



A comparative analysis of approaches to data integration



Speaker: ***Iryna Sierova,***

Simon Kuznets Kharkiv National University of Economics,

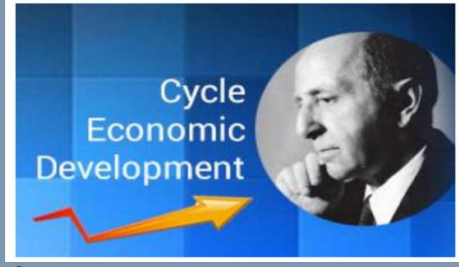
61165 Kharkiv, Ukraine



System of National Accounts 2025

- **Aim of the Research**: to develop a general framework for the analytical assessment of the role of comparative analysis in the process of data integration for the SNA 2025
- **The research object** is the processes of data integration within the system of national accounts
- **Research subject:** methodological approaches, models and data integration tools used to compile and reconcile national accounts in accordance with the SNA 2025.
- **Research hypothesis**: consideration of data integration for the compilation of national accounts in accordance with the SNA 2025, as a combination of technical compatibility between data sources and conceptual and methodological harmonization of data, metadata, and quality control procedures.

KhNUE
Simon Kuznets



- **A conceptual diagram illustrating the relationship between comparative analysis and SNA assessment tools**
- ***Comparative Analysis***: Enables the comparison of macroeconomic aggregates over time, across institutional sectors, and between countries.
- ↓
- ***A Key Tool for SNA Assessment***: identifies structural shifts, assesses the efficiency of resource utilization, verifies the quality of statistical calculations, and enhances the international comparability of national accounts indicators.



Key Dimensions of Comparison in the System of National Accounts (SNA)

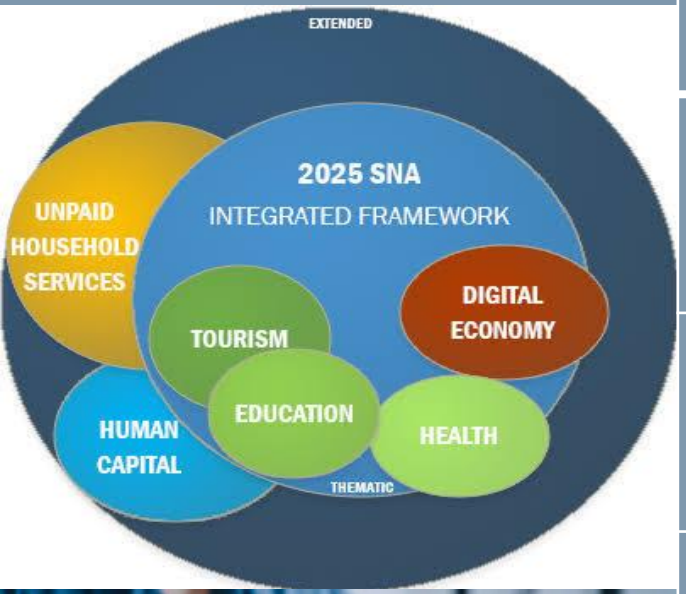
Dimension of Comparison	Assessment Focus	Examples of Indicators
Over time	Economic dynamics and cyclicality	Real GDP, growth rates, investment share
Across institutional sectors	Economic structure and interrelationships	Shares of households, general government, financial corporations, and non-financial corporations
Across countries	Relative level and model of economic development	GDP per capita, share of government expenditure, saving rate
Between theoretical frameworks and observed data	Adequacy of statistical estimates	Production, income, capital, and financial accounts balances
Across versions of the standard	Data quality and comparability	Transition from SNA 2008 to SNA 2025



The Challenges of the SNR 2025 and Their Content

Challenge	Content
New data sources	Use of supplementary surveys, administrative registers, and big data to meet new data requirements.
Methodological complexity	Need for new methods of estimation, modeling, and data quality assessment.
Disruption of comparability	Increased complexity in international comparisons of macroeconomic indicators and other aggregates.
Historical time-series revisions	Need to reconcile historical data with the new methodology.
Resource constraints	Additional financial resources and investment in staff training are required.
Unequal national capacities	A dual burden for countries that have not yet completed the transition to SNA 2008, as they must implement SNA 2008 while preparing for the transition to SNA 2025.

Methodological conditions for the validity of comparisons and key features of the transition to the SNA 2025





Terms of comparison	Key Features of the Transition to SNA 2025
Verification of the SNA version used to compile indicators for each country	The phased transition to the SNA 2025:
Calculation of prices adjusted for inflation or using purchasing power parity (PPP)	Consideration of differences in the pace of implementation of the new standards across countries
Harmonization of classifications of economic activities	Determination of country-specific priorities for implementing methodological changes
Assessment of data quality across countries	Identification of country-specific changes that are both relevant and practically feasible to implement
Retrospective recalculation of indicators following changes in the accounting standard or base year	



Distribution of errors in the interpretation of the results obtained from a comparative analysis, depending on the type of problem

Task Type	Common mistakes	Interpretation of Results
Diagnosis of the state of the research object	Violations of comparability conditions	Comparing incompatible objects, disregarding methodological differences, and failing to account for specific price formation mechanisms.
Typology and segmentation	Failure to account for the comparability of time periods	Comparing data from different periods without adjustment, interpreting short-term fluctuations as long-term trends, and neglecting seasonality and cyclical patterns.
Scientific research and theory verification	Logical and cause-and-effect errors	Treating correlation as causation, overlooking confounding factors, and misinterpreting the underlying research question.
Performance assessment	Statistical and methodological errors	Ignoring measurement errors and the statistical significance of differences, and incorrectly generalizing findings based on sample data.
Forecasting and scenario modeling	Oversimplification of intermediate results	Mechanically transferring successful practices to other countries and assessing performance using a single indicator.
Decision-making and policy justification	Data visualization errors	Selective presentation of data and insufficient explanation of the methodology used.

Key characteristics of the interpretation of the results of a comparative analysis



Statistical Significance	Practical Significance
The probability that an observed result is due to chance: assessment of the probability that an observed effect (e.g., a difference in means, a correlation, or a regression coefficient) could have occurred by chance under the null hypothesis.	The real-world relevance of a result for decision-making in a specific context: assessment of how important and useful the observed effect is in practice for informing decisions, taking into account costs, time, objectives, and other relevant factors.

Tools for analytical assessment in comparative analysis

Data Analytics		Comparative analysis	
Level	**Tools**	**Opportunities**	**Tools**
Descriptive analytics	Excel, Power BI, Tableau, Google Analytics, Looker Studio, Qlik Sense	Indicator comparison; comparison with plans and contractual targets; benchmarking; sorting and ranking	Excel, Power BI, Tableau, Google Looker Studio, Qlik Sense
Diagnostic analytics	Power BI (drill-down), Tableau, Qlik Sense, TIBCO Spotfire, KNIME	Correlation analysis; factor analysis; anomaly detection; drill-down comparison; segment comparison	Power BI (drill-through), R, Tableau, Qlik Sense, Python, KNIME
Predictive analytics	RapidMiner, Python (scikit-learn, pandas), R, SPSS, SAS Advanced Analytics, Statistica, SAP HANA	Model comparison; forecast comparison; trend comparison; model validation	Python, R, SPSS, SAS, SAP HANA
Prescriptive analytics	Board, Halo, RapidMiner, machine learning platforms, operations research tools	Scenario comparison; decision comparison; strategy comparison; optimization-based comparison	Python, Board, RapidMiner, Halo, operations research tools
Level of Data Analytics	Main Data Analytics Tools	Capabilities of Comparative Analysis	Main Comparative Analysis Tools

Risks associated with AI implementation in a comparative analysis of data integration

Comparative Analysis	Artificial Intelligence (AI)
Comparability issues	Inconsistencies in security standards
Variable equivalence	Data privacy and protection risks
Biased case selection	The model may disproportionately misclassify certain groups as threats
Excessive number of variables	Difficulty in verifying AI decisions and lack of clear explanations for the results
Lack of transparency in conclusions	Absence of transparent methods for verifying the correctness of decisions
Generation of erroneous generalizations	Inaccurate or false information
Unjustified trust in the model	AI may perform harmful actions as a result of excessive trust in its capabilities



Key issues in comparative analysis and their implications for SNA 2025

Key issues	Implications for SNA 2025
Separation of data-related expenditure	Difficulty in distinguishing between expenditure treated as investment in assets and expenditure classified as current operating costs.
Quality and substance of data	The cost-based approach does not reflect the actual economic value of the information content of data.
Free data	Difficulty in verifying the accuracy of valuation estimates and substantiating imputed costs.
Duplicate data entry	Risk of counting the same data or related expenditure more than once in GDP and capital formation.
Data reuse	Difficulty in determining whether reused data should be classified as a single asset or as multiple distinct assets.
Data depreciation	Reliance on assumptions regarding the potential rates of data depreciation.
Cross-country comparisons	Limited comparability due to differences in accounting practices, levels of digitalization, and access to supplementary data.







KhNUE
Simon Kuznets

Thank you for your attention!

