PROCEEDINGS OF THE SIXTH STATISTICS CONFERENCE

"The Power of Data for a Smart World"

SANTIAGO, JUNE 2025

banco central
Chile

Contributions in this volume were prepared for the Conference "The Power of Data for a Smart World", organized by the Central Bank of Chile in June 2025. The views expressed are those of the authors and do not necessarily reflect the views of the Central Bank, its members and the institutions represented at the meeting. This publication is available on the website (www.bcentral.cl)

PROCEEDINGS OF THE SIXTH STATISTICS CONFERENCE "THE POWER OF DATA FOR A SMART WORLD"

SANTIAGO, JUNE 2025

Seminar overview

Building Trust Through Data: A Shared Global Commitment, Gloria Peña, Statistics and Data Director, Central Bank of Chile

Opening remarks

Rossana Costa, Governor, Central Bank of Chile

Keynote speech

Power of Data for Smarter Economic Policymaking

Bert Kroese, Chief Statistician, Data Officer, and Director, IMF Statistics Department

Session 1 – Data for Crucial Decisions

The relevance of long, historical, series for policy analysis.
Vincent Bignon, Senior consultant researcher, Bank of France

Increasing the information value of data and statistics with science.
Patty Duijm, Head of Department Securities, Sustainability and Payment Statistics, De Nederlandsche Bank

Chilean Economy: Evidence-Based Decisions?
José Díaz, Professor, Pontificia Universidad Católica de Chile

Granular data: a giant leap for decision-making.
Remigio Echeverría, Deputy Director General Statistics, European Central Bank

Session 2 – New and Enhanced Indicators

New dynamics in housing loans in Portugal: the use of microdata.
Ana Reis, Adviser, Bank of Portugal

Better insights in household inequalities - The role of Distributional national accounts.
Jorrit Zwijnenburg, Head of Sectoral and National Accounts within the Statistics and Data Directorate, Organisation for Economic Co-operation and Development (OECD)

Geospatial Data and Multimodal LLM for Validating Company Data.
Susanne Walter, Data Scientist, Deutsche Bundesbank

Session 3 – How We Listen to Our Audiences

Communication Strategies for Macroeconomic Statistics.
Sanjiv Mahajan, Supporting Editor to the 2025 SNA and lead author of the Joint SNA/BPM Chapter on Communicating and disseminating macroeconomic statistics. Head of Methods and Research Engagement, Office for National Statistics

How Do We Listen to Our Audiences? We Anticipate Their Needs.
Andrew McCallum, Group Manager, International Finance, Federal Reserve Board Statistical literacy:

Strategies for Making Statistics More Accessible to Users.
Ricardo Vicuña, National Director, National Statistics Institute

Session 4 AI: A Powerful Tool

Developing and Harnessing generative AI tools for Public Institutions.
Giuseppe Bruno, Economics and Statistics Directorate, Bank of Italy

Dealing with artificial intelligence: a central banking perspective.
Bruno Tissot, Head of Statistics and Research Support, Bank for International Settlements

What are AI Agents? Why do they matter?
Fernando Pérez-Cruz, Senior Adviser, Bank for International Settlements

Panel: Information Insights: Navigating Challenges and Opportunities

Bert Kroese, Chief Statistician, Data Officer, and Director of the Statistics Department, International

Monetary Fund

Bruno Tissot, Head of Statistics and Research Support, Bank for International Settlements

Luis Ángel Maza, Director of the Statistics Department, Central Bank of Spain

Gloria Peña, Statistics and Data Director, Central Bank of Chile

THE POWER OF DATA FOR A SMART WORLD[1]/

Overview of the Sixth Statistics Conference

Gloria Peña, Head of Statistics and Data Division

1. INTRODUCTION – BUILDING TRUST THROUGH DATA: A SHARED GLOBAL COMMITMENT

We have explored how data, when used responsibly and innovatively, can empower better decision-making, foster transparency, and strengthen public trust. Institutions worldwide are embracing new technologies, particularly artificial intelligence, to enhance the quality, timeliness, and relevance of the statistics we produce.

This year holds special significance as we celebrate the centennial of our institution. It has been a true honor, on behalf of the Central Bank of Chile, to host this important gathering and to welcome such a distinguished group of professionals dedicated to building better statistics for a smarter, more inclusive, and more resilient world.

Throughout the conference, we discussed the emergence of new indicators, the integration of microdata, the use of geospatial tools, and the development of distributional national accounts, all of which enrich our understanding of economic and social dynamics. These innovations are not merely technical; they are transformative, enabling us to respond more effectively to today's complex challenges.

A recurring theme has been the importance of listening, to our users, to society, and to one another. Whether through improved communication strategies, audience engagement, or ethical considerations in data governance, we are reminded that statistics are not just numbers; they are powerful tools for dialogue and understanding.

As AI becomes increasingly embedded in our work, we must remain vigilant about privacy, misinformation, and the ethical use of data. Our final panel emphasized the need for collaboration, shared standards, and a renewed commitment to quality and integrity. International organizations play a crucial role in addressing these challenges and fostering global cooperation.

All these insights aim to ensure that we continue delivering timely, reliable, and comprehensive statistics that meet society's evolving needs.

[1]/ The presentations were prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the ECB, IMF, OECD, BIS, or the central banks and other institutions represented at the meeting.

2. DATA FOR CRUCIAL DECISIONS

The first session, chaired by Luis Felipe Céspedes, Board Member of the Central Bank of Chile, emphasized the importance of data for timely decision-making. Vincent Bignon, Senior Consultant Researcher at the Bank of France, presented the work of the Historical Monetary and Financial Statistics (HMFS) Group, created under the Irving Fisher Committee and coordinated by the BIS. This initiative seeks to develop long-run data series and methodologies to inform monetary and financial policy decisions. While granular and real-time data are essential for monitoring policy implementation, historical data play a critical role in shaping decisions before they are made. Longer time series help policymakers identify exceptional circumstances, reducing the risk of short-term bias.

The group's initial work includes a review of academic practices for constructing historical series on credit aggregates, interest rates, and property prices, culminating in a roadmap for collaboration between statisticians and historians. Historical evidence increasingly influences research and central bank debates on monetary and macroprudential policies. Seminal works, such as those by Reinhart and Rogoff on debt thresholds and Schularick and Taylor on credit booms, underscore the importance of long-run data in understanding systemic risks and crisis dynamics. However, these studies often rely on fragmented series with limited comparability. The HMFS initiative addresses this gap by promoting methodological rigor and transparency, ensuring coverage, representativeness, and consistency. Examples such as the evolution of the Bank of France's balance sheet since 1800 illustrate how historical context informs decisions on quantitative easing and tightening, offering benchmarks against past crises like World Wars I and II.

Looking ahead, the HMFS Group advocates creating a "Statistorium", a collaborative framework combining historical expertise with statistical skills to produce robust, long-run datasets. This involves leveraging archives, harmonizing institutional histories, and publishing historical series through official portals to enhance accessibility and quality assurance. By doing so, central banks provide a public good that benefits academia and policy-making, mitigating reliance on unofficial or unverified data. The roadmap includes methodological discussions, national contributions, and integration with BIS statistical resources. Ultimately, making historical data widely available strengthens evidence-based policy, supports macroprudential decision-making, and fosters a deeper understanding of financial stability over time.

The second presentation, delivered by Patty Duijm, Head of Securities, Sustainability and Payment Statistics at De Nederlandsche Bank, highlighted the growing role of data science in enhancing the informational value of granular financial data collected since the global financial crisis. The Dutch Central Bank has established a dedicated Data Science department to support projects across the organization. Some of the projects relate to sustainability-related data and promote the development of sustainable finance statistics that are produced in cooperation with the European Central Bank. These efforts focus on climate risk indicators, such as carbon intensity and physical risk exposure, by combining financial portfolios with

data on companies' emissions and geographic vulnerability to climate scenarios. Unlike traditional statistics based on reporting agents, sustainable finance statistics rely on public sources and existing datasets, requiring innovative approaches to integration and timeliness.

Three strategies underpin this work: combining, enriching, and expanding data. Combining involves linking granular internal datasets (e.g., securities holdings and credit exposures) with external sources to calculate indicators like Weighted Average Carbon Intensity (WACI), using machine learning and name-matching algorithms to resolve inconsistencies. Enriching data includes projects such as the "digital twin," mapping real estate exposures against climate scenarios in collaboration with international partners and the BIS Innovation Network. Expanding focuses on improving timeliness through nowcasting methods, predicting missing elements like company revenues and emissions using machine learning and credit ratings, enabling more up-to-date assessments of climate-related financial risks.

This experience demonstrates that applying data science techniques, such as machine learning, nowcasting, and automated data integration, can significantly improve statistical processes beyond sustainability metrics. Centralized tools for external data collection, like the "data fetcher" package, enhance efficiency and consistency across institutions. Moreover, fostering collaboration between data scientists and domain experts ensures methodological rigor and practical relevance. Looking forward, the adoption of global identifiers such as the Legal Entity Identifier (LEI) and continued innovation in data integration will be critical to overcoming current limitations. Ultimately, leveraging advanced analytics and hybrid data sources strengthens the capacity of central banks to monitor systemic risks, support evidence-based policy, and address the urgent challenges posed by climate change.

The third presentation, delivered by José Díaz, Lecturer at Pontificia Universidad Católica de Chile, offered a historical perspective on how statistical data has shaped public policy decisions in Chile, emphasizing the country's long-standing tradition of data collection since the 19th century. Early efforts, such as the first GDP estimates for the 19th century and the establishment of the Central Bank in 1925, underscore the importance of historical data for understanding economic and institutional evolution. Despite limitations in data quality, these records provide valuable insights into monetary and banking developments, including inflation trends, exchange rate depreciation, and the expansion of the banking sector prior to the Central Bank's creation. Historical evidence reveals prolonged periods of monetary instability and banking crises, highlighting the relevance of reconstructing long-run series to inform current debates on financial stability and policy design.

Beyond monetary history, the presentation examined major public health and infrastructure initiatives, notably the development of potable water and sewerage systems during the late 19th and early 20th centuries. Initially motivated by high infant mortality rates and infectious diseases, these interventions required prioritization based on demographic data, such as urban population estimates derived from census records. The analysis exposed significant regional disparities and the gradual urbanization process, which influenced investment

decisions and resource allocation. These historical cases illustrate how data availability, or its absence, can shape policy outcomes and how systematic efforts to collect and harmonize information enable more effective responses to social challenges.

Finally, the discussion addressed Chile's dependence on natural resource exports, focusing on nitrate in the late 19th and early 20th centuries and copper in the mid-20th century. Fiscal reliance on export taxes during the nitrate boom transformed the country's revenue structure, reducing the role of domestic taxation and prompting major reforms in the 1920s, including the introduction of income tax. Later, copper raised questions about foreign investment, technological change, and nationalization policies, illustrating the complex interplay between economic data, political decisions, and global market dynamics. These historical episodes underscore the need for robust, accessible data to anticipate structural shifts and design resilient policies. The overarching message is clear: investing in historical and contemporary data systems is essential for evidence-based policymaking and for addressing emerging challenges in a rapidly changing world.

The fourth presentation, delivered by Remigio Echeverría, Deputy Director General of Statistics at the European Central Bank (ECB), examines the evolution of the ECB's statistical framework and its pivotal role in supporting monetary policy and financial stability. Since its inception in 1999, the ECB primarily relied on traditional macroeconomic statistics, such as monetary aggregates (M3), integrated financial accounts, and balance of payments, to set interest rates and maintain its medium-term inflation target of 2%. During the first decade of the euro area, policy implementation was relatively straightforward, requiring limited data beyond these aggregates.

The global financial crisis of 2008 marked a turning point. The collapse of Lehman Brothers and the subsequent sovereign debt crisis exposed significant information gaps, particularly regarding credit and market risk. These deficiencies hindered the ECB's ability to answer urgent questions about exposures to failing institutions and systemic vulnerabilities. Later shocks, including the COVID-19 pandemic and disruptions linked to the war in Ukraine, further underscored the need for more timely and granular data.

In response, the ECB developed a new statistical toolkit to support unconventional policy measures such as long-term refinancing operations and large-scale asset purchases. Key initiatives include AnaCredit, a loan-level database covering credit exposures to legal entities; securities databases providing comprehensive information on issuances and holdings of financial instruments linked to entity registers for integrated analysis; and money market statistics collecting daily transaction data from unsecured, repo and derivatives markets. These datasets have transformed the ECB's capacity to respond to crises, allowing rapid assessments of exposures during episodes like Credit Suisse, assessing differentiated lending conditions for SMEs, and managing collateral scarcity during asset purchase programs. Granular market data also enabled the creation of the €STR benchmark rate following the LIBOR crisis.

Looking ahead, the ECB plans to expand granular data collections to include household loans and sustainability-related attributes. The future vision emphasizes producing aggregate statistics from granular data and integrating supervisory and statistical reporting to reduce costs for reporting institutions while enhancing analytical capabilities.

In summary, the ECB's experience demonstrates that granular, high-frequency data is indispensable not only for setting policy rates but also for ensuring effective implementation, safeguarding financial stability, and responding to systemic shocks in an increasingly complex economic environment.

3. NEW AND ENHANCED INDICATORS

The second session, chaired by Claudio Soto, Board Member of the Central Bank of Chile, focused on the importance of granular data for developing cutting-edge indicators to support central bank decision-making. The initial presentation, delivered by Ana Reis, Adviser at the Bank of Portugal, illustrated how granular data can reveal emerging financial phenomena, using the case of mortgage loan transfers. In 2022, sharp interest rate increases raised concerns because over 90% of housing loans in Portugal were on floating rates, creating immediate financial pressure on households and prompting government measures such as banning early repayment fees. These developments highlighted the need to go beyond aggregate statistics and leverage microdata for deeper analysis.

Using its central credit register, which records all loans above €50, the Bank of Portugal identified a growing trend: loan transfers, cases where households repay an existing mortgage and simultaneously take out a new loan for the same property transfers, became significant in late 2022 and surged through 2023 and 2024. An algorithm matched loans with the same borrower and similar amounts (within 5%), accounting for reporting delays and collateral information. This approach enabled accurate detection without additional reporting burdens.

Findings showed that in 2023, Portuguese banks issued €12.8 billion in new housing loans, a 13% decline from 2022. However, 27% were transfers, meaning the true reduction in new lending was closer to 30%. Transfers were concentrated among younger, highly educated households, typically holding larger, recent loans originated after 2016, when interest rates were historically low. Simulations revealed that by end-2024, households saved an average of 13% on monthly payments, totaling €26 million for loans transferred in 2023. Savings were achieved through better interest rates and, in some cases, extended maturities. This case demonstrates how granular data acts as a "magnifying glass" for understanding market behavior and improving policy analysis without increasing reporting burdens.

The second presentation, delivered by Jorrit Zwijnenburg, Head of National Accounts within the Statistics and Data Directorate at the Organisation for Economic Co-operation and Development (OECD), focused on the development of Distributional National Accounts, an

initiative to provide systematic, internationally comparable data on economic inequality. While traditional national accounts offer aggregate household data, there is growing demand for insights by different household groups, for example clustered by level of income, consumption, and wealth. Aligning distributional results with national accounts ensures comprehensive coverage, including elements often missing from microdata (e.g., social transfers in kind), and enhances comparability under the System of National Accounts (SNA).

This work responds to G20 Data Gaps Initiative recommendations and the updated 2025 SNA, which call for household distributional results. Targets include publishing estimates by income and wealth quintiles for 2021–2023 by early 2025 and moving to decile-level results by 2026. These efforts are led by the OECD in collaboration with the ECB, Eurostat, IMF, UN, World Bank, and participating countries worldwide, including Chile. Two expert groups lead this effort: one on income, consumption, and saving, and another on wealth distribution. Methodology involves five steps: adjusting national accounts totals to limit the scope to private households; linking relevant microdata variables to national accounts items; bridging gaps between micro and macro aggregates through imputations; clustering households into household groups, for example by income or wealth; and deriving aligned distributional results. Challenges include incomplete coverage in microdata, missing items such as informal income, and underrepresentation of top wealth holders. Bridging these gaps is critical for improving accuracy and enabling finer granularity (deciles or percentiles).

Preliminary results reveal significant inequality across countries: Mexico and the U.S. show the highest income disparities, with top quintiles earning up to eight times more than the lowest quintiles, while Ireland and Sweden exhibit much lower ratios. Wealth inequality exceeds income inequality, whereas consumption shows relatively smaller gaps. Saving ratios highlight vulnerabilities among lower and middle quintiles. Social-demographic breakdowns add further insights, showing younger individuals often concentrated in lower income quintiles and older groups in higher quintiles, with variations across countries.

Future steps include expanding country coverage, improving timeliness and granularity, refining methodologies, and developing joint distributions of income, consumption, and wealth for a holistic view of household material well-being.

The third presentation, delivered by Susanne Walter, Data Scientist at the Deutsche Bundesbank, explores an innovative use case applying generative AI and multimodal large language models (LLMs) to enhance the quality and completeness of company data by leveraging unstructured sources, particularly geospatial data. This initiative addresses common challenges faced by central banks when working with granular microdata: inconsistencies, missing values, and regional biases, especially for small firms and physical risk assessments. Traditional data cleaning and record linkage methods are resource-intensive and often insufficient, prompting the search for AI-driven solutions.

The project integrates novel unstructured data, such as high-resolution aerial imagery, 3D building models, and textual information from company websites, financial reports, and social media, with existing datasets. Geospatial data from Germany's Federal Agency for Cartography and Geodesy includes overflight images with 20 cm resolution and digital twin models capturing building volume, function, and structural details. These features enable plausibility checks, such as verifying whether a building's size aligns with reported employment or whether its function matches declared economic activity.

The core objective is to automate data validation and quality checks using AI. Multimodal LLMs process visual and textual inputs to confirm company attributes, detect anomalies, and classify economic activities. For example, satellite imagery combined with building function data can reveal discrepancies, such as manufacturing firms registered at residential addresses, indicating potential errors. OpenStreetMap data supplements this process by providing facility boundaries and contextual information.

Initial experiments tested LLM capabilities in two ways: (1) validating company activity types against satellite images and (2) predicting activity codes solely from images. Results were promising, models accurately identified industrial complexes and provided confidence scores and reasoning, such as noting heavy machinery or storage areas. When given detailed descriptions of activity types, classification performance improved significantly, even at granular NACE code levels. However, multimodal integration—processing multiple data types simultaneously—remains challenging, with models performing best on single-modality tasks like image analysis.

The project also explored record linkage, where LLMs demonstrated strong ability to match company names despite noise, abbreviations, or formatting differences, reducing reliance on traditional string-matching techniques. Comparative tests revealed trade-offs between smaller open-source models and larger proprietary ones: smaller models achieved higher match rates but produced more false positives, while larger models delivered fewer but more accurate matches. Incorporating reasoning improved interpretability but increased computational costs due to token-based pricing.

Key lessons include the importance of building a ground truth dataset for training and evaluation, leveraging confidence probabilities for model assessment, and tuning parameters such as "temperature" to minimize hallucinations and ensure reliability. Future steps involve scaling experiments to full datasets, benchmarking AI-based approaches against traditional methods, and refining multimodal architectures to exploit complementary information from diverse sources.

This pioneering work demonstrates the potential of AI and geospatial data to transform statistical processes in central banks, offering a pathway to more accurate, timely, and cost-efficient data validation for economic analysis and policymaking.

4. HOW WE LISTEN TO OUR AUDIENCES

The third session, chaired by Stephany Griffith-Jones, Vice Governor of the Central Bank of Chile, discussed the importance of communication and user-centric approaches in statistical dissemination. The first presentation, delivered by Sanjiv Mahajan, Supporting Editor to the 2025 SNA and lead author of the Joint SNA/BPM Chapter on Communicating and Disseminating Macroeconomic Statistics, and Head of Methods and Research Engagement at the Office for National Statistics, focused on the critical role of communication and dissemination in macroeconomic statistics. He highlighted innovations introduced in the 2025 System of National Accounts (SNA) and the seventh edition of the Balance of Payments Manual (BPM7). These updates reflect a growing recognition that effective communication is essential for improving user understanding, preventing misinterpretation, and enhancing trust in official statistics. Historically, these standards lacked guidance on communication. The new chapter fills this gap by providing principles for clear, user-focused dissemination while avoiding prescriptive technologies to ensure adaptability. The approach emphasizes tailoring language and formats to diverse audiences, from policymakers and journalists to researchers and the public, without altering technical definitions.

Key innovations include:

- **Alignment Framework:** Harmonizes taxonomy, terminology, and branding across international standards.

- **Common Glossary:** Consolidates over 12,000 terms into 850 harmonized definitions, improving consistency.

- **Compilers Hub:** A collaborative platform for sharing methodologies and best practices, enabling countries to leverage existing resources and avoid duplicative efforts.

Effective communication strategies emphasize clarity on what to share, when to release, how to deliver, and who the audience is. Recommendations include proactive engagement through release calendars, feedback channels, and training programs for media and statistical staff. Practical tools showcased include user-friendly websites, visualization dashboards, infographics, videos, podcasts, and mobile-friendly formats. Examples from Canada, Costa Rica, Estonia, and the UK illustrate diverse approaches, including interactive dashboards and multilingual content. Resource allocation and dedicated teams are critical for success. Implementing these strategies requires dedicated teams with expertise in communication, IT, and statistical production. Continuous improvement, supported by user feedback and international collaboration, ensures that dissemination practices remain relevant and effective transforming them from technical outputs into trusted resources for informed decision-making.

The second presentation, delivered by Andrew McCallum, Group Manager for International Finance at the Federal Reserve Board, emphasized anticipating user needs rather than relying solely on feedback. Using the Treasury International Capital (TIC) System as a case study, he illustrated strategies for proactive design and dissemination. TIC collects and publishes data on cross-border banking and securities positions and transactions, including derivatives. Three main actors manage this process: the U.S. Treasury, which holds legal authority and publishes the data; the New York Fed, which supervises reporters and compiles data; and the Federal Reserve Board, which reviews data, develops methodologies, and prepares reports for dissemination. TIC data serves multiple audiences, including policymakers at the Fed, the Bureau of Economic Analysis (for balance of payments and international investment position), the IMF, and the public via the Treasury's website.

The presentation highlighted two strategies for anticipating user needs: addressing potential questions and improving data consumption. For example, when new data collections were introduced in April 2023 and April 2024, the team released extensive documentation and FAQs alongside the data to clarify unfamiliar aspects and explain methodological changes. They also published a Feds Note, a short analytical paper, to provide deeper context and comparisons with previous data.

To enhance usability, TIC publications were redesigned to be machine-readable, facilitating automated ingestion by data aggregators such as Bloomberg and Haver. Tables were reformatted for mobile accessibility, recognizing that many users—especially policymakers—consume data on smartphones. Long annual reports were also improved through better visualization: replacing dense tables with charts, expanding coverage to include more debt categories, and adopting formats that display multiple countries and years simultaneously. These changes aimed to make complex information more digestible and visually intuitive.

Further refinements included Mobile Accessibility redesigning tables and graphics for smartphone use. Accessibility Compliance was also prioritized by standardizing fonts across text and graphics to comply with U.S. government accessibility standards (Section 508) and improve readability for users with visual impairments.

In conclusion, small, incremental improvements, such as better documentation and mobile-friendly formats, can significantly enhance user experience. Anticipating needs requires deliberate planning and resources, transforming data from technical outputs into actionable insights.

The third presentation, delivered by Ricardo Vicuña, National Director at the National Statistics Institute (INE), shared Chile's experience in addressing two major challenges: improving communication of official statistics and adapting to the growing role of artificial intelligence (AI) in data dissemination, issues that reflect global concerns among statistical producers striving to maintain relevance and trust in an evolving information ecosystem.

Regarding the first challenge, enhancing Communication and Statistical Literacy, INE emphasized that official statistics must be understandable and usable not only for experts but also for the public. Promoting statistical literacy fosters informed citizenship and evidence-based decision-making, in line with the UN Fundamental Principles of Official Statistics. To achieve this, INE adopted a segmentation strategy to reach diverse user groups, government, private sector, academia, journalists, and students, through tailored products and multiple dissemination channels. Barriers such as technical jargon, low statistical literacy, and public distrust were addressed through clear language, accessible narratives, data visualization, storytelling, and differentiated products.

Examples include online courses, citizen conferences, newsletters, and school visits to foster a participatory culture. INE also developed thematic microsites, such as the Gender Statistics Portal (with the Ministry of Women and Gender Equality) and the Labor Market Information System (CMAIL) in collaboration with the Ministry of Labor and technical assistance from the ILO. For specialized audiences, INE offers tools like the R Quality Package for survey accuracy and an Automatic Coding API using neural networks for classifying occupations and economic activities.

The communication strategy for the 2024 Population and Housing Census included spokespersons, media campaigns, hotlines, social media engagement, and partnerships with universities and radio stations. Special outreach targeted immigrants, indigenous communities, and vulnerable groups. Stakeholder management involved ministries, municipalities, civil society organizations, and parliamentarians, ensuring broad participation and trust in the census process.

The second challenge concerns adapting to AI-Driven Data Access. Users increasingly rely on AI tools rather than visiting official websites, reducing direct interaction and feedback for statistical offices. While this trend poses risks, such as reliance on inaccurate secondary sources, it also offers opportunities for collaboration. INE stressed the need to adopt standards and tools that enable official data to interact with AI systems, citing the IMF's StatGPT initiative as an example. This system uses SDMX standards to convert natural language queries into structured requests, ensuring accurate retrieval of official data.

INE plans to develop prototypes allowing citizens to query official databases through language models, ensuring reliable responses to questions like unemployment rates. Collaboration with major AI platforms could also provide feedback on data usage, helping agencies evaluate impact and relevance.

In conclusion, INE underscored that adapting to AI and enhancing communication strategies are essential to maintain the relevance of official statistics. Failure to innovate risks diminishing the role of statistical offices in an increasingly digital and automated world.

5. AI: A POWERFUL TOOL

The fourth session, chaired by Luis Óscar Herrera, General Manager at the Central Bank of Chile, explored the transformative potential of artificial intelligence in central banking, focusing on generative AI (GenAI), large language models (LLMs), and AI agents. The first presentation by Giuseppe Bruno, Head of Division at the Economics and Statistics Directorate at the Bank of Italy, examined opportunities and challenges in leveraging GenAI within public institutions. The discussion focused on recent advancements in LLMs, practical applications developed by the Bank of Italy, and the importance of interdisciplinary collaboration in implementing these technologies responsibly. The presentation highlighted rapid advancements in LLMs since 2023, including GPT-4 and open-source alternatives like LLaMA, and tools enabling web-based knowledge augmentation. These developments have improved model accuracy and expanded capabilities in natural language processing and reasoning. However, risks such as toxic content and hallucinations underscore the need for robust governance and ethical frameworks.

A key theme mentioned was the dual nature of LLM intelligence, language and mathematics, and the progress in mathematical reasoning alongside linguistic tasks. Scaling laws show that larger models trained on extensive datasets deliver better performance, though resource constraints and diminishing marginal returns impose limits. Training requires significant computational power, while inference tasks, such as answering questions, consume less but still notable resources.

The Bank of Italy's experience illustrates how GenAI can enhance efficiency and innovation in central banking. Some practical applications are:

- **Document interaction and summarization:** Using tools like Google's NotebookLM, the Bank enabled chat-based interaction with internal documents, allowing users to request summaries, identify inconsistencies between versions, and extract key details.

- **Legal and regulatory analysis:** High-capacity models (up to 405 billion parameters) deployed on supercomputing platforms to query complex legal frameworks, such as Italian banking laws, and produce structured outputs.

- **Programming language translation:** AI-driven solutions were tested to convert code from proprietary languages (e.g., Stata) to open-source alternatives like R, supporting the transition to more accessible technologies.

- **Tone adjustment in communications:** Generative models were applied to refining clarity and style in public speeches and reports, improving engagement with diverse audiences.

A hybrid infrastructure combining Microsoft Copilot, cloud environments, and supercomputing resources ensures scalability, security, and flexibility for AI experimentation. Despite these advances, human oversight remains essential to mitigate errors and avoid hallucinations.

Future priorities include prompt engineering, domain-specific training, and secure on-premises deployment of open-source models to prevent data leakage.

In conclusion, generative AI offers transformative potential for central banks, from automating routine tasks to enabling sophisticated analytics. However, success depends on interdisciplinary collaboration, ethical safeguards, and continuous evaluation to balance innovation with accountability.

The second presentation by Bruno Tissot, Head of Statistics and Research Support at the Bank for International Settlements, was based on the results of a recent survey on the use of AI in central banks. As outlined in IFC Report no 18, innovative technologies hold significant potential, offering efficiency gains, generating new analytical insights, and supporting policy decisions. A strategic priority for central banks is to explore generative AI to enhance a wide range of operational tasks, including information retrieval, computer programming, and data analytics. However, many institutions remain in the initial adoption phase, highlighting a central question: how can AI be used effectively and responsibly in production processes?

A first insight is that the deployment of AI must duly take into consideration concerns about privacy protection, cyber security, skills shortages and ethical biases.

A second critical area involves IT considerations, such as computational capacity, the choice between closed versus open-source AI models, the decision to adopt in-house solutions versus off-the-shelf products, and the use of cloud services.

Finally, further progress is needed on more "traditional" data management issues, given that AI-generated outputs intrinsically depend on the quality of their underlying data inputs – the well-known "garbage in, garbage out" principle. This underscores the need to improve all phases of the data life cycle, from production, validation, integration and storage to dissemination and use. Priority areas include: (i) ensuring data and metadata quality to support transparency, traceability, and machine readability; (ii) enhancing data access, appropriate sharing, and the exchange of best practices; (iii) developing modern, metadata-driven and standardized data processes and systems; and (iv) strengthening user literacy in AI and data-related issues.

The last presentation, delivered by Fernando Pérez-Cruz, Senior Adviser at the Bank for International Settlements (BIS), examined the implications of AI and AI agents for labor markets, data confidentiality, and governance frameworks, as well as the role of statistical institutions in monitoring these rapid changes. The discussion highlighted diverse perspectives from central banks, financial regulators, and international organizations, addressing both opportunities and risks associated with AI adoption. He offered a historical perspective on the evolution of artificial intelligence, clarified misconceptions between AI and machine learning, and discussed how AI agents could transform routine tasks and compliance processes. The evolution of large language models (LLMs), which progressed from predictive text systems in 2018 to versatile tools capable of answering complex questions, culminating in the release of ChatGPT in 2022. The concept of Artificial General Intelligence (AGI) was introduced,

referring to systems able to perform most cognitive tasks that humans can do. However, current models remain limited in scope.

Two definitions of AI agents were explored:

1. **LLMs using external tools**—limited by data preparation requirements.

2. **LLMs controlling computers like humans**—a promising approach enabling automation of routine tasks without pre-structured data. Combined with zero-shot learning, this capability allows models to learn workflows by observing human actions and replicating them efficiently.

The presentation clarified the distinction between machine learning and statistics, noting that both aim to predict outcomes but differ in priorities: statisticians focus on causal, interpretable models, while machine learning prioritizes minimizing prediction error, even using over-parameterized models and embedding techniques. These approaches underpin innovations such as self-supervised learning, which trains models on large unlabeled datasets for general tasks before fine-tuning for specific applications. Practical implications include automating compliance processes such as anti-money laundering checks, reducing manual work without costly system overhauls. This approach offers a pragmatic path to efficiency without major infrastructure changes. However, challenges remain, memory limitations and error rates require human oversight. The analogy with self-driving cars illustrated that achieving near-perfect reliability for continuous decision-making is difficult; widespread deployment will likely be limited to specialized, high-value tasks rather than universal automation. Looking ahead, modular AI architectures combining fast, intuitive components (System 1) with analytical ones (System 2) may emulate human cognition. Success will depend on targeted training, robust governance, and careful integration into workflows.

Finally, the session considered the role of statistical offices and central banks in monitoring AI-driven changes. Suggestions included developing indicators to monitor labor market dynamics, technology adoption, and productivity shifts. Such metrics would support evidence-based policymaking and help institutions anticipate structural shifts.

In conclusion, AI is not merely a tool for efficiency but a catalyst for structural change, raising critical questions about employment, data governance, and institutional readiness. Harnessing its benefits will require a balanced approach, combining technological innovation with robust safeguards and proactive policy frameworks.

6. INFORMATION INSIGHTS: NAVIGATING CHALLENGES AND OPPORTUNITIES

The seminar concluded with a panel discussion on the challenges and opportunities arising from new information. The session was chaired by Alberto Naudon, Board Member of the Central Bank of Chile and Chair of the BIS Irving Fisher Committee (IFC). The panel featured Bert Kroese, Chief Statistician, Data Officer, and Director of the Statistics Department at the International Monetary Fund; Bruno Tissot, Head of Statistics and Research Support at the Bank for International Settlements; Luis Ángel Maza, Director of the Statistics Department at the Bank of Spain; and Gloria Peña, Director of Statistics and Data at the Central Bank of Chile.

The discussion began with the impact of AI and digital transformation. Panelists highlighted the growing use of machine learning and generative AI in statistical work, including applications such as satellite imagery for economic estimates and tools like *StatGPT* to enhance data accessibility. They stressed the importance of leveraging existing data sources, such as central bank speeches and commercial datasets, and noted that AI applications are expanding beyond statistical imputation to areas like legal document search. The conversation underscored the need to build internal capacity and infrastructure to support broader AI adoption, extending its use beyond research departments to administrative functions, while addressing legal and ethical considerations.

The panel then turned to the evolution of trust and communication. Panelists agreed that building public trust in statistics requires transparency, methodological rigor, and proactive engagement. They emphasized the importance of legal frameworks that safeguard statistical independence and highlighted the role of metadata, openness, and effective communication strategies, including social media and direct dialogue with users. The discussion also addressed the growing challenge of misinformation and the need for statistical institutions to help the public interpret data accurately.

Next, the panel explored collaboration and governance issues. They discussed persistent barriers such as legal constraints, institutional silos, and lack of interoperability, noting that the main obstacle is often insufficient institutional commitment. Panelists cited successful European efforts to reconcile trade data and emphasized the importance of leadership and international coordination. They also highlighted the role of international organizations in promoting standards, metadata sharing, and global data infrastructure, and called for stronger cross-border collaboration and the development of universal identifiers to unlock the full potential of microdata.

The session concluded with reflections on future priorities for statistical producers, emphasizing the need to remain agile, collaborative, and user-focused in an increasingly complex data environment.

SIXTH STATISTICS CONFERENCE "THE POWER OF DATA FOR A SMART WORLD",

SANTIAGO, JUNE 2025

OPENING REMARKS BY ROSSANA COSTA, GOVERNOR, CENTRAL BANK OF CHILE

Good morning, ladies and gentlemen. It is an honor and a privilege to welcome you to the Sixth Statistics Conference of the Central Bank of Chile, which we are hosting in a particularly meaningful year for our institution as we celebrate its centennial. Over the past 100 years, the evolution of the economy, and of statistics, has been closely interlocked with dynamic and ever-changing environments and faced with a future that is advancing rapidly and full of promising and challenging changes.

This sixth edition of the Statistics Conference, whose title, "The Power of Data for a Smart World" resonates strongly in this moment of our history. If there is one lesson the past century has taught us, it is that knowledge, grounded in solid information and rigorous analysis, guides the path to progress and stability. We are also opening the window to new developments that offer us opportunities and challenges, because they provide tools, but at the same time demand more information, of better quality and greater timeliness.

In recognition of these one hundred years, I would like to revisit them through the lens of statistics and from there, open the door to the next century.

The Bank and Statistics

The production and publication of statistics have gone hand in hand with the institutional development and the economic and financial evolution of our country, as well as with global events.

At least three decades before its official establishment in 1925, there was widespread debate about the idea of having a central bank that would contribute to macroeconomic stability. By the late nineteenth century, economic instability, and a fragile monetary system, evidenced by frequent banking crises and the near-continuous depreciation of the Chilean peso, highlighted the need for an institution capable of providing stability. These significant economic and social tensions culminated in the enactment of a new Constitution in 1925. Within this context, the idea of an autonomous central bank, including its governance and its relationship with the Executive and Legislative branches, became an important topic of national debate.

From its inception, our Central Bank recognized the importance of information for decision-making. It established a specialized unit to collect statistical data that would enable analysis and study of the country's commercial and economic conditions, essential for the Board at the time.

In January 1928, the Central Bank of Chile began publishing its Monthly Bulletin. This publication, that compiles national statistics, continues to this day. Its first issue included figures on nitrate production, exchange rates, railway traffic, total amounts of checks processed in clearinghouses, components of money (currency and deposits), and loans. These data reflected the Chilean economy of the time, characterized by a strong dependence on nitrate as the main export product and the driving force of domestic activity.

During the Great Depression of the 1930s, Chile suffered a severe economic contraction due to the sharp decline in nitrate and copper exports, which weakened international reserves, public finances, and economic growth. The fiscal surplus of the late 1920s turned into a deficit in 1931, GDP plummeted, and a deflationary period unfolded, culminating in the abandonment of the gold standard and the suspension of external debt service. In this context, banking statistics (that is, loans, interest rates, currency in circulation, and bank reserves); foreign exchange statistics (official exchange rate and foreign currency operations); fiscal data (on government revenues and expenditures), foreign trade data (exports and imports of copper, nitrate, and coal) and international reserves gained importance.

During World War II, the Chilean economy faced external restrictions, rising inflation, and increased state intervention. The war disrupted international trade, negatively affecting exports, except for copper, which remained strong due to demand for arms production. One of the main statistical innovations at the time was the systematic publication of the balance of payments. Although its compilation began in 1930 by the General Directorate of Statistics, it was in 1943 that the Central Bank of Chile took an active role in its preparation, analysis, and dissemination.

Toward the end of World War II, in 1944, the Bretton Woods Monetary and Financial Conference was held in the United States to design a new international financial system. This meeting led to the creation of institutions such as the International Monetary Fund (IMF) and the International Bank for Reconstruction and Development (now part of the World Bank Group). The IMF was tasked with promoting international monetary cooperation, supporting trade expansion and economic growth, and discouraging policies that could harm global macroeconomic stability and nations' welfare.

The Bretton Woods agreements required the Central Bank to submit a "Foreign Exchange Budget" each November, which included estimates of the foreign currency the country could use the following year to cover imports, capital account liabilities, and other obligations or needs. This was first presented in 1946.

The need for a common methodology to measure countries' economic activity led the United Nations to publish the first international standard for the System of National Accounts (SNA) in 1953. Chile adopted this system in 1967 with the creation of the National Accounts Department under the National Planning Office (ODEPLAN). This enabled the annual publication of statistics on Gross National Product (GNP), investment, consumption, and productive and institutional sectors. This new information marked a significant advancement in the availability of national macroeconomic data.

The decades following World War II were marked by Central Bank financing of the Treasury, which led to recurrent inflationary problems, and balance of payments crises. The economic authorities of the time designed diverse and consecutive macroeconomic stabilization programs through price controls, including wages, and applied different exchange rate schemes, without correcting the source of the imbalance. In this context, it became necessary to have balance of payments and monetary statistics.

In the 1970s, the Bretton Woods agreements were abandoned globally, which brought new challenges for the oversight of the international financial system. The collection of statistics now took place in a world of flexible exchange rates and increasing financial integration. Locally, in the early years of the decade, the Chilean economy was largely closed, facing significant macro- and microeconomic imbalances, marked by substantial fiscal expansion that led to hyperinflation.

In the second half of the 1970s, Chile began a process of trade liberalization, gradually eliminating tariff barriers, reducing subsidies to domestic production, and dismantling exchange and price controls. In this new context, the Central Bank strengthened its role in statistical compilation, expanding foreign trade coverage through the publication of export and import statistics by product and country. This allowed for more precise analysis of trade flows and better monitoring of trade openness and the competitiveness of local industry. Monetary and financial statistics were also enhanced to reflect the incipient activities and operations emerging in the capital market, distinguishing between monetary and fixed-income assets and, in the latter, identifying the various inflation-indexed instruments and their issuers.

The 1980s marked a turning point. In July 1981, the Central Bank officially assumed responsibility for compiling and publishing the National Accounts, which had previously been the responsibility of ODEPLAN. The Bank thus established itself—alongside the National Statistics Institute —as one of the main institutions responsible for producing and disseminating Chile's macroeconomic statistics.

The 1982–83 debt crisis had a profound impact on economic activity and employment, severely affecting the population's well-being. Its severity led to the suspension of external debt payments, the collapse of nearly the entire domestic financial system, the imposition of capital controls, and the adoption of an adjustment program with the IMF to support economic stabilization. This program included fiscal and structural adjustment measures, such as reducing the fiscal deficit, controlling inflation, imposing capital controls, and restructuring the financial system.

These developments demanded greater statistical transparency, particularly regarding external accounts. Thus, in 1983 the Central Bank began publishing external debt data disaggregated by maturity and counterparty. That same year, to enable more timely monitoring of the economy, the Bank also launched the Monthly Index of Economic Activity (IMACEC), a key tool for short-term economic analysis that continues to this day.

In 1989, a milestone occurred that has shaped the history of our institution over the past 35 years: the enactment of the Constitutional Organic Law, which established the Central Bank's autonomous status. In this context, the Bank's role was consolidated as the institution responsible for compiling and publishing the country's main macroeconomic statistics in a timely manner, including statistics of monetary and exchange rates, balance of payments, and national accounts.

In line with international best practices, in 1992 the 1986 Input-Output Matrix (IOM86) was published. It was considered one of the most comprehensive in Latin America by its breakdowns. It was updated in 1996 and 2003, and since 2008 has been calculated annually, along with the publication of Supply and Use Tables. Additionally, the Bank updates the benchmark years every five years since 1996. The next benchmark revision, for the year 2023, will be published in March 2027.

The globalization and increasing financial integration of Chile into the world economy during the 1990s prompted the development of new statistics. Measuring external debt alone was no longer enough. It became necessary to also include the international investment position, capturing investments made abroad by various domestic economic agents, with institutional investors playing an increasingly prominent role.

Institutional strengthening and technological advances enabled improvements in the frequency, coverage, and disaggregation of data, as well as the dissemination of new statistical series since the second half of the two thousand. In fact, to consolidate all macroeconomic statistics compilation and collection functions within the Central Bank, the Statistics Division was created in 2008.

Likewise, we have gradually integrated into the international community, actively participating in organizations such as the IMF and the BIS. Chile's admission to the Organization for Economic Co-operation and Development (OECD) in 2010 further enhanced the quality and scope of macroeconomic statistics. This has involved adopting the highest international standards in data collection and dissemination, ensuring both the comparability and reliability of our statistics. At the regional level, we are also affiliated with the Center for Latin American Monetary Studies (CEMLA).

Technological progress has paved the way for the statistics produced by the Central Bank of Chile to evolve continuously to meet the growing demand for information to support monetary and financial policy, as well as the needs of the Chilean society. In 2005, Quarterly National Accounts were introduced, followed in 2011 by the Quarterly Institutional Sector Accounts (QISA), integrating real and financial dimensions of the economy with greater frequency.

The 2008 Global Financial Crisis highlighted the need to improve transparency and oversight of derivatives markets. In Chile, this led to the development of the Integrated Derivatives Transactions Information System (SIID-TR), a centralized data repository that enables financial authorities to monitor systemic risks, identify cross-institution exposures, and prevent liquidity crises or contagion.

In critical contexts such as the COVID-19 pandemic, the urgency of having more timely and detailed data became evident, enabling a deeper understanding of economic and social phenomena, as well as the heterogeneous impacts across different segments of society. In response, the Bank, using microdata, increased both the frequency and timeliness of its publications, along with the territorial and thematic disaggregation of data. These improvements have facilitated more agile and precise decision-making, especially in periods of heightened uncertainty. One example of this was the launch, in mid-2022, of the quarterly Regional GDP publication and the development of experimental statistics, which provide more timely insights into the evolution of economic activity.

Currently, we continue working to enhance the use of statistics and data. We are committed to producing statistics on household disposable income, consumption, and savings by income quintile, in line with international recommendations. Additionally, we are developing natural capital statistics, and within regional national accounts, we plan to include data on investment and foreign trade.

In our commitment to serving society and making the Bank's information more accessible, we have continuously innovated in how we publish and share our statistics in different formats and channels. In 2004, we launched the Statistical Database (BDE). Later, we introduced a mobile simplified version and an API, to enable bulk data downloads and seamless integration with analytical and visualization tools. We also launched a Regional App, improving access to this data. As part of this effort, in 2024 we also announced the first public call for research projects using microdata, promoting academic collaboration.

Statistics of the Future

As we have seen, the Bank has played a key role in providing information to society, guided by the principles of transparency, trust, and integrity. Our statistical work has evolved through the use of data and technology, the implementation of new methodologies, and the exploration of new data sources.

As we cross the threshold of our centennial, it is time to envision the statistics of the future. As I mentioned earlier, economic and social events have historically driven the development of new statistics, as well as advances in computing led to more comprehensive, higher quality, and more timely statistics. A similar pattern can be expected going forward. However, the future also brings new challenges, given the availability of microdata, advances in artificial intelligence and quantum computing, and improvements in regulatory frameworks. Certainly, economic statistics are poised for a radical transformation.

Asking ourselves: what statistics and data might look like 100 years from now? is more than an interesting experiment—it is a compelling exercise. Some current aspects will only improve. For example, the immediacy with which we access information today suggests that, in the future, every economic transaction could be recorded in real time, allowing us to instantly gauge economic activity. This would bring economists closer to their long-standing aspiration

of detecting turning points in key variables as they happen. Likewise, data collection sources will multiply, and historical series could be continuously updated, eliminating the need to wait years for final versions.

The use of unstructured data—captured from our digital devices such as smartphones or smartwatches—could offer more accurate, timely, and cost-effective insights than traditional surveys. This could, among other things, facilitate the measurement of informality.

Statistical portals, which currently allow users to collect information based on their needs, could evolve into platforms where users, by inputting their characteristics, would receive customized information powered by AI. Machines may communicate autonomously, potentially leading to previously unthinkable fields—such as the development of statistics for robots.

The future is uncertain but promising. However, these advances must be accompanied by regulations that ensure the appropriate use of data and protect privacy, placing the society's well-being at the center, promoting transparency, and ensuring the ethical use of information to strengthen public trust.

In this rapidly evolving world—one that may reach unimaginable heights—our 2023–2027 Strategic Plan includes a key objective, that is, to continue deepening the use and analysis of data across all areas of the institution. We aim to be at the frontier when major changes occur, with a robust statistical system that respects individual and social rights. This will help the Central Bank to continue to fulfill its role as a guardian of macroeconomic stability in the decades to come.

Conference 2025

Our commitment to staying informed about innovations in statistics and data has been enduring. This led us to organize our first Statistics Conference in 2015, bringing together the interests of both compilers and users, aligned with the international agenda and our strong commitment to promoting, disseminating, and using statistics. Today, with pride —and after organizing five conferences, including one during the pandemic in 2021, with participation from experts at statistical offices and central banks at the forefront of knowledge—, I can say that the Statistics Conference has become a regional benchmark event.

This year's conference is an opportunity to reflect on the past, celebrate current achievements, and look to the future with determination. Over the next two days, we will explore successful experiences in the use of data and artificial intelligence for more informed and timely decision-making. We will see how data is being used to gain deeper insights into economic and social phenomena, anticipate trends, identify opportunities, and mitigate risks. I believe the discussion on how artificial intelligence is transforming decision-making will be of particular interest.

In a world where information flows at unprecedented speed, organizations must know how to listen to their audiences and adapt to their needs. Moreover, in a landscape filled with multiple sources and indicators competing with traditional statistics, ethics, privacy, and responsibility—both in data management and information dissemination—are essential considerations. For this reason, we will explore how various institutions work to maintain credibility and strengthen public trust through effective communication.

Final Reflections

Since their earliest uses, statistics have been—and will continue to be—essential for interpreting our reality, guiding research, and, above all, enabling informed decision-making. In an increasingly interconnected and data-saturated world, their relevance not only endures but grows stronger.

The role of central banks and statistical offices in providing the information needed to understand the economy and design evidence-based policies is more vital than ever. On the occasion of its centennial, the Central Bank of Chile reaffirms its commitment to producing relevant economic information, leveraging the growing analytical capabilities offered by granular data and artificial intelligence tools.

A century of history has taught us that stability is built on solid foundations of knowledge and transparency. That crises and abrupt changes in the environment are best addressed with timely, high-quality data and rigorous analysis—not improvisation or political pressure. That autonomy is not an end in itself, but a means to ensure technical decisions focused on the well-being of citizens.

We look to the future with the certainty that the power of data—managed with responsibility, ethics, and vision—will enable us to build a more prosperous and resilient world. Yet we face a number of complex and multifaceted challenges, such as designing strategies to manage and analyze large volumes of data while ensuring their reliability, accuracy, and confidentiality, and balancing data utility with privacy rights.

As producers of statistics, we must maintain a constant commitment to learning and a willingness to think outside the box—to be creative, explore new methodologies, and strengthen communication and collaboration skills in an increasingly interdisciplinary world, all while meeting the growing demand for more accurate, rigorous, timely, and granular information.

As for the limits we can reach, there are concepts today that remain only partially measurable—such as well-being—which may become fully measurable in the future. This would represent a revolution in the way public policies are designed. Whether or not we reach that point, we must begin preparing now.

The Sixth Statistical Conference of the Central Bank of Chile represents a valuable opportunity to learn, share, and collaborate in addressing the new challenges we face as producers and users of statistics.

We sincerely thank all speakers and participants present for your valuable contributions to this event and for allowing us to share and learn from one another. We are confident that the discussions and presentations over the coming days will be of great value and inspiration in continuing to strengthen the fundamental role of statistics in our countries. I hope this conference proves productive and enriching for each one of you.

Finally, I would like to thank the organizing team of this Conference—Patricia Medrano, Helen Parker, Antonia Silva, Vania Vargas and Romina Villarroel from the Statistics and Data Division, and Dafne Guiloff and Maria José Reyes from the Institutional Affairs Division.

A very warm welcome to you all!

SIXTH STATISTICS CONFERENCE "THE POWER OF DATA FOR A SMART WORLD",

SANTIAGO, JUNE 2025

POWER OF DATA FOR SMARTER ECONOMIC POLICYMAKING

BERT KROESE, CHIEF STATISTICIAN, DATA OFFICER, AND DIRECTOR AT THE IMF STATISTICS DEPARTMENT

Thank you, Governor, for your great speech and for emphasizing the importance of data and statistics. It is a real honor to speak on a topic that is central to my work and passion. I also congratulate you on the Centennial of the Central Bank of Chile—a highly respected institution whose statistics Division enjoys international recognition, thanks to your strong commitment to data and statistical excellence.

It is a pleasure to be here in person. During the fourth Statistics Conference in 2021, I participated virtually from my home in the Netherlands due to travel restrictions. While that was a valuable experience, nothing compares to meeting and exchanging ideas face-to-face over the next two days.

Central banks depend on timely and accurate data to fulfill their core functions—from monetary policy and financial stability to supervision and payment systems oversight. High-quality data is not a luxury; it is essential for effective and trustworthy policymaking.

Today, I will provide an overview of the role of statistics and data in central banks, focusing on emerging areas in the 21st century—particularly artificial intelligence and big data.

The Role of Central Banks in Data

Central banks are both producers and users of data. Mandatory reporting by financial institutions ensures reliable information for compiling key statistics such as monetary aggregates, balance of payments, and financial soundness indicators. In some countries, national accounts are compiled by central banks, in others, by national statistical offices. These traditional datasets remain essential, complemented by indicators on inflation, unemployment, fiscal accounts, and sentiment.

Historical Perspective

Early central banks, such as Sweden's Riksbank (17th century) and the Bank of England, operated without formal statistics, relying on merchant reports and gold flows. The Industrial Revolution introduced official statistics, followed by milestones such as the creation of the Federal Reserve (1913) and the IMF (1944). Mid-20th century advances in computing

enabled forecasting and inflation targeting, while IMF dissemination standards (SDDS and SDDS Plus) promoted transparency and quality.

The Data Revolution

The 21st century has transformed statisticians into data scientists, but the core mission remains, turning data into actionable insights. Big data, IoT, satellite imagery, and digital platforms now complement, not replace, traditional statistics, offering timeliness, granularity, and new perspectives. COVID-19 accelerated innovation, highlighting the need for rapid, alternative data sources. Microdata, such as credit card transactions and loan-level datasets, has become critical for understanding household and firm behavior. While some areas, such as CPI measurement, can leverage scanner data or web scraping, most applications enhance rather than replace conventional statistics. This evolution is reshaping how central banks and policymakers make decisions in an increasingly complex and uncertain world.

Added Value and Innovation

Big data enhances:

- **Timeliness:** GDP nowcasting uses alternative datasets (nighttime lights, Google Trends, emissions) to provide early signals of economic trends.
- **Granularity:** Job postings and wage data from online platforms offer detailed insights by occupation and region.
- **Advanced Analytics:** Text mining and semantic clustering gauge sentiment and anticipate risks.

Examples of Innovation

Recent advances illustrate how big data and technology are transforming economic measurement.

- **PortWatch (IMF):** Tracks global trade using AIS signals from ships, enabling real time monitoring and scenario analysis for shocks and climate risks. It even supports countries like Somalia in producing official trade statistics.
- **GDP Nowcasting:** Combines satellite imagery (e.g., nighttime lights), Google Trends, emissions data, and other indicators for near real-time estimates of economic activity. Nighttime lights provide insights into infrastructure and urbanization, while search trends and emissions data help track tourism, trade, and informal activity. These tools are increasingly used by the IMF and national authorities to improve timeliness and accuracy in economic monitoring.

- **Geospatial Risk Analysis:** Maps flood risks against population and GDP exposure to assess financial stability implications. The IMF Geospatial Data Dissemination Platform combines hazard probabilities with economic exposure, guiding adaptation strategies and systemic risk assessments. This initiative is part of the G20 Data Gaps Initiative.

- **Text Mining & Sentiment Analysis:** Uses semantic clustering of news and reports to monitor uncertainty and financial stress in real time. Indicators such as the World Uncertainty Index and Financial Stress Index provide early signals of economic trends. There's a lot of use of that information because it gives this early indicator of where the economy is heading.

- **Labor Market Analytics:** Leverages web-based job postings and LinkedIn data to track emerging skills (e.g., AI expertise), labor market tightness, and green job trends. These datasets provide timely insights into structural changes that traditional surveys cannot capture due to cost and time constraints. However, coverage limitations exist, such as incomplete representation of the population, requiring careful interpretation to ensure representativeness.

- **Crypto Monitoring:** Monitors cross-border flows and foreign exchange contracts to identify risks such as capital flight, currency substitution, and money laundering. Data collection relies on crypto exchanges, wallet providers, and blockchain analytics, but challenges remain due to the absence of unified regulatory frameworks and a single issuer.

Current efforts under the G20 Data Gaps Initiative include defining standardized templates and exploring solutions such as blockchain analysis and commercial data providers. These initiatives aim to improve transparency and risk assessment in an evolving financial landscape.

AI and Data Readiness

Making data AI-ready requires open access, clear metadata, APIs, and interoperability standards like SDMX. The IMF's new data portal consolidates datasets with consistent metadata and API capabilities, enabling machine-readability and integration with AI tools. These initiatives are part of a broader strategy to modernize economic data governance, including updated methodological manuals (SNA, BPM, MFS, GFS) and a new quality assessment framework that underpins dissemination standards and audits.

AI-driven tools under development include:

- **StatGPT:** Converts natural language queries into SDMX-based requests to trusted databases, reducing hallucination risks and ensuring reliable outputs. We hope to announce it by the end of 2025.

- **Talk to the Manuals:** Assists compilers and users in navigating complex statistical standards by answering questions and linking to relevant sections of the manual where it gives the answer, and you can read it yourself.

Together, these efforts aim to make official data findable, interoperable, and policy-ready, shaping the future of global data governance.

Partnerships and Governance

Innovation depends on collaboration among international organizations, national authorities, and the private sector. Initiatives such as the G20 Data Gaps, Project and partnerships with technology companies enable access to private data while ensuring privacy and ethical standards. The IMF participates in a global framework involving 30 private companies (e.g., Google, Microsoft, LinkedIn) and 12 international organizations to facilitate secure data sharing. This framework addresses legal, governance, and technical aspects, ensuring compliance with copyright and privacy requirements. Secure techniques like remote code execution and homomorphic encryption enable analysis without compromising confidentiality. Tools such as the Data Access and Sharing Maturity Assessment, based on Eurostat's work, help countries benchmark and improve their data governance strategies.

Concluding Remarks

Data is at the core of central banking and policymaking. The IMF is committed to enabling innovation through tools, partnerships, and global standards. Together, we can build a modern, interoperable, policy-ready data ecosystem that meets the challenges of a rapidly evolving world.

Thank you.

All the presentations are available at the web page: Statistics Conference - Banco Central de Chile (bcentral.cl)

Proceedings of the Sixth Statistics Conference
"The Power of Data for a Smart World"