan, Y., Asudeh, A., & Jagadish, H. (2020). Identifying insufficient data coverage in databases with multiple relations. Proceedings of the VLDB Endowment, No. 13 Pp. 2229–2242. Retrieved from: https://dl.acm.org/doi/10.14778/3407790.3407821.

9. World Bank Group to Discontinue Doing Business Report (2021). World Bank Group. Retrieved from: https://surl.li/jxeeuv.

10. UN Documentation: Development. Introduction, 2000–2015. United Nations. Retrieved from: https://research.un.org/en/docs/dev/2000-2015.

11. What are the Sustainable Development Goals? (2016). United Nations Development Programme. UNDP. Retrieved from: https://www.undp.org/ukraine/sustainable-development-goals.