PROCEEDINGS OF THE FIFTH STATISTICS CONFERENCE

"Information to unlock the future"

SANTIAGO, SEPTEMBER 2023



banco central
Chile

Contributions in this volume were prepared for the Conference "Information to unlock the future", organized by the Central Bank of Chile in September 2023. 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)

This publication is available on the Conferencia Site at the Central Bank website (https://www.bcentral.cl/en/content/-/detalle/quinta-conferencia-estadistica).

This publication is protected under Chilean Law 17336 on intellectual property. Hence, its contents may not be copied or distributed by any means without the express permission of the Central Bank of Chile. However, fragments may be reproduced, provided that a mention is made of source, title, and author.

PROCEEDINGS OF THE FIFTH STATISTICS CONFERENCE "INFORMATION TO UNLOCK THE FUTURE"

SANTIAGO, SEPTEMBER 2023

Seminar overview

The value of information in central banks, Gloria Peña, Statistics and Data Director, Central Bank of Chile

Opening remarks

Rossana Costa, Governor, Central Bank of Chile

Keynote speech

Granular, timely and beyond GDP: thoughts on the new statistical landscape.

Paul Schreyer, Chief Statistician and Director of Statistics and Data Directorate, OECD

Session 1 – Measuring the Digital Economy

Measuring the U.S. digital economy: current progress and ongoing research.
Tina Highfill, Senior Research Economist, BEA

Incorporating the economic value of data into the system of national accounts.
John Mitchell, former national accountant at the OECD, international expert

The digital economy and productivity.
David M. Byrne, Principal Economist Industrial Output Section, Federal Reserve Board

Digital economy measurement: Challenges and opportunities.
Martha Tovar, Senior Economist, IMF

Session 2: Artificial Intelligence: Applications at Central Banks

The AI revolution: Present and near future.
Álvaro Soto, Head of National Center for Artificial Intelligence (CENIA)

Trustworthy AI.
Maciej Brzezinski, Head of the European Central Bank Data Office, ECB

Artificial intelligence and credit risk: Accuracy vs. interpretability.
José Manuel Carbó, Senior Economist, Financial Innovation Division, Banco de España

Session 3 Panel discussion: Natural Capital

Measuring the environmental goods and services sector in the U.S.
Scott Wentland, Senior Research Economist, Office of the Chief Economist, BEA

How the environmental accounts can be used in the decision-making process.
Mónica Rodríguez Z., Executive, Environmental Statistics Area, Banco Central de Costa Rica

Economic and environmental accounts in Chile: insights and challenges.
Enrique Calfucura, Senior Economist, Macroeconomic Statistics, Banco Central de Chile

Keynote speech

Information to unlock policies.
Mário Centeno, Governor, Banco de Portugal.

Session 4: Data Governance and Communication

Data governance: It's different but it matters.
Sandra Cannon, Data Management Expert, Cannon Analytics, Robert Gordon University

Statflix: a Banco de Portugal's original series.
Luís Teles, Deputy Director, Statistics Department, Banco de Portugal

New channels, new approaches, new data formats.
Fernando Rocha, Head of Statistics Department, Banco Central do Brasil

Opportunities and challenges of social media.
Aldo Mascareño, Senior Researcher, Centro de Estudios Públicos (CEP)

Panel: Challenges and opportunities of new information

Paul Schreyer, Chief Statistician and Director of Statistics and Data Directorate (OECD)

Bruno Tissot, Head of Statistics and Research Support, BIS and Head of the Secretariat, Irving Fisher Committee (IFC)

Maciej Brzezinski, Head of the European Central Bank Data Office (ECB)

Ricardo Vicuña, National Director, National Statistics Institute of Chile (INE)

INFORMATION TO UNLOCK THE FUTURE

Overview of the Fifth Statistics Conference[1]/

Gloria Peña, Statistics and Data Director

1. INTRODUCTION – INFORMATION ISSUES FOR CENTRAL BANKS

We are living in times of profound digital and technological transformation. "Information" is a key aspect of this process. This transformation has accelerated the provision and use of massive information, and has opened the door for more timely, granular, and varied data. We have the chance to change the way we produce statistics, making them more efficient and flexible. The rapid advances in technology use and the availability massive information and new sources now, allow us to address and expand our scope to topics like distributional national accounts, natural and human capital, and create alternatives to existing official statistics, such as experimental statistics. This new environment also poses new challenges in terms of data privacy, data governance and addressing ethical dimensions.

During this conference, we had the opportunity to learn from each other's experiences, exchange knowledge and expertise, and broaden the scope of our analyses regarding digital economy, artificial intelligence, communication, data governance, and the opportunities and challenges associated with the future of information.

We also reviewed ways to enhance the accounting of natural capital, aiming to contribute to a more comprehensive analysis on sustainability issues that address emerging user needs in this area and its economic impact. Furthermore, we explored the significance of data, especially microdata, within our organizations, highlighting its critical role in analysis and decision-making to gain a better understanding of economic developments, design public policies, and, in turn, assess their impact on the economy.

We want to be part of these changes. We need to stay relevant and enjoy the advantages of these opportunities. In addition to changing and expanding the scope of our statistics, we also need to improve communications and reach new audiences. We emphasized the importance of effective communication with diverse audiences, promoting active listening practices when engaging with data users.

At the same time, we have learned these days that there are many aspects and challenges to consider. Artificial intelligence (AI) provides many benefits, but we also need to protect the society from deepfakes, fake news, and, in general, the misuse of these new technologies.

[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 OECD, BEA, IMF, BIS, or the central banks and other institutions represented at the meeting.

We still have many questions to answer. How can we protect personal information? Should personal information be used to train these kinds of models? We will need to adopt common standards and procedures to continue improving data management and dissemination, as well as to promote innovation to keep pace with these advancements while mitigating their associated risks. At the same time, we will have to maintain quality, trust and safeguard ethics in our work. The role of international organizations is key in addressing these challenges and fostering extensive collaboration. All these insights are aimed at ensuring that we can continue to provide timely, reliable, and comprehensive statistics and better serve society.

2. DIGITAL ECONOMY

The first session, chaired by Stephany Griffith-Jones, Board Member of the Central Bank of Chile, discussed the importance of measuring the digital economy. The initial presentation, by Tina Highfill, Senior Research Economist at the BEA, highlighted the digital economy statistics and the research areas related to the digital economy. The BEA's digital economy report covers 4 categories of goods and services based on the OECD's framework: Infrastructure (ICT hardware and Software), E-commerce (Business-to-business and Business-to-consumer e-commerce), Priced digital services (Cloud services, Telecommunications services, Internet and data services, and all other priced digital services), and Federal non-defense digital services. They use supply-use tables to for most of these categories. However, they also use external data from the US census bureau to isolate the E-commerce portion. The results show that for the year 2021, the Digital economy represented 3.7 trillion dollars, which represents almost a 9% of US gross output. In terms of value added it is over 10% of GDP. The report also shows that inflation had a much bigger impact on the total US economy, than it did for the digital economy for both, gross output and value added.

Even though, the report has a comprehensive view for digital economy, they intent to continue to improve and expand on these statistics. Some areas on emerging topics in the digital economy include measuring "data as an asset", new price indexes for software and cloud services, valuing "free" digital services and developing digital trade statistics. Another one relates to **digital intermediation services**, such as rideshare, travel services, where a digital platform only connects producer and consumers, but the digital intermediator doesn't own the product. Therefore, it is not possible to obtain this information from supply-use tables. The Highfill paper uses two approaches, **bottom up** where they look at publicly available financial report for the ride-share companies; and the second is a **top-down** approach where they start with total gross output for taxi services and try to narrow down to the part that is specific to the digital intermediation. Even though the yearly data obtained for the 2017-2021 period move mostly in the same direction, the total growth rate does not coincide. The same results apply for the travel services companies. In this case the author shows a clear preference for the top-down approach because the bottom-up approach is probably underestimating, given it is hard to identify all firms involved.

There are some concern national accounts are not valuing **"free" media,** like free TV, newspapers of magazines, Google, Facebook, etc. The Rachel Soloveichik paper is an experimental approach to measuring "free" digital media as a **payment to consumers**. They describe advertising as a three-way transaction where users barter their viewership to the media company, the media company resell the viewership to outsiders and then that viewership is used by outsiders. Using this experimental method to construct measures, they find that it would actually increase growth rates in GDP.

The last topic is related to the **new questions** (cloud computing services, digital intermediation services, and digital orders) added to the **survey of digital activity within multinational enterprises** that allowed to understand better what companies think are producing digital intermediation services and using and producing cloud services.

The second presentation, by John Mitchell, national accountant expert (former OECD), was about incorporating the **economic value of data** into de System of National Accounts (SNA) and GDP more broadly. There is an IMF-Eurostat task force, compiling a manual that will assist countries with incorporating data assets into GDP in line with the revision to SNA that is occurring.

Data is becoming a critical input into new economic products, and it is crucial to recognize **data as an asset**. Data has become fundament to the creation of new products such as digital advertising and artificial intelligence, but it also has become fundamental to business processes to become more efficient. Given the importance of data, many people in statistical offices and academia have tried to compile estimates to come up with the value of data. They can distinguish at least three ways: based in **market capitalization** o venture capital valuation, is bigger than the sum of the fixed assets, there is an additional asset worth value in there; **revenue from sales of data products** or data intensive products; or **willing to pile or accept surveys** where people would put a value on the pros of giving the information to digital platforms.

Conceptually, data would appear to align with other produced assets in GDP. For the purpose of the SNA data will be defined as "Information content that is **produced** by accessing and observing phenomena; and recording, organizing and storing information elements from these phenomena in a **digital format**, which provide an **economic benefit** when used in productive activities." Data is very different from other assets in the SNA. It is produced on an own account basis, it doesn't depreciate in a conventional sense, it is extremely heterogenous, and quantity is often unrelated to value. A third of firms are increasingly using big data analytics, produced by themselves. There is not a lot of use in purchasing data your competitor across the street can also buy. Given market transactions are unlikely, and there is limited information to formulate assumptions to estimate net present value. The recommendation that is going forward is to measure data assets as sum of cost, like with other intangible assets in the SNA. This method has been successfully implemented by NSO's and academia. It has confirmed the feasibility of the approach they are recommending. The

value of the data asset, range con 0.8% to 2.7%. Even though data value has been growing faster than general GDP, given it small percentage there is no substantial impact on GDP so far. Data is in a similar position as "Research & Development" when it was incorporated in the SNA for the first time in 2008.

However, a conceptual framework is only half the journey. Existing work validates the possibility of data's inclusion in the SNA. This work on Data measurement is part of a broader OECD-wide "Going Digital" project. Research, testing, refinement, clarification will continue before and after data's inclusion in the 2025 SNA.

The third presentation, by David Byrne from the Federal Reserve Board, talked about the digital economy from a productivity perspective and how it enters the potential output of the economy. The first challenge is to define what digital economy is. The 2025 SNA revision defines digitalization but leaves the boundaries of digital economy uncertain. Therefore, the digital economy satellite account is defined as the production of infrastructure (IT hardware and software, e-commerce and priced digital services). It is a growing share of the US economy. As a growth accountant, we include the total factor productivity for industries producing IT capital (including components) and the productivity contribution of the use of IT capital throughout the economy.

The digital economy plays a central role in the study of sources of economic growth, innovation and technical change, and globalization. **Sources of economic growth**, the production and use of IT capital boosted aggregated productivity growth in the late 1990s and early 2000s, but that contribution has dropped off in recent years. This is relevant because to anticipate the capacity of the economy, we have to forecast productivity. We need to understand better the process of **innovation and technical change**, to deal with emerging technologies like AI. How long is it going to take between the innovation happening and its impact on productivity. There are a lot of things that happen in between the invention and the kick in productivity (Paul David example on dynamo invention). The digital economy is a preeminent example **of globalization**. The electronic supply chain has spread all over the world. The production of electronics moves across many countries before there are finished. Consequently, local ecological, epidemiological and political disruption is magnified through the supply chain.

Accurate and useful measures of the digital economy need to have three different types of consistencies: Capital types, time-series, and geographic. Consistency across **different types of capital**, tangibles vs intangibles, different kind of tangibles and intangibles: symmetric treatment for the high-volume production like cell phones and low volume production like MRI machines. We must account for the digital contribution to the quality of those goods symmetrically across all these things to get analytical results you can believe. We need **time-series consistency** over a very long period. The Paul David example shows that it takes over a decade for the effect to play out. Finally, we need **geographic consistency** for supply chain questions.

How do we integrate indexes that differ in data sources, index methods, scope and timespan? There is no enforcement mechanism to enforce consistency. There is a focus on narrow projects because there is not enough time to wait for perfectionism. Finally, he mentioned the recency bias, which is the tendency to believe current innovations are more remarkable than historical ones.

The fourth presentation, by Martha Tovar from the Statistical Department at the IMF, reviewed the challenges and opportunities in the digital economy measurement. There is a clear impact of the digital revolution on our daily life. Some examples are the transformation of the financial services industries, the value generated with the data of a football team or cloud computing. Therefore, a huge challenge is how to include these new digital activities in the GDP and make them visible.

The pervasive nature of digitalization makes it difficult to isolate and quantify the economic impact of this activity. The standard classifications do not define a "digital sector" or set of "digital establishments" whose growth and evolution can be compared to other sectors such as manufacturing and agriculture. The pace of innovation and the application of digital technologies to existing economic activity makes it a challenge to ensure comprehensive coverage of economic activities that are digitally impacted. While the task of quantifying digitalization is difficult, there is value in quantifying this activity to better understand its evolution relative to overall economic activity and its impact on labor markets, inflation, societal progress, etc.

GDP can be considered an adequate concept to measure market production, but some concerns have arisen over several areas. Are we properly capturing the prices and volumes behind the digital products? What are the new forms of intermediation services? What is happening about the free and subsidized consumer products? What is happening with the consumers as producers of data? Certain assets (data) or even cross border flows among countries, are not being measured.

There are three main concepts about digitalization. Digitization (from an economic perspective) is the process of converting products into a digital format to be used in production or consumption (e.g. the ideas of an author are digitized using a computer). Digitalization is the use of digital technologies to facilitate economic activity (e.g. the digitized book is sold via a digital platform and digitally delivered to the consumer who consumes the book using an e-reader). So, the "Digital Economic Activity", all of this together, reflects digitally facilitated economic activity as a whole.

The handbook to measure the digital supply and use tables, is based on three elements:

- Digital Industries that are those firms that rely on and/or utilize digital technologies to produce goods and services.

- Digital Transactions, which are facilitated by digital technologies.

- Products that are either digital themselves (e.g. e-books) or that facilitate the production of digital products and digital exchanges (e.g. communication switches, cell phone towers)

The statistics department of the IMF conducted an experimental exercise in 2022 with five countries and used three countries that already disseminating digital economy estimates to compute indicators of digital industries. The update of the standards includes guidance on quantifying the impact of digitalization on the economy. The recommendation for the countries is to develop a suite of indicators such as: Digital transactions, products, and industries, and also indicators related to e-commerce, digitally delivered products; digital intermediation platforms, expenditures on information, communication and technology, artificial intelligence, cloud computing and data.

The international statistics community find the conceptual framework robust, but the measurement in some areas may require improvement and new approaches. Recognize that complementary statistics to bridge gaps between consumer surplus and GDP may be necessary. In response to user demand for data, making the digital revolution more visible, developing guidance that countries can use to quantify the impact of digitalization on their economies. The guidance is designed to meet a wide range of demands and needs. The updated standards are also addressing user needs.

The Statistics Department has provided technical assistance and training to IMF Members. It implemented a multipronged Capacity Development (CD) strategy:

- Organize regional workshops: Latin America and the Caribbean, Africa, Asia and others to come.

- Provide technical assistance and training on digital economy or digitalization topics.

- Develop methodology handbooks and compilation guides: Crypto currencies, digital money, Fintech, data.

- Coordinate CD with other international organizations.

- Provide Teaching Assistance (TA) on the use of a new range of non-traditional data sources that can potentially be used to improve estimates of GDP and for the development of high frequency economic indicators.

They are using a new digital platform for collaboration; transiting to electronic publications, share common concepts across all statistical domains; hosting a forum for interaction between compilers and users and peer-to-peer collaboration; and the development of a compilers' hub to facilitate sharing of best practices. The key ingredients for a successful transition to updates standards, considers being engaged in the global consultations, developing revision process, engage with reporters and users of data, ensure sufficient resource allocation and request IMF TA if needed.

3. ARTIFICIAL INTELLIGENCE

The second session, chaired by Alberto Naudon board member of the Central Bank of Chile, reviewed artificial intelligence applications at Central Banks. The initial presentation, by Álvaro Soto, Director of the National Center for Artificial Intelligence, CENIA, presented his view about the current state of the art in AI. In his view there are two main technologies behind the revolution in artificial intelligence. The first revolution occurred in 2012 and allowed the implementation of associative memory, transforming any pattern in a memory address. It is a big advantage because you do not need to search in the whole memory, you can take advantage of deep learning. It is a powerful pattern recognition ability through effective associative memories. The second revolution occurred around 2020 with the **reasoning-based machines**. If you input a definition or an image, the machine can produce a coherent text. It is a powerful reasoning ability (but still unstable) through large scale models. It is a new way to access knowledge. In the 1980's we had to attend a class or go to the library to access knowledge. In the 1990's we could search in the web. Now we have a new scenario to transform information into knowledge. It has an impact in the near future in a new way to access knowledge and on a new way to build machines with these new technologies. This new technology also has an impact on the type of machines we can build. For example, robots for retail stores. We are facing a huge technological revolution, with new cognitive tool. In a partnership with humans, it can become key to boost productivity and business value.

The second presentation, by Maciej Brzezinski, Head of Data Office at ECB, illustrated us about **trustworthy AI**. According to IBM Global AI Adoption Index 2022, only one out of five companies is not using AI. While organizations have embraced AI, few have made tangible investments in ways to ensure trust or address unintended bias. Some organizations have already defined the way how to adopt AI, they put in place governance and developed policies that are formally approved and operationalized. While other organizations have not started their AI journey yet and they have not started defining their governance or guidelines for AI use. Most organizations are somewhere in the middle, they are considering if they should use AI, they evaluate risks and benefits, experiment, discuss governance structures, prepare guidelines and policies. It is not about rushing with AI adoption because everyone else is doing so. It requires careful considerations where to use it, should it be used, and how to ensure it is used in a trustworthy way.

The **principles of trustworthy** AI, include being:

- Lawful, ensuring compliance with all applicable laws and regulations.
- Ethical, demonstrating respect for, and ensure adherence to ethical principles and values.
- Robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm.

Trustworthy AI concerns not only the trustworthiness of the AI system itself but also comprises the trustworthiness of all processes and actors that are part of the AI system's life cycle.

Some of the **challenges** organizations are likely to face when adopting AI, include to build a robust data management and data governance and the operationalizing of AI services. The main challenges of AI Governance are: balancing regulations vs innovation, choose between principles based vs detailed guidance, and enhanced regulations to tackle bias and privacy issues. A challenge for central banks is to star using this technology, even though it is still not regulated by governments. AI is inheriting global technology; therefore, we need to make sure it is governed globally. The main recommendations for the AI governance for central banks include having a wide range of stakeholders, be transparent about plans for AI, create mechanisms for stakeholders to provide input, be accountable for decisions, literacy and learn from others.

The third presentation, by José Manuel Carbó, Senior Economist of Banco de España, illustrated us about the effect of artificial intelligence applied to a use case in credit risk, that shows the tradeoff between accuracy and interpretability.

The financial industry is in a transition from traditional approaches such as linear regressions and dictionary methods to advanced algorithms like deep neural networks and Large Language Models (LLMs). Some possible examples at central banks on natural language processing are analysis of news for real time economic sentiment tracking, and automated extraction and understanding of terms in climate reports, loan guarantees. Also, the use of image recognition to design/production of banknotes or look for signature verification. A third topics would be identifying outliers or irregularities in both transactions and databases for preventive action, and for prediction of corporate bankruptcies or credit default.

Credit risk is central to financial stability. Accurate credit risk assessment is critical not only for individual financial institutions, but also for the overarching stability of the financial system. By focusing on AI and credit scoring, we can illustrate the opportunities and risks of AI in finance.

It clearly has a superior **predictive accuracy** in assessing credit risk, when using a tree base model (like random forest and XGBoost) or deep learning. Given that credit default is influenced by a multitude of factors that interact in a complex way, the gains can go up to 20% in different statistical metrics, with huge heterogeneity. The economic impact of a better prediction of default could decrease loss from default up to 10%. Better prediction of default could yield regulatory capital savings up to 20%. However, there are many **risks for validating** a machine learning model. Some reports on the risks of applying ML in credit (Bafin 2022, Dupont et al. 2020). Alonso y Carbó 2022a, group the risks into 3 categories: statistics, technology, and market conduct. Many of these risks apply to traditional econometrics. Nonetheless, due to its novelty and its importance, the presentation focused on interpretability problems in credit risk assessment.

There is an ongoing discussion about consumers' right to receive explanations for credit decisions, whether it's a denial or unfavorable terms. Even though there is no legal mandate to "open the black boxes" of AI in credit decision-making, the decisions shouldn't be based on special categories of personal data that we are not able to understand and explain.

Traditional econometric models like logit are inherently explainable, whereas interpreting machine learning models is complex. There is an emerging field aimed at enhancing transparency and trust in ML models. Explanations could be local (for a particular loan) or global (for the whole dataset). Some post hoc interpretability techniques can be applied to any model after it has been trained. The basic idea is to evaluate the model's performance with and without a feature to measure that feature's importance. Some techniques are:

- Data Shuffling: For each feature, its values are randomly shuffled (permuted) across all data points, destroying any correlation with the target variable.

- Performance Drop: The model's performance (e.g., accuracy, F1 score, etc.) is measured before and after the permutation.

- Importance Metric: The drop in performance indicates how important the feature is; bigger drops signify greater importance.

- Repeat and Average: The process is often repeated multiple times and averaged to get a more robust measure.

SHAP is a technique that measures the contribution of a variable to the predicted outcome, on a given day compared to the average prediction. These contributions are called the Shapley values. Suppose we have as variables "Income", "Age", and "Homeowner", and that we want to know the importance of the "Income" in the probability of default of individual i. The Shapley value of "Income" for the probability of default of individual i is the weighted average of those marginal contributions. To obtain the global importance of "Income" in the test sample, we repeat the process for all days and take the mean of the Shapley values.

Although there are techniques that interpret machine learning, many questions arise: How to define the discrepancy between interpretability techniques? (Krishna et al 2022) How worrisome is the discrepancy? Is it enough as an explanation? (Miller, 2019). The literature is still incipient and trying to improve the techniques. In Alonso-Robisco & Carbó (2022c) they propose the generation of synthetic data to create a stress test for post hoc interpretability techniques. In conclusion, machine learning brings great benefits in credit risk models, but it also has unintended effects in terms of biases and interpretability.

4. NATURAL CAPITAL

The third session, chaired by Sofía Baducco, Manager of Economic Studies at the Chilean Central Bank was a panel discussion on Natural Capital. The first presentation, by Scott Wentland, Senior Research Economist at the BEA, reviewed the methodology related to the measurement of the environmental goods and services in the US. The Central Framework of the System of Environmental-Economic Accounting (SEEA) provides a structure for organizing data about environmental activity to quantify the green economy. The first SEEA-based pilot of environmental activity in the U.S. economy developed a proof-of-concept on environmental activity accounting following a satellite account approach, drawn on detailed internal supply-use data, obtained primarily from Census's Industry and Product data along with other supplemental sources. They break down the categories that are consistent with how other countries have done it: CEPA (Classification of Environmental Protection Activities) and CReMA (Classification for Resource Management Activities).

They estimate that the U.S. Environmental Goods & Services Sector in 2015 was about US$ 620 billion dollars and in 2019, US$ 724 billion dollars (about 1.9% of the total gross output of the US economy), which is broadly in line with the EU numbers. A lot of what is considered the green economy is Waste management (26%), Management of water (17%), Wastewater management (13%) and Protection of biodiversity and landscapes (13%). Another take away from the pilot study is that around 28% comes from de public sector, while some subcategories are mostly in the private sector (Waste management) and others in the public sector (Protection of biodiversity and landscapes).

In August 2023, the White House has released guidance on accounting for ecosystem services and benefit-cost analysis that are different from the SEEA EA framework. National accounts are often focus on the price times quantity to reflect the exchange values of an economy, but benefit-cost analysis is a broader analysis which account for welfare and consumer surplus. And so, the guidance has some key differences with the national accounts, where we don't count for consumer surplus.

The second presentation, by Mónica Rodríguez, senior economist at the Central Bank of Costa Rica, was about how the **environmental account** can be used in the **decision-making process**. Traditionally, we relate the wealth of a country with its physical, human, and financial capital, but now we know the relevance of the natural capital that represent about 10% of the global total wealth with great heterogeneity across countries. Costa Rica started the development on environmental account with the support of the World Bank in a project called WAVES. It had a steering committee with the financial system, the Central Bank of Costa Rica, the Ministry of Planning, the Institute of Statistics and the Ministry of Environment and Energy. The Central Bank established the Environmental Statistics Unit that follows the environmental accounts guidelines aligned with the United Nations' System of National Account. They started with three accounts: Energy, Forest and Water in 2011. Last year, they added Material flows (physical) and Environmental protection expenditure accounts. Now, they are working on two new products. Ecosystem services and environmentally related tax revenues. The inventory of indicators is huge, so to illustrate you can get the forest contribution

to the economy, the physical water losses, the increase in forest cover and the household emission related to the energy use. The main contributions from the environmental account are disaggregated information by economic activity which allows the development of input-output analysis; access to hybrid indicators so they can mix with the same principles physical and monetary value and environmental and economic information; allow the monitoring system of environmental data in the country; and to respond to international request (OECD, UNSD-UNEP and SDGs).

To finish the presentation, the presenter reviewed some **environmental account application** that can be found on the Environmental account's website. The first one, is to respond to specific sustainable development goals (water, energy and life on land). The second application is the Integrated Environmental Economic Modelling for Costa Rica (IEEM-CR). This is a general equilibrium model, that allows forward-looking analysis of public policies and understanding of the impact of decisions prior to their implementation. They also use the account in the National Decarbonization Plan 2018-2050, in line with the Paris agreement. It allows to evaluate a green tax reform and, in the strategy, related to transparency, metrics and open data.

This year Costa Rica published the National Circular Economy Strategy, to maximize the use of the material while it is minimizing the impact in the environment. The conceptual framework for measuring the transition to a circular economy uses specific indicators from the environmental accounts (Material flows, Water and Energy). To finalize, they developed a satellite account for the National Strategy of Bioeconomy.

The third presentation, by Enrique Calfucura, Senior Economist at the Central Bank of Chile, described the insights and challenges on economic and environmental accounts in Chile. The presentation highlighted what we plan to measure in the Chilean case, given the more relevant sectors, the limitation in terms of information and the indicators that can be estimated.

In the Chilean framework, the Central Bank of Chile (CBC) is responsible for the compilation of National Accounts according with guidelines from the UN SNA. The CBC is part of the OECD informal Expert Group on Natural Capital and member of the Network of Central Banks and Supervisors for Greening the Financial System (NGFS) where it is involved in the development of climate and nature-related financial risk frameworks and data. The CBC also participates as technical advisor to the Natural Capital Committee (Ministry of Environment, Ministry of Finance and Ministry of Economy). The Ministry of Environment compiles many environmental indicators that are relevant for natural capital accounting.

The objective is to have a hybrid implementation of SEEA CF and SEEA EA where methodologies and data are available (land use, marine resources, minerals, energy, water, emissions; ecosystem extent accounts and condition accounts, provisioning ecosystem services). Regulating ecosystem services are very important but intangible. The CBC is working on dealing with the methodological and informational issues for this kind of ecosystem services. There is a collaborative work: the NCC is leading a pilot of natural capital intended

to support public policies (GEF-World Bank project in partnership with the Natural Capital Project/Stanford University).

The main product in the short term is to compile: Land use and land cover mapping, Land use and land cover change and transition mapping, and Land use physical and monetary accounts.

The development of a semi-automated algorithm that allows the replication of land cover and use for different dates will allow to develop other ecosystem accounts, such as:

1. Land cover mapping as input for Ecosystem Extent Accounting
2. Land cover mapping as input for Ecosystem Condition Accounting
3. Land cover mapping as input for Ecosystem Services (Flow and Asset) Accounting

Other products that can be compiled in the short to medium term, based on methodological and data availability, are accounts for marine resources, minerals, energy, water, emissions, and provisioning ecosystem services.

There are different challenges related to SEEA CF Accounts: the adoption of new classifications for measuring mineral resources and reserves identify the contribution of each of the power sources - hydro, thermo, solar and wind – for the supply – use table for renewable energies account. On the other hand, the SEEA EA Accounts also face challenges, such as modeling physical ecosystem services for a territory where climates and ecosystems are diverse; and valuing regulating services, which is still a work in progress. Moreover, there are additional challenges that apply to both SEEA Frameworks: the valuation of natural assets and provision ecosystem services requires estimates of physical capital at the resource level, and an appropriate private rate of return to the physical capital; dealing with the institutional framework for the valuation of natural and ecosystem services; and modelling and imputing data for small and medium agents where information is less abundant and precise.

5. DATA GOVERNANCE AND COMMUNICATION

The fourth session, chaired by Luis Felipe Céspedes board member of the Central Bank of Chile, reviewed central banks' Data Governance and Communication. The first presentation by, Sandra Cannon, Data Management Expert at Cannon Analytics and Robert Gordon University, highlighted the practices and processes that help to ensure the formal management of data assets within an organization, and in particular in central banks. The core of data governance activity is the notion of decision right and the scope of proper use of data.

We need to keep in mind what data governance is not. In the private organization, data management usually is implemented when they must fix data leakage. It is about compliance and risk management. It is not about technology. Data cannot work without IT, but data

needs to be govern separated from IT. Data governance does not have one size fits all. **Different kinds of data need different kinds of rules and support**.

What does it look like? Nearly every expert will have a version of a pyramid with three levels: Strategic conversations, tactical activities and operation thinks that need to be done. What does it cover? What is the scope? Nearly every expert will have a version of the data governance "house". Data governance is the roof that keeps everything together. Underling is technology, processes, and support services. The ethos that are in the atmosphere are education and training a change management.

The dichotomy between corporate and central banks work, starts with the data that are in the scope for data governance program. For most organizations, this is a moot point: the data that commercial entities collect, create, and manage are all about running the business. These "Enterprise Data" are widely discussed at conferences, written about by consultants, and are generally important for every organization. However, for many government agencies and mission-driven organizations, there are other data that are core to their mission but not necessarily their business. And these data are often overlooked by traditional data governance practitioners.

The main difference is that "Enterprise data", "corporate administration data", "operational data" are all generated by an internal business process, and the control over the data is entirely within the organization. In central banks data are collected from external stakeholders, acquired from external partners, or purchased from external providers, and may come with rules that the organization cannot readily change.

Another difference is the different people. The data owner has accountability for various aspects of data; often charged with making decisions on permission and access. For example, the head of HR is usually the data owner for employee data. But who "owns" the data that are brought in to serve a mission?

The theory of data management recommends the strategic decision should consider individuals that consider different **domains**. The domains are often based on key business processes: Finance, Marketing, Sales, etc. Central Banks don't have that, the business processes and topics for research and statistics organizations look very different.

There are different c**hallenges**. Data quality is often one of the reasons organizations start doing data governance. With internal data, quality can be addressed at the source: if an employee attribute in the payroll system is missing, someone in that function can investigate and correct it. However, with external data, quality issues often can't be fixed.

Finally, we have the different **motivation**. Enterprise data governance is about increasing the bottom line. The "360 View of Customer" has been the standard motivational presentation for ages. However, mission-driven organizations do want to be good stewards of their resources and do have binding budget constraints. But data governance for the "other" data has nothing to do with growing the customer base or finding new revenue sources.

Therefore, it is necessary to adapt the framework. Governance is still important. The house analogy still works, but the pillars may need to be revised. Data evaluation and **quality assessment**, policies, and guidelines, and sometimes standards. We need to align the incentives. Data governance is often seen as punitive: someone will now tell staff what they CAN'T do. Instead, the conversation can be about **enablement**: what can be done with data governance that couldn't be done otherwise. It is necessary to make the address the concerns because any change is hard. It may help to find community.

The second presentation was by Luís Teles, Deputy Director, Statistics Department, at Banco de Portugal. He showed how Bank of Portugal developed **"StatFlix: the statistical course students won't want to miss!"**. For many years, Banco de Portugal has regularly gone to various economics and management faculties to talk about the statistics it produces. Some of these sessions focused more on concepts and the students commented that they would like to know more about how they can access the statistical series. Other sessions focused more on how to explore the Bank's statistics portal and thus easily access the statistical series, and in these sessions, students commented that it would also be useful to talk more about statistical concepts. The needs of students seemed unlimited and the resources to satisfy them – time and people – were scarce. After some brainstorming on the best way to address those needs, the winning idea was to develop an e-learning course on official statistics that combines concepts and how to access the data.

The course was designed with **six "seasons"** dedicated to explaining how GDP, inflation and public debt are calculated, whether families are more indebted or saving more, how corporate profitability is evolving or how the Portuguese economy interrelates with the economies of other countries. Each season has **five "episodes"** that help decode concepts and metrics and understand how statistics have evolved over time. During approximately three and a half hours, the students can explore the videos and flipbooks, listen to the podcasts, and challenge themselves in the quizzes that count towards the final grade of the course. Most of the contacted schools in different universities, entered with enthusiasm. In total, the project lasted around one year, considering: design the course structure, write texts, review, and rewrite them numerous times, produce scripts, graphics and other visualizations, make audio and video recordings, make translations into English, hire an external company to produce the course, test extensively the materials produced, fix bugs, etc. Around 50 people were involved, not only from Statistics, but also with the collaboration from other departments in specific subjects and from the National Statistics Institute. The project was greatly supported by the Board, and, above all, it was highly motivating for everyone involved.

The course was launched in October 2022, but only began to be used massively in the 2023-2024 school year. It was distributed by means of a **Shareable Content Object Reference Model** (SCORM) file, that is a zipped file that contains all the files needed to make online courses available in the Learning Management System of each school. Currently, 16 schools are on board and have integrated StatFlix in the curricula of their courses. The course is complemented with "special episodes" which are synchronous sessions on specific courses organized by Banco de Portugal. This **communication product** targeted the students of economics and management and was developed as a "user-first" approach. It addresses the proximity to society and the promotion of statistical literacy.

The third presentation, by Fernando Rocha, Head of Statistics Department at Banco Central do Brazil (BCB), was about **dissemination of statistics**, considering new channels, new approaches and new data formats. The UN Fundamental Principles of Official Statistics is on the roof of the governance system and refers to actions by statistical authorities to "facilitate a correct interpretation of the data". Compilers should be in a constant dialogue with user and target a broader the audience (young and non-experts). The strategy from the Central Bank do Brazil is to use new channels to reach this new audiences. Adopt new outreach approaches to data dissemination to enhance the understanding of datasets and **increase their use and targeted audiences**. Develop new data formats of a same dataset to disseminate the same information to different audiences, albeit with different levels of detail. In this regard, new channels and approaches are not ends but rather means of a data dissemination framework. The new approaches are essentially how the Bank seeks to engage users in the use of its data. This includes, for example: placing more emphasis on disseminating concepts and definitions, informing potential users of the service availability of the statistics disseminated by the BCB; and asking data users' and providers' views on data needs and data quality. The new formats seek to broaden the audience, catering to those who otherwise would not make use of statistics, and to enhance the service ability of statistics to those who already use it.

The old fashion is that all statistics disseminated by the BCB are published on the website: www.bcb.gov.br The Bank also resorts to several social media channels for the dissemination of data. The social media channels generally convey headline data, providing links to detailed statistics.

The **Open Data Portal** is machinery readable format. It also feeds into the website with interactive graphs and ability to download some data.

The new channels and formats consider Facebook, X (former Twitter), YouTube explaining some key concepts (inflation, market expectations and the focus report), Telegram, Blogs, LinkedIn and Instagram.

The main challenges include enhancing these aspects and ensuring that they lead to the service availability of data, such as: identifying the target audiences; pairing target audiences with the correct mix of format/channel/statistics; identifying the most cost-effective initiatives; ensuring that the format/channel/type of data are a means to an end, not an end in itself (careful with the "novelty frenzy"); and maintaining the focus of compilers on the quality of statistics, first and foremost.

The last presentation by Aldo Mascareño, Senior Researcher at Centro de Estudios Públicos (CEP) highlighted the opportunities and challenges of social media from a sociological approach. His argument is that after a decade and a half of actual global functioning of digital social media (including social network, internet, digital platforms, algorithm, IA) we are facing a radical change in our personal and social relations. We move from communication among persons, to communication between entities (algorithms and electronics) where we cannot identify human mediation. It has profound consequence on at least three level: impossibility of forgetting, politics, post democratic age and the level of truth. Putting things into perspective: **Evolution of social communication**.

Taking a brief look into the evolution of social communication, we observe every new communication media, produces a new form of society. The **printing press** lies at the basis of modern society. The diffusion of meanings through books decentralizes interpretation. The **mass media** in the 20th century led to massive propaganda and democratic pluralism in the western world. The **digital communication** of the 21st century turned into an actual experience the fact that we live in a one single world society with global experiences (financial crisis, climate change, pandemic). On the other hand, digital social media has produced multiple **identarian echo chambers** (believers in flat earth, or believers in God).

Our knowledge has clearly improved in the last decades. However, the more specialized our knowledge becomes, the more ignorance we face. So, there are no fake news, but a small social world communicating in social media without coordination among them.

The first application is related to **memory**, the right to be forgotten. Memory operates in the present. It is a present selection of past events. Its main function is forgetting, to release the capacities of the system. Since the spread of Web 2.0, with its virtually unlimited capacity to store and process data, the Web seems to allow for a form of perfect remembering. Indeed, our society seems to be able to remember everything. However, privacy is more relevant than the access to information. Internet does not forget. This is a new experience from the last century. Social memory must forget to remember and remembers different. Algorithmic memory cannot forget anything and can retrieve the same information once and again.

The second application is on **politics**, given knowledge has always been relevant to politics. The very foundation of the nation-state does not rely on violence – which is only applied in particular cases– but on data and the possibility of representation. Today we can complement our knowledge with artificial intelligence. However, the challenge is that with digital social networks in democratic contexts, politics loses its capacity for top-down communication and, therefore, its capacity for representation and intermediation of interests between the public and the administration. Political parties renounce their doctrines and become a platform for particularism –non-knowledge based politics, identity politics, post-democracy. Today, we have a wonderful opportunity to supplement our knowledge through artificial intelligence, to identify the biases in our datasets, and to train efficient algorithms that do not reproduce historic segregation.

The third application is to the **truth**. The expression "fake news" is just a manifestation of an ignorance gap between the well-known conditions of our daily experience as individuals and the world as a whole. The challenge is to determine the unacceptable risk. Regulation should include an iterative risk management system, with permanent updating and special consideration of technical knowledge, experience, and user training.

To conclude, while the printing press produced a symbolic differentiation that led to the emergence of modern society, digital communication has just begun transforming society. With the expansion of digital communication and the differentiation of modern society, the

production of both specialized knowledge and ignorance about the complex consequences of interdependence increases. Society becomes unpredictable. The question is not when there will be a generative AI –we already have GPT4–, but what is the minimum level of risk we can accept in dealing with social media and artificial intelligence.

6. CHALLENGES AND OPPORTUNITIES OF NEW INFORMATION

The seminar ended with a panel discussion on issues related to challenges and opportunities of new information, chaired by Pablo García, Vice Governor of the Central Bank of Chile and Chair of the BIS Irving Fisher Committee (IFC). The panel was composed by Paul Schreyer, Chief Statistician and Director of Statistics and Data Directorate at the OECD; Bruno Tissot, Head of Statistics and Research Support at the BIS and Head of the IFC Secretariat; Maciej Brzezinski, Head of the European Central Bank Data Office at the ECB; and Ricardo Vicuña, National Director, National Statistics Institute of Chile (INE). The discussions reviewed the impact of new data sources and techniques and digitalization. In the second round the panelists mentioned the evolution on their data governance frameworks triggered by the new challenges posed by accessing non-exploited data. Afterwards, they talked about the benefits of the availability of new data sources, including the possibility to construct alternative types of statistics, their ability to address new public demands for timely information, and the opportunities they provide in times of crisis such as during the recent pandemic. This also poses several challenges, for instance to communicate to different audiences about the relevance of the new types of statistics being produced and to avoid misinterpretation and limit the spread of fake news. Finally, they discussed the future priorities for statistical producers looking forward.

FIFTH STATISTICS CONFERENCE "INFORMATION TO UNLOCK THE FUTURE"

SANTIAGO, SEPTEMBER 26-27, 2023

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

Welcome, good morning, everyone and welcome to the Fifth Statistical Conference "Information to unlock de future", organized by the Central Bank of Chile. It is an honor to host this event that, since 2015. It gives us the opportunity to share experiences and reflect on the future of statistics and the challenges that information brings us to improve our analysis and decision making. This Conference brings together distinguished experts in the field of macroeconomic statistics, data and information production, and their users. I thank our international speakers and our local speakers, as well as all the attendees and those who are follow us via streaming.

We are living in an era of data and information. This period in history is marked by an unprecedented volume of information generated, on a massive scale. The digitization of our activities has given rise to a vast amount of data, covering extensive aspects of our lives. From the way we interact with our devices, the way we shop, we make purchases or transactions online. Every action leaves a digital footprint that adds up to an interconnected network of information. One of the most remarkable aspects of this revolution is the massification of information. Internet has become an immense resource, accessible for millions of people around the world. In fact, in July it was estimated that 5.192 million people (approximately 65% of the world's population) were using the Internet, with an average of six hours and forty minutes per day.

Furthermore, dealing with critical situations, such as the COVID-19 pandemic or climate change whose effects are becoming increasingly evident on the planet, together with the acceleration with which they have occurred, generates a complex scenario that poses important challenges. On the one hand, we must better measure and explain economic and social phenomena as they occur, properly reflecting the heterogeneous way in which this impact on the different agents of society. Digitalization of administrative processes and activities is a powerful source of background information, of great statistical richness and immensurable academic value. All of this comes hand in hand with the challenge of safeguarding the privacy of its owners, who generate the digital footprint. On the other hand, we must strengthen the publics' confidence in official statistics in the face of the disruption from multiple sources.

Therefore, at this Conference, we want to explore the opportunity offered by this new scenario, empowering, and promoting data-driven decision making, not only by governments and firms, but also by the general public. Until now, the abundance of data and the possibilities of connecting them has allowed us to improve analysis and decision making. Currently, many central banks and statistical agencies have been able to adapt their analytical thinking to the use of unconventional sources and big data techniques for processing.

The Central Bank of Chile is no exception. In our strategic plan for the period 2023-2027, we continue to deepen our use and analysis of data in all areas of the institution, in addition to developing new statistics that link economic activity with new dimensions, and in this way, we hope to continue contributing to the country with better analysis and decision-making regarding monetary policy and financial stability. Clearly, we are only at the threshold of the potential. An example of this, at a time when our biggest challenge is the convergence of inflation to the target, the characterization of the dynamics of firm's price adjustments, which through microdata from administrative records, has contributed to understanding the determinants of inflation. Recently, a decrease in the frequency of price increases has been observed, approaching historical averages, and an increase in the frequency of negative price adjustments by firms, which adds evidence to the inflation convergence process towards the 3% target. During this time, to understand this phenomenon we worked on a structural decomposition of CPI for goods, by distinguishing between supply and demand shocks which we complemented with and estimation of the half-life of the shocks to determine the persistence of inflation.

Likewise, during the pandemic access to information such as electronic invoices integrated with data of economic performance, access to banking credit and margins, made it possible to monitor at firm level the shocks that hit the economy and the effect of macroeconomic policies implemented during that time. Also, thanks to mentioned information, we have improved timeliness of our official statistics. These is the case of more disaggregated monthly results for activity and a quarterly frequency for GDP and household consumption at the regional level as of last year.

Most recently, in view of the reduction of the monetary policy rate the question rose to as how this lower MPR would be transmitted to market interest rate. Using, once again, microdata on commercial bank loans, it was possible to decompose the variation in rates into the extensive and intensive margin recognizing the predominant role of the latter. A pattern of interest rates was observed and that was been pass-through according to usual patterns for different size of firms determined by sales levels and different credit risk classifications. Therefore, this allows for early monitoring of a very important channel of monetary policy pass-through.

The wealth of this information has also led to new opportunities like the development of a set of experimental indicators that complement traditional macroeconomic statistics. Among others, there are the Regional Purchase and Sale Indicators (ICVR), published last May, which show transactions between firms, with breakdowns such as the region of sale or purchase, the sector of economic activity and size of the company; or the Retail Trade Daily Sales Index, which serves as a leading indicator of the monthly trade activity index (IAC), calculated by the National Statistics Institute, which is updated every two weeks, with daily data.

The diversity of economic, social and environmental challenges that we face demands an expansion in the production of statistics, to incorporate the measurement of emerging phenomena, such as climate change and its impact on economic analysis.

In this context, the Central Bank of Chile has been participating in international initiatives, such as the Network for the Greening of the Financial System (NGFS), among others, and has joined two initiatives based in Chile. The first is the Natural Capital Committee, an organization created by the Ministry of the Environment and supported by the National Council of Science, Technology, Knowledge, and Innovation for Development (CTCI), whose objective is to advise and provide recommendations for the measurement of natural assets in Chile. The second is an initiative led by the Ministry of Finance to develop a taxonomy, that is a system for classifying environmentally sustainable economic activities.

Another demand that has been highlighted by the sanitary crisis caused by Covid-19 is the need for an integrated multidimensional framework to better understand the shocks and the effect of policies, which can differ significantly across heterogeneous households. Thus, having distributional information on household income, consumption, and wealth, consistent with national accounts totals, provides a comprehensive view of economic phenomena, strengthening its understanding and analysis. Therefore, following the international recommendations, included in the System of National Accounts (SNA-2025), we also plan as an objective for this five-year period, to prepare and publish of disposable income, consumption, and household's savings by quintile.

This new information ecosystem, while opening the possibility of calculating new statistics and recognizing new patterns, also raises ethical questions regarding the privacy and confidentiality of information. In this context, the Central Bank of Chile has been, for some years now, promoting the appropriate and secure use of data, which we hope to continue strengthening. Starting in 2020, we began implementing an integrated and robust data management system aligned with international best practices. Our data model centralizes microdata bases and the administration of existing information exchange agreements, in addition to improving the technological infrastructure for queries involving large volumes of information.

But all this effort to offer new statistics that contribute to a better understanding and more timely monitoring of economic phenomena and, consequently, a more accurate evaluation of policies, will have little effect, if it is not known and understood by the public, constantly bombarded by data, often partial and biased, and which tarnishes their confidence in statistics. During this situation, communication of official statistics has taken on an increasingly critical role in society. Central banks and statistical agencies have a duty to be open to dialogue with users, especially to ensure that new generations see official statistics as a reliable and high-quality source of information.

In this goal of making statistics available to users in a massive, clear, and easy to understand manner, we have considered a strategy that includes getting closer to the regions and to segments of the population that are less familiar with statistics, especially the young. In the middle of this year, taking advantage of social networks, we launched the statistical account on Instagram, LinkedIn, and X (ex-Twitter), where we disseminate the main macroeconomic and financial indicators that we generate. We invite you to follow us on all these platforms.

This opens enormous possibilities for the development of statistics and a better response to economic challenge, contributing to the better design of public policies. These horizons multiply with the prospect of artificial intelligence. There is a vast road ahead of us with enormous potential and at the same time this poses great challenges. This process involves the handling of individual information which we must carry out under strict cannons of content and information security. We must be careful and self-regulate because it is an ethical duty and because the trust required by statistical processes merits it.

To conclude, I would like to stress that the role of central banks and statistical agencies in providing the information and data needed to understand the state of the economy, to help in the formulation of more informed policy and maintain the public's trust in institutions, is today more relevant than ever. The Central Bank of Chile will continue to contribute to the understanding of relevant macroeconomic phenomena based on the growing analytical possibilities opened by the availability of microdata.

I have no doubt that this conference will enrich our knowledge and broaden our perspectives on what it means to have reliable, comparable, and comprehensible information for everyone and show us the way to the future of statistics. I also like to thank the speakers, our guests and wish you a very fruitful conference. Thank you so much.

FIFTH STATISTICS CONFERENCE "INFORMATION TO UNLOCK THE FUTURE"

SANTIAGO, SEPTEMBER 26-27, 2023

GRANULAR, TIMELY AND BEYOND GDP: THOUGHTS ON THE NEW STATISTICAL LANDSCAPE

PAUL SCHREYER, CHIEF STATISTICIAN AND DIRECTOR OF STATISTICS AND DATA DIRECTORATE, OECD

Good morning, everyone. Its great to be here, actually to be back here, because I already had the pleasure of participating in this conference a couple of years ago and I have continued to find it extremely interesting, the topics very well chosen and the same is true for the current program. What I would like to do in the next half hour is to share a few thoughts with you on the statistical landscape as I perceive it, but I like to stress that there is no claim of completeness or an exhaustive treatment of the various topics in there. I was very pleased to note however that quiet a few of the thinks I will be saying trying very well with what we heard from the governor in terms of microdata, new sources and tools that are being used.

I will give you a little bit of background. I will then say a few words on what the main areas are, or on how we should see the contents of this new landscape of data that we are pursuing, move on to speak a little bit to the tools and sources that are available and then finish up with some considerations regarding the institutional setup that is needed to bring all these thinks to fusion.

The environment in which statistician are operating is evolving quickly, shaped by the digital transformation with issues of volume and ownership of data, and the debate about AI, we have new policy challenges – climate change, inequality… - that require cross-cutting policies. There are more diverse views by society on evidence and scientific knowledge.

The various shock and structural changes that we have experiences came with a strong demand for more timely, granular, and trusted data. For example, the financial crisis and pandemic required detailed and timely information to state how it affects different firms, locations, etc.

Where does this leave official statistics? The presenter structures three spheres to organize the thoughts: National Account, current well-being, and assets.

Beneath the well-established National Accounts, we need to pay more attention to household aggregates instead of economywide measures because they don't always move in the same direction. Another beneath GDP is to compute de Distributional National Accounts consistent with the national aggregates. Social research shows that health, income, and social connections are relevant for people. Therefore, it is fundamental to capture them correctly in well-being indicators, as already many OECD countries have developed dashboards of well-being indicators.



And better measures of assets to include non-produced- non-financial assets, like land and its value subsoil assets, other environmental assets (soil, water, etc). We can go further and think of ecosystem accounting where social capital and trust can be included. **Assets** are partially included in the updated of the SNA, especially when it comes to natural resources and the greater granularity that is been developed on environmental economy assets.

We want to develop data beneath and beyond GDP. Think of the linkages of this spheres to capture different environmental social economic aspects. The effect of digitalization and globalization that are partially and still not fully in the production sphere. Assets and how they change, not just income and expenditure. Finally spatial and socio-economic granularity that society and policy-makers are inquiring.

The **new data landscape** allows us to respond better the policy questions. For example, **rural-urban connectivity** divide. New timely and granular data show disparities in Internet speeds. Gaps in fixed download speeds experienced by users, by degree of urbanization. A second example is from geo-spatial data to assess land-use in OECD cities. AI is used to trainee the satellite information on with atlases, monitoring in near real-time land-use in OECD functional urban areas. A third example is a combination of data from satellite imaging with census data, to get measures of **exposure to air pollution,** combining new satellite data on air pollution (PM2.5) concentration with traditional census, survey or other spatial data on population distribution.

The main concern is **quality issues** when you use these new sources. Granularity entails challenges for privacy. Quality is multi-dimensional (intrinsic data quality, timeliness, accessibility, reproducibility, security). This means that the traditional quality frameworks may have to be revisited.

A word on **AI and statistics**. There are many opportunities associated with AI replacing human-intensive statistical activities (classification, imputation, NLP, user support,…), opening new methods (AI instead of econometrics for house prices, nowcasting), accessing data ecosystems, findability of official and other sources. It opens new possibilities for the official statistic institutions to retain, regain and revive if needed its role as an enabler of societal progress; counter-balancing misinformation. However, there are also challenges. It requires massive investments and innovations in statistical systems and efforts, new types of quality assurance, and change management in operations. It also requires the proper institutional setting. The discussion of **data stewardship** has been around for some year. Data stewardship is really the implementation of policies, standards, and principle of data governance. **Data governance** is the broader decision-making authority for data related matters in society. In this context is a discussion on who should be data steward in each country. The members of the national statistic system should be operating as data stewardship because they have a lot to offer legal basis, history of data ethics, quality frameworks, data protection, modern methods for data management, integration, metadata, visualization, and classifications. Involving members of the national statistical bodies in data governance helps reducing the trade-off between generating evidence with new data and protecting privacy on the other hand. Because you have professional bodies to deal with data protection.

If you think of the ingredients in this data ecosystem and the institutional setup the first one is a **strong legal and ethical basis**. You need the social acceptance and trust in the institutions. You need a Statistics Law and its proper implementation, data protection laws and their implementation, ethical standards that are credibly implemented and an ongoing dialogue with society's stakeholders. The second ingredient is **Common standards and inter-operability** of the national statistical system. We need to work on common classifications between different parts of the administration, metadata standards, structured access to administrative micro-data, common data quality frameworks and inter-operability framework. The access to administrative micro-data in Switzerland is centralized at the Data Science Competence Center (DSCC). The third ingredient is **Co-investment** given we are all facing the same challenges. Some examples where this is already happening is the HLG-MOS, a group of committed chief statisticians to steer the modernization of statistical organizations (CAN, AUS, IRL, ITA, MEX, NLD, NZL, POL, KOR, UK, Eurostat, OECD); the Statistical Information System Collaboration Community at the OECD with the co-financing of 15 members; the Development Data Partnership that facilitate the use of third-party data in research and international development cofounded by international organizations in partnership with private firms.

To conclude, there is a whole new world of data for evidence. There is an enormous demand for trusted and granular evidence. Digitalization brings many opportunities but also challenges. We needed strong legal and institutional basis for data governance, common standards, and interoperability. And finally, the co-investment NSOs, official statistics and IOs have an important role to play forward.

FIFTH STATISTICS CONFERENCE "INFORMATION TO UNLOCK THE FUTURE"

SANTIAGO, SEPTEMBER 26-27, 2023

INFORMATION TO UNLOCK POLICIES

MÁRIO CENTENO, GOVERNOR, BANCO DE PORTUGAL

Good morning, I want to thank the Central Bank of Chile for the invitation to be here today. I also try to unlock your spirit to the wonders of data and information and statistics which is a very important topic for central banks and certainly for Bank of Portugal.

Yesterday morning I was in "Museo de Arte Precolombino" and it was quite interesting to see how important it was already in the pre-Columbian time in this region the treatment of information. It was recorded in textiles in inventive ways of transporting information across territories with very strict and detailed methods using colors, using all sort of thinks that they had available to carry information over time, space, regions. Information is a challenge in our society, and we need to think of that with a good spirit, to take the challenges, to use more and more information and new technologies, to deal security personal data protection. The story of humanity is one dealing with data and information and we need to continue with that.

I used to say there will be no good policies if you don't have good data and a little bit of good economic analysis. Good data goes end on end with good policies. We need to take care of the basis. We like to see Bank of Portugal as an increasingly data-driven central bank. During the presentation I will also deep dive on the importance of data to ensure a smooth transition to low carbon economy and the challenges of artificial intelligence.

Good data is a prerequisite for good policies. Bank of Portugal is using data as a major asset for monetary policy, supervision, research, and the production of statistics. This returns value to the society.

One of the key issues we were able to develop is internal data sharing. We developed an **integrated data management program**, implemented a data governance model, and developed a data warehouse, supported by a data catalogue and a master and reference database. That allowed everybody within the Bank to use the same data. Additionally, we developed science labs, units that centralizes the information and teaches how to use the data. The Bank of Portugal Academy is a data science school to improve data skills to improve the quality of data usage.

There are always new data challenges. Microdata is the mainstream and has become more granular. The analyses and exploration of these new data landscape, requires new knowledge and techniques, which moved the bank to a big data project with the IT department. Of course, security and privacy continue to be a major concern.

AI and advanced analytics for Central Banks are a challenge. They may generate the idea of lack of trust. Some people may think information is being exposed to too many people with too few checks and balances. But it also generates tones of opportunities. AI are algorithms that need some guidance. They allow faster and efficient analysis of a growing volume of data, facilitating the decision-making process.

In the Bank of Portugal, we build some **business use cases** to generate creative solutions with **responsible AI frameworks**. This requires promoting data culture among the business areas. In Statistics AI is used to detect anomalies in data report in large datasets, e.g. with Central Credit Register, where IA has been of great use. In supervision AI is used in the classification of information requests and complaints from the public. The market analysis of media and research articles to extract market sentiment has also benefited from AI analysis.

To finalize, some thoughts on the importance of data to study **climate change**: the new source of financial risk. Firstly, it is important to get data . Data will always be imperfect, but we are working to further develop new ways of mapping the transmission channels to the economy. Given the experience of Central Banks dealing with data, we need to step up and provide a strong contribution to get reliable, granular, comparable, and relevant information. It is a global problem, so it requires global cooperation. The Network for Greening the Financial System (NGFS) is working on **bridging data gaps** and provide a common benchmark to regulators and supervisors at international level. Bringing public providers with public sourced data together with financial institutions' data is very important. All efforts are necessary to ensure international consistency and minimize overlapping climate disclosure standards.

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



Proceedings of the Fifth Statistics Conference
"Information to unlock the future"