



SDG ACCELERATION ROADMAP

UNLEASHING THE POWER OF
PRIVATE-SECTOR DATA IN THE GLOBAL SOUTH



Using Retail: Scanner Data to Estimate CPI in Chile

Using Retail Scanner Data to Estimate CPI in Chile

Reproduction

This publication may be partially or fully reproduced, in any format, for educational or nonprofit purposes, without special permission from the copyright holder, provided the source is cited. Cepei would appreciate a copy of any publication that uses this text as a source.

Main author

Hernán Muñoz, Cepei

Associated researcher

Tom Orrell, DataReady

Edition

Nicolás Marín, Cepei

Ricardo Corredor, Cepei

Executive office

Philipp Schönrock, Director, Cepei
Alexandra Roldán, Strategic Planning, Cepei

Translation

Interpreting Colombia

Graphic design

David Pachón Cortés

Credits

© Maps, photographs, and illustrations as specified within



Attribution 4.0 International

This work was carried out thanks to a grant from the International Development Research Centre (IDRC), Ottawa, Canada. The views expressed herein do not necessarily represent those of IDRC or its Board of Governors.

The SDG Acceleration Roadmap was developed in partnership with LIRNEasia and the support of the Mona School of Business & Management at the University of the West Indies in Jamaica, Local Development Research Institute (LDRI) in Kenya, and the Center for Continuing Education (CCE) at Birzeit University in Palestine.

Table of Contents

- 1. Acronyms and abbreviations -----4
- 2. Executive summary -----5
- 3. Using Retail Scanner Data to Estimate CPI in Chile: Findings
and Recommendations -----8
- 4. Part 1: Background and context-----10
- 5. Part 2: A public-private partnership for Scanner Data sharing -----12
- 6. Part 3: Findings and lessons learned -----17
- 7. Part 4: Conclusions and recommendations-----21
- 8. References ----- 24
- 9. Annex I - Methodological Note ----- 26
- 10. Annex II - Interview guide-----27

Acronyms and abbreviations

CPI	Consumer Price Index
CSR	Corporate social responsibility
ESG	Environmental, social, and governance
IDRC	International Development Research Centre
INE Chile	National Institute of Statistics of Chile
MoU	Memorandum of Understanding
NSO	National Statistical Office
SDG	Sustainable Development Goal

Executive summary

In 2021, Cepei and LIRNEasia launched the Sustainable Development Goal (SDG) Acceleration Roadmap project,¹ supported by the International Development Research Centre² (IDRC), and in partnership with the Local Development Research Initiative (LDRI) in Kenya,³ the Caribbean Open Institute (COI) in Jamaica,⁴ and the Centre for Continuing Education (CCE) in Palestine.⁵ With the midway point⁶ of the Sustainable Development Goals (SDGs)⁷ fast approaching, the project aims to understand just how far the private sector's data-related contributions to public policy in the Global South extend. Our primary research question is: **what is the private sector doing to make more and better data available to achieve and monitor the SDGs in the Global South?** To answer this question, over the course of 2022 and 2023, we are undertaking and publishing research and fostering global dialogues about the value of public-private partnerships in the data and digital policy spaces.

This document provides insights into the partnership between several Chilean retail companies that share scanner data⁸ with the National Institute of Statistics of Chile (INE), which combines these data with traditional survey data to produce the Consumer Price Index (CPI).

¹ Cepei. 2023. SDG Acceleration Roadmap Homepage. Online at: <https://cepei.org/en/initiatives/sdg-acceleration-road-map/>, accessed 2 March 2023.

² IDRC. 2023. IDRC – CRDI Homepage. Online at: <https://idrc.ca/en>, accessed 2 March 2023.

³ LDRI. Local Development Research Institute Homepage. Online at: <https://www.developlocal.org>, accessed 2 March 2023.

⁴ COI. 2023. The Caribbean Open Institute Homepage. Online at: <https://caribbeanopeninstitute.org>, accessed 2 March 2023.

⁵ Birzeit University. 2023. Center for Continuing Education Homepage. Online at: <https://www.birzeit.edu/en/community-affairs/institutes-centers/center-continuing-education>, accessed 2 March 2023.

⁶ UN. 2023. High-Level Political Forum on Sustainable Development. Online at: <https://hlpf.un.org/sdg-summit>, accessed 2 March 2023.

⁷ UN. 2023. Department of Economic and Social Affairs: Sustainable Development. Online at: <https://sdgs.un.org>, accessed 2 March 2023.

⁸ Scanner data refers to the electronic records of transactions captured at the point of sale, usually by bar code scanners at the checkout of retail stores. These records include information on the products purchased, their prices, and the quantities sold.

About the SDG Acceleration Roadmap Project

There is a significant shift taking place within the world of business. In recent years, environmental, social, and governance (ESG) has risen in prominence among an increasingly socially and environmentally conscious consumer base. ESG investing has attracted substantial interest, with reporting on environmental and sometimes social metrics becoming a must-do in many companies' annual reporting and shareholder meetings. Viewed from a public policy perspective, the turn towards more socially and environmentally conscious capitalism creates opportunities for alignment between business and government; for instance, between ESG metrics and political targets such as the SDGs that center on people, the planet, and prosperity as the three pillars of sustainable development.

One area of alignment is in the field of data and digital transformation. Whether framed as ESG, the SDGs, or corporate social responsibility (CSR), companies around the world are providing support to public sector institutions in ways that help to improve their capacity for evidence-based decision-making. Activities -or data actions- being taken range from the transfer of actionable data to the public sector directly through to the provision of tools or services that help improve the public sector's ability to capture and utilize insights from data repositories.⁹

Examples of public-private data actions abound. *Our Mapping Private Sector Contributions to the Data Revolution for Sustainable Development: Insights from the Global South*¹⁰ report ("mapping report") is based on insights from 394 data actions that have been mapped across 94 countries. These data can be analyzed and accessed directly through the project website.¹¹ What our mapping work uncovered is that there is a vibrant ecosystem of public-private activity around data.

⁹ See here for a full list of 'data actions' identified within this project: https://cepei.org/wp-content/uploads/2022/09/Terminologia_Data_Actions-ENG.pdf

¹⁰ Cepei. 2022. Mapping Private Sector Contributions to the Data Revolution for Sustainable Development: Insights from the Global South. Online at: <https://cepei.org/wp-content/uploads/2022/09/Mapping-private-sector-contributions-to-the-data-revolution-for-sustainable-development.pdf>, accessed 2 March 2023.

¹¹ Online here: <https://cepei.org/en/initiatives/sdg-acceleration-road-map/>

Our Case Studies

To further explore the themes identified in our mapping report and uncover how the mechanics of public-private data partnerships operate – what the incentives for partnership are, how the impact is measured, what enabling environment needs to exist, and other factors – we undertook a series of eight case studies in late 2022 and early 2023.

Our case studies showcase examples of public-private data partnerships and document how companies' data actions can help public institutions in the Global South to respond to major public policy challenges such as climate change, the promotion of gender equality, improving employment opportunities, and digital literacy, among others.

We selected our case studies in a way that ensured that we covered multiple: types of data, types of partnership/partnership facilitation, types of data action, company size, and thematic area. Thereafter, each project partner independently produced their case study(ies) using a common semi-structured interview guide and desk-based research. Case study-specific methodological considerations are explored in more detail within each study. Through our case studies, we have further developed and refined the themes identified in our mapping report and proposed recommendations based on them.

In summary, the five overarching themes that have emerged from our case studies are:

1. There is real-world value being produced from public-private data partnerships.
2. Companies that can offer solutions to more than one data-related challenge are more attractive partners for public bodies.
3. Identifying and measuring the impact of public-private data partnerships is challenging and it takes time and energy for processes to effectively do this to emerge.

4. Business support to public initiatives and core business operations are not mutually exclusive.
5. A significant barrier to initiating, implementing, monitoring and scaling-up public-private data partnerships is the lack of standard operating procedures to develop partnerships.

Using Retail Scanner Data to Estimate CPI in Chile: Findings and Recommendations

The National Institute of Statistics (INE) of Chile is the institution responsible for producing and publishing Chilean official statistics, including monthly updates to its Consumer Price Index (CPI). CPI is “an economic indicator that measures monthly variation of prices in a basket of goods and services representative of the consumption of urban households [...]” It is a widely used indicator of inflation. The CPI is relevant for monitoring the Sustainable Development Goals (SDGs) because it reflects the cost of living for consumers, which is an important aspect of the economic and social development of a country. In particular, the CPI can provide information on the affordability of basic needs, such as food, housing, and healthcare, which are key elements of SDG 1 (no poverty) and SDG 3 (good health and well-being). Moreover, the CPI can be used to monitor progress towards SDG 8 (decent work and economic growth), as it provides information on the impact of inflation on wages and the purchasing power of consumers.

Traditionally, data needed to calculate CPI have been collected through in-person surveys, a process that is time consuming and expensive to repeat monthly. However, in recent years, countries including Australia and France among others have been experimenting with the use of the scanner data from retailers on each sale as a means of supplementing or even replacing data collected in person. Scanner data provides a unique source of systemic data on price and sales volume that can be disaggregated by point of sale (location) and time. As such, they are a rich and accurate source of data for statisticians working on CPI.

In Chile, as in many countries, the COVID-19 pandemic severely limited the ability of INE to collect data through traditional methods such as in-person surveys. As a result of this, learning from the models used elsewhere, INE decided to innovate with the use of scanner data to produce its CPI. INE's journey in this regard started in 2019 and is still evolving. It took time for relationships to be formed with key private sector data providers and for memorandum of understanding (MoUs) to be negotiated with them.

The outcome is impressive. INE now has working relationships with four major retail conglomerates. The sales records that are relayed from private companies can provide up to 40% of the data needed for INE to calculate the CPI. Moreover, while other types of big data tend to complement official statistics rather than replacing them, scanner data can directly replace survey-based data meaning that statistical concepts and methodologies do not need to be modified. Moreover, scanner data offer richer insights into consumer behaviour that can help guide economic and fiscal policy. This is because scanner data records provide price information on more products than it would be feasible to collect data on in person (i.e., they offer insights into a broader basket of goods) and they also provide data on the quantity of goods sold and purchased. The data are also near-real time, making them extremely valuable not just for CPI calculation but potentially also for the tracking of short-term fluctuations in inflation.

Notwithstanding the benefits to policymaking of datasets such as those produced through the scanning of barcodes, in Chile –as in many other countries around the world– there are barriers to the use of business-produced data. To establish and formalise partnerships with its private sector data providers, INE has had to individually negotiate MoUs with each one, securing: access to relevant datasets at no cost, the inclusion of a broad basket of goods within the dataset, a sufficiently clear aggregation of goods to ensure comparability across data providers, an agreed schedule for the timely sharing of updated datasets, and clarity around who the primary point of contact at each data provider should be. The time and effort taken to negotiate and enter into these agreements was one of the major limitations relayed to us by interviewees when exploring this case study.

Despite the challenges, the use of scanner data by INE in Chile to calculate CPI monthly has been a great success. It is contributing to the creation of statistics that are invaluable for both government officials responsible for setting economic and fiscal policy and reporting on SDGs indicators, specially under Goal 1, 3 and 8. The public private partnerships that INE has entered now offer a blueprint for potential future expansion and replication in other areas where private sector data are deemed to have public value.

The document is structured as follows: A first part describes the background and the context in which the initiative took place; a second part contains the characteristics of INE and the data initiative, as well as, the reasons why this case is relevant to understand the role of the private sector in the region; a third part contains the findings and lessons learned that emerged from the interviews; and finally a series of conclusions and recommendations for future actions. Methodological notes on how this case was built are included in annexes 1 and 2.

Part 1: Background and context

The 2030 Agenda for Sustainable Development posed a series of challenges in terms of data needs that, in turn, represented an opportunity to make innovations and improvements, often neglected, in the supply of data for sustainable development. The use of non-traditional sources such as financial transaction data, mobile networks, satellite imagery, sensors, or Internet platforms can substantially increase the quality and reduce the cost of generating the information required for monitoring and implementing the SDGs.

In addition, “traditional” statistical surveys and censuses can be complemented with new sources of information. This trend gained momentum during the COVID-19 pandemic, when data producers faced sudden interruptions of traditional collection sources and had to develop new methodologies for using

¹² Data ecosystem is the name usually given to the place where the processes of production, consumption, and transfer of data are generated.

alternative sources, thus spontaneously extending the data ecosystem.¹²

However, the data revolution not only implied a quantitative leap in data supply, but also a series of new unavoidable demands: greater relevance (data on new phenomena), timeliness (real-time data), and coverage, including granularity (new data sources allow high levels of disaggregation), in addition to the reduction of costs and response burden.

In this context, and to find out what the private sector is doing to make more and better data available for the SDGs in Latin America, Cepei, in the framework of a global project funded by IDRC, has mapped private sector data initiatives that contribute to strengthening data capacities to monitor and achieve the SDGs.¹³ On that basis, a series of case studies have been chosen that can provide relevant lessons and learning for the data community and have replicability, scalability, and co-creation characteristics.

Based on these premises, in Latin America, we have been able to identify 42 data initiatives distributed in 14 countries,¹⁴ as well as some from large multinational companies with a region-wide reach. It is likely that many initiatives that are not visible or are visible, but in ways that do not fit the search methods used for this exercise, have not been captured. Most of the initiatives identified are related to data sharing, either directly or through actions such as data analysis, capacity building, or impact reports that include some levels of private data opening.

In cases where public-private partnerships have been identified, it is useful to distinguish between partnerships that seek to enhance an alternative data source to measure one-off phenomena from partnerships that aim to measure them on a regular basis. In this sense, data initiatives that aim to measure phenomena on a regular basis require sustainable collaboration, which must be backed by

¹³ For the purposes of this report, this concept includes capacity building, skills sharing, data collection, data sharing, data analysis, data infrastructure, data governance, data mapping, funding, among others.

¹⁴ Argentina, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, México, Nicaragua, Panama, Peru, Uruguay and Venezuela

enforceable rights and obligations between the parties.

The launch of barcode technology during the 1970s, and its growth in the 20th century, made it possible to capture detailed information about transactions at the point of sale. Scanner Data records are high-volume and contain information on individual transactions, date, quantities and values of products sold, and product descriptions. As such, it constitutes a rich source of data for National Statistical Offices (NSOs) that can potentially be used to improve their statistics and reduce the reporting burden and associated costs in physical data collection. This motivated the interest in leveraging Scanner Data in official inflation indices in countries such as Australia, France, the Netherlands, New Zealand, Sweden, Switzerland, and Italy which have already implemented it in their official measurements (See [6]).

Part 2: A public-private partnership for Scanner Data sharing

2.1 About the public-private stakeholders

The National Statistical Institute of Chile (INE Chile) is the office in charge of conducting the general population and housing censuses, and producing, compiling and publishing the country's official statistics. INE has a staff of more than 2,500 people, of which 600 are field surveyors who travel the territory in search of data that will later be processed, analyzed, and disseminated by professionals and technicians of the Institute. Its mission seeks the elaboration and dissemination of reliable, timely, accessible, relevant and comparable information at a national and international level that serves the elaboration of public policies and accountability to citizens. In total, it periodically disseminates more than 70 statistical products, of which one of the main ones, both for its own relevance and its impact on other indicators, is the Consumer Price Index.

The partner companies

In order for INE to benefit from the data held by the private sector, it is important to maintain an open and close dialogue with focal points in the companies, who

were generally the entry point for establishing Memorandums of Understanding (MoU). In those MoUs, responsibilities, and rights were established, as well as it was established that the use of the shared data is limited to the production of official statistics and protected by Statistical Secrecy and information security policies.

The network of retail companies that currently contribute their data to the INE has been growing since the first contact with the main companies in the sector between 2019 and 2020. By the end of 2022, MoUs have been signed with four major supermarket chains, while agreements continue to be managed for the transfer of data from another supermarket, a multi-pharmacy, a retail company, and two pharmacies with a national presence. The names of the companies involved cannot be explicitly mentioned for reasons of confidentiality of INE Chile.

2.2 About the Data Initiative

Scanner Data is collected by large retail stores when consumers go to pay for their products. It provides a unique source of systematic data on prices and quantities of specific products disaggregated by point of sale and time. This means for NSOs, they have the possibility to improve their statistics, and reduce the reporting burden and costs of the physical data collection process, which usually covers only a subset of products and does not contain information on disaggregated product sales or quantities (See [7]). These innovations, however, must be carefully managed, both in terms of the impact on the statistical process and on the relationship and communication with users and key stakeholders.

In the case of INE Chile, the initial motivation for accessing Scanner Data was to achieve greater efficiency in the production of the Consumer Price Index, reducing the need for face-to-face visits to establishments. This objective was reinforced during the Covid-19 pandemic, given the difficulty of conducting the surveys. The pandemic situation also opened a window of opportunity for public-private partnerships, given the increased awareness of companies on the need to contribute their data for the common good. Since 2019, INE had been trying to make arrangements with various companies to obtain data on their daily sales and prices for products that are part of the Consumer Price

Index basket. However, it was not until 2020, that these efforts became a priority for both parties.

The sequential process of incorporating Scanner Data into the CPI began in 2020 with the transfer of data sets from some of the main supermarket chains and convenience stores, the implementation of a secure data transfer process (SFTP), and the incorporation of this data into the calculation of some CPI divisions. In 2021, the frequency of data transfers from supermarkets was increased from bi-weekly to weekly, and several MoUs were signed with the companies. During 2022, the efforts were extended to new sectors such as pharmacies and other retail chains, and improvements were implemented in the technological support for information storage.

The transmission of the data sets is done through SPSS folders where the companies leave weekly the files containing the price, volume, and type of product of all transactions carried out during that week disaggregated by point of sale. The company must prepare the file and load it into standard SQL tables, which implies an allocation of extra resources for this purpose.

The volume of data available allows for building an algorithm that recognizes the description of the CPI varieties and extracts the barcode information from the external source to update the data in the internal list and ensure the match between the data sources. In this regard, some difficulties have been encountered due to problems with the key variable (product barcode), which indicates the need to keep the internal inputs for processing the transferred data updated periodically, ideally through an automated mechanism for assigning codes, but so far has been done mainly manually.

2.3 Rationale for the case selection

Currently, countries such as Australia, France, the Netherlands, New Zealand, Sweden, Switzerland, and Italy have incorporated Scanner Data in their official CPI measurements. However, the data initiative being undertaken by INE Chile is an innovative experience for the Latin American region, not only in specific terms of improving the CPI with Scanner Data, but also in terms of a concrete and lasting public-private data collaboration, given that in both aspects the

region is far behind in taking full advantage of the opportunities that the data revolution offers (see [5]).

This collaboration will result in reduced costs for INE and less burden on private sector informants. Further, the availability of Scanner Data offers a number of opportunities to improve CPI accuracy. Scanner data sets typically contain complete coverage of the items sold by a retailer in all of its stores; as well as the quantities of items sold and the revenue received by the retailer for these items. This information has the potential to: improve the accuracy of prices used to compile the CPI by calculating unit values for homogeneous products; improve the samples of items to which the price is applied, with the potential to use a census of items sold to compile the CPI; and use quantity/revenue information to weight items according to their economic importance [6].

The use of the scanner datasets in the CPI calculation may represent a significant change from the data collection practices and price index production methods traditionally employed by the INE, which estimates that the CPI collection system can draw up to 40% of its records from Scanner Data held by the private sector. While other forms of big data use tend to aim at producing new statistics that complement existing public statistics rather than replacing them, a specific aspect of Scanner Data is that they can actually replace survey data without the need to modify the concepts or methodological framework of what is to be measured [7].

The Director of INE Chile has stated [9] that “having administrative records, whether from the public or the private sector, allows gaining in efficiency and opportunity that undoubtedly improves the lives of all people”. At the same time, she emphasizes “the richness of sharing public and private information to make the production of information that contributes to improving public policy more efficient” (See [9]).

The COVID-19 pandemic proved to be an accelerator for this process given the need to reduce field data capture due to mobility constraints. Therefore, this

case has lessons on the correct detection of windows of opportunity for the implementation of innovations.

2.4 Regional context for the use of private sector data

From the mapping exercise carried out in the framework of this project, together with a recent diagnosis conducted by Cepei on data ecosystems in the region [5], we can conclude that few private initiatives have been systematically incorporated into data ecosystems for the SDGs in the countries. Currently, the involvement of the private sector to strengthen data ecosystems is rather focused on individual and isolated actions by companies, with a greater or lesser degree of partnership with the public sector and that in general do not have guarantees of continuity.

Likewise, we have found that efforts to contribute to the public good through the use of non-traditional data sources from the private sector are, in general, in exploratory stages and are often diluted over time. Therefore, the sustainability of these types of initiatives is still uncertain. The generation of reliable and sustainable information within data ecosystems requires an enabling environment, understood as all the resources, processes and systems, legal norms, incentives, and information on the actors involved in the data generation processes. Many of the current challenges must be resolved through the generation of incentives and legislation that promotes public-private partnerships.

In this regard, it has been identified in the region that regulatory frameworks, in general, must evolve in order to support this practice. The development of laws that enable the reuse of privately owned data is a fundamental aspect for the future of publicly available data. Likewise, the intensive use of administrative records and geospatial data are among the strategic priorities of official data producers, such as NSOs (See [5]).

In the region, there are two major determinants that affect the participation of the private sector in the data agenda for the SDGs, such as a possible loss of

momentum and interest in the 2030 Agenda by the private sector, as well as the lack of an enabling environment conducive to public-private initiatives. That is, there is data and there are initiatives related to the SDGs, but there is a lack of data initiatives. Companies, in general, are leaning towards reporting their own impact on the SDGs through data collection and analysis, but this does not necessarily imply a strengthening of the data ecosystem for monitoring and implementing the 2030 Agenda.

Part 3: Findings and lessons learned

3.1 The use of Scanner Data: Pros and cons

Traditionally, national statistical institutes have produced official measures of CPIs using price in-person collection, and expenditure weights for product groups that are based on household expenditure surveys. The resulting index provides a picture of inflation experienced by a “representative” consumer, which implies limitations to understand the effects of inflation on households.

Studies that have taken advantage of scanner data to measure inflation have been able to address several of the limitations inherent in official CPIs. Scanner Data records provide price information on many more products than it would be feasible to collect prices in person but also contains information on the quantities acquired. Thus, this data can provide a measure of the consumer’s ability to adapt to changes in relative prices, and expenditure weights at the disaggregated product level, allowing the importance of products within their groups to be weighted and expenditure weights to be updated frequently. Another important advantage of scanner data for measuring inflation is that it is available in near real-time. Finally, Scanner Data records can allow the development of new statistics (e.g., regional indexes and spatial price comparisons) and the measurement of phenomena related to household consumption patterns [10].

¹⁵ The traditional CPIs update expenditure weights based on Consumer Expenditure Surveys, which are usually conducted every 5 years and are available at least one year in delay, given the need for processing.

In terms of the statistical process, scanner datasets present many potential advantages, as they can be used to: (i) cross-check with physically collected data; (ii) replace field-collected prices; (iii) expand price samples; (iv) weight products at lower disaggregation levels; and (v) apply new methods that possess desirable properties for price indices and allow for process automation (see [6]).

To do so, the NSOs must be able to negotiate appropriately (see [7]):

- the provision of the dataset at no cost;
- a broad scope of items included in the dataset;
- a level of aggregation of the items to ensure homogeneous information;
- an agreed schedule for the provision of the dataset to meet CPI processing requirements and dissemination calendar; and
- a contact person within the retail company who is familiar with the dataset to respond to queries.

All these points have been taken into account by INE Chile to establish the current agreements with the reporting companies.

3.2 Challenges and opportunities

INE Chile CPI team has stated in the interviews that they identify short-term challenges in at least two main areas: the consolidation of the collection method and its progressive integration into the official CPI calculation.

Bilateral negotiation of datasets is one of the main challenges. The experiences of the Netherlands, Switzerland, and Australia indicate that these negotiations can take several months (See [11]). The negotiations concern a wide range of issues: from computer systems and formats to confidentiality issues.

In INE's experience, it has also proved difficult to agree with companies on regular data delivery, according to the requested structure. The agreements reached between INE and the retail companies were formalized in Memorandums of Understanding containing the rights and obligations of each party with the objective of ensuring a continuous supply of data according to a schedule and

completeness. In this way, an attempt was made to control the risk of not receiving in due time and form the information required to address the replacements in the calculation.

In cases where there is no signed agreement, it has been more difficult to obtain a sustainable commitment from the companies. It is necessary to evaluate strategies to strengthen the relationship between the Institute and the private sector, emphasizing the responsibility for the delivery of data for statistical production, and being able to transmit the solutions that are designed and implemented to solve any doubt or problem regarding information security, as well as the importance of the use of records for statistical production.

The consolidation of the collection method is directly dependent on the strengthening of public-private cooperation for the production of statistics based on registers; therefore, awareness of the relevance of public-private cooperation for the production of official statistics is required. In this sense, **the Covid-19 pandemic was a catalyst that accelerated this awareness, but from now on, concrete actions should be taken to promote data sharing and generate an environment of trust among the stakeholders.**

Likewise, this initiative requires the development of institutional technical and technological capacities in the INE to improve coverage in the use of data. On the one hand, it is necessary to coordinate internal coordination for the implementation of a data lake to store and extract the transferred data. Work is also underway to improve the linkage mechanism between Scanner Data and CPI products. This process requires designing an automated coding and matching automatic process for the selection of data to be incorporated in the month-to-month calculation.

In some countries, such as France (see [7]), the incorporation of Scanner Data into official statistics required the modification of the statistical legislation regarding the possibility and/or obligation to exchange certain data held by the private sector. In the case of INE Chile, as reported in the interviews, no change in legislation was necessary.

According to the current law, companies are obliged to provide information to INE. **In any case, the strategy of establishing agreements aims to underpin the sustainability of the process through collaboration and leave legal enforcement to a later stage.**

Once the data has been obtained, the biggest challenges focus on converting these datasets into information that can be used in the CPI, through a technological system suitable for processing, classifying the data received homogeneously because the datasets often have product classifications that are unique to each company and must be assigned to a single CPI classification, and ensuring the quality of the Scanner Data records received. Among others, special attention must be paid to changes in the composition of products sold and their qualities.

Some companies required a process of reorganization of their internal protocols since they must deliver databases with fields defined by INE. The work of reconciliation and homologation of the descriptions was quite arduous, due to the matching required with the INE's "product master" and the use of the bar code as a binding variable. The INE team stated that the data exchange and database interoperability format still has many manual elements and must be improved.

In the case of INE Chile, Scanner Data currently constitutes 3.5% of the total prices that enter the monthly calculation of the CPI and INE estimates that it can reach a coverage of 40% of the CPI with Scanner Data (See [9]). For this to happen, INE considers that it is important to design a plan for the progressive incorporation of data in the calculation, to advance in the definitive change of the collection method (from face-to-face to scanner data), in addition to establishing mechanisms that take advantage of these records for updates (base year change projects) and definition of variety replacements in establishments where possible and to increase the coverage of replacements [9]. Given that, this plan implies important changes in the ways of producing the CPI, a project in this sense implies working on change management, with the purpose of establishing the use of this technique as an official collection method.

There is also the opportunity of putting the scanner datasets to other uses to gain a greater understanding of consumption patterns or to develop new price indicators. But INE Chile stated that, although the value of having disaggregated consumption information for other economic surveys has been discussed, for the time being, they will not consider the information to be of any other use. There are also no plans to have a CPI with a greater periodicity (biweekly or weekly).

Part 4: Conclusions and recommendations

The CPI is relevant for monitoring the Sustainable Development Goals (SDGs) because it reflects the cost of living for consumers, which is an important aspect of the economic and social development of a country. In particular, the CPI can provide information on the affordability of basic needs, such as food, housing, and healthcare, which are key elements of SDG 1 (no poverty) and SDG 3 (good health and well-being). Moreover, the CPI can be used to monitor progress towards SDG 8 (decent work and economic growth), as it provides information on the impact of inflation on wages and the purchasing power of consumers.

The case of INE Chile presents a series of lessons and recommendations for similar initiatives to be undertaken in other countries that could have a similar impact on their SDGs indicators.

As stated by the INE Chile CPI team, the current advantages of this public-private cooperation for the sharing of Scanner Data implied for INE Chile a reduction of costs of the CPI survey operation, new information about consumption data yet to be explored, the possibility of more active cooperation with companies, exploration and introduction of working with high volumes of data, development of the team's own skills, storage capacities, streamlining of the process, and the possibility of having real-time data. From the private sector perspective, it will imply less burden as reporters of official statistics and the possibility of being able to contribute their data to the public good.

A key success factor is associated with the strategy used to engage with companies. In the case of INE Chile, it consisted of taking advantage of existing contacts with the reporting focal points in the companies for the statistics that INE produces. These people generally already knew the working protocols with INE and paved the way for building trust between the two parties.

In this way, they were able to obtain a positive response from a first company that understood the importance of this contribution. Other companies have been more reticent, but another learning in this sense is related to the herd effect that was generated as more companies joined in sharing their scanner data. Thus, barriers of mistrust have been broken down and the implementation of learning has been perfected. Nowadays, contacts are mostly made with letters from and to the highest hierarchical level of the institutions. But it started with the operational level, both from INE and from the companies.

From the perspective of the INE Chile team, the tool for approaching companies should include contextualizing them about statistical production and the need to modernize their databases, as well as, incentives to reduce the burden on the informant. For example, INE Chile offers the opportunity to free the company from answering the structural survey once a year.

It is also very important to guarantee the security that the information will not be leaked by establishing SFPT protocols for data encryption. In this sense, INE Chile commented in the interviews that one partner company, after the delivery of the first dataset, entirely changed the people who were doing the exchange and could not resume contact, even insisting from the legal area. They believe that there was some episode of distrust about the use of this data from the management area. For this reason, they consider that signing Memorandums of Understanding between the highest hierarchical levels is the most effective way to ensure the flow of data. Therefore, **a recommendation in this sense refers to avoid relying only on the willingness to collaborate and formalize it as long as the conditions are in place.** People and organizational structures usually change, so the source of information may be put at risk or the relationship may have to start from scratch, something that has happened to INE Chile on several occasions.

Another recommendation arising from this case study refers to the identification of windows of opportunity for the implementation of this type of initiative. In this regard, although INE Chile had already started during 2019 to conduct tests with one of the large supermarkets, it was only after the start of the Covid-19 pandemic that companies began to understand the real need for their data and could definitely move forward in consolidating data transfer.

Finally, in terms of interoperability, it is important to take into account that for data sharing, companies do not generate their internal databases based on statistical needs. Some companies require a process of reorganization of their internal protocols, which should be foreseen and presented as an investment that will result in a reduction of the reporting burden in the future by automating the transfer.

References

[1] United Nations Independent Expert Advisory Group. 2014. "A world that counts--Mobilising the data revolution for sustainable development". Available at:

<https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>

[2] United Nations. 2013. "A new global partnership: eradicate poverty and transform economies through sustainable development. The Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda".

[3] United Nations. 2020. "The Age of Digital Interdependence, Report of the United Nations Secretary-General High-Level Panel on Digital Cooperation". Available at:

https://www.un.org/en/content/digital-cooperation-roadmap/assets/pdf/Roadmap_for_Digital_Cooperation_EN.pdf

[4] United Nations Secretary-General. 2020. "Road map for digital cooperation: implementation of the recommendations of the High-level Panel on Digital Cooperation: report of the Secretary-General".

[5] Cepei. 2022. "The state of Data Ecosystems in Latin America". Available at:

<https://cepei.org/documents/estado-ecosistemas-datos-america-latina/>

[6] IMF. 2020. "Consumer price index manual : concepts and methods". Available at:

<https://www.imf.org/en/Data/Statistics/cpi-manual>

[7] Marie Leclair. 2019. "Using Scanner Data to Calculate the Consumer Price Index" INSEE. Available at:

<https://www.insee.fr/en/information/5014725?sommaire=5014796>

[8] DellaVigna, S. and M. Gentzkow (2019). Uniform pricing in US retail chains. Quarterly Journal of Economics 134 (4).

[9] Quijada S., 2022, "Uso de registros de Scanner Data en el cálculo del Índice De

Precios al Consumidor - INE Chile”, presentado en la “Conferencia de Alto Nivel sobre el Intercambio Efectivo de Datos Públicos y Privados para la Formulación de Políticas basadas en evidencia” organizada por el INE de Uruguay y GPSDD en noviembre de 2022. Main messages available here:

<https://www.ine.gob.cl/sala-de-prensa/prensa/general/noticia/2022/11/23/directora-nacional-del-ine-expone-en-uruguay-sobre-producci%C3%B3n-estad%C3%ADstica-y-scanner-data>

[10] Pierre Dubois, Rachel Griffith and Martin O’Connell. 2022. The Use of Scanner Data for Economics Research. Annual Review of Economics, Available at:

https://pierredubois.github.io/DGO_ScannerData.pdf

[11] Australian Bureau of Statistics (ABS) (2013), “Feature Article: The Use of Transactions Data to Compile the Australian Consumer Price Index”, 6401.0, 23 October 2013, Canberra: ABS.

[12] Eizenberg, A., S. Lach, and M. Oren-Yiftach (2021). Retail Prices in a City. American Economic Journal. Economic Policy 13 (2), 175–206.

[13] Ivancic, L., K. J. Fox, and E. W. Diewert (2011). Scanner data, time aggregation and the construction of price indexes. Journal of Econometrics 161 (1), 24–35.

[14] Jaravel, X. and M. O’Connell (2020). Real-time price indices: Inflation spike and falling product variety during the Great Lockdown. Journal of Public Economics 191, 104270

[15] van der Grient, H.A. and J. de Haan (2010), “The Use of Supermarket Scanner Data in the Dutch CPI”, Paper presented at the ECE/ILO Workshop on Scanner Data, 10 May 2010, Geneva, Switzerland.

Annex I - Methodological Note

This Case study is part of a larger research project that is seeking to understand how far the private sector's data-related contributions to public policy in the Global South extend, especially as it relates to the Sustainable Development Goals (SDGs). The primary research question is: **what is the private sector doing to contribute to more effective data ecosystems that help to achieve and monitor the SDGs in the Global South?**

The Research has examined examples of private sector support to public institutions across Latin America and the Caribbean, The Middle East and North Africa (MENA), Sub-Saharan Africa, and Southeast Asia, initially through the structured mapping of a range of public-private data partnerships. In the second phase of the study, emerging insights from this mapping have been examined in more detail through the production of eight in-depth case studies of public-private data partnerships. This data initiative on the use of scanner data for the CPI at INE Chile has been selected as one of the two Latin American case studies.

The elaboration of this case study was based on a review of documentation on the initiative, media articles, and a series of interviews with the protagonists. In that sense, two interviews were held with the INE Chile CPI and International Relations team. No interviews could be held with the private sector counterparts since the names of the companies and the focal points involved cannot be informed by INE Chile for reasons of confidentiality. The below listed triggering questions were used in the interviews to INE officials.

Interview Sessions INE Chile – Dec-Jan 2023

1st Interview International Relations and CPI Team - 20/12/22

2nd Interview CPI Team – 10/1/23

Participants: Alejandra Paz García Mozo, Rodrigo Andrés Glade Gonzales, Felipe Andrés Lopez Borges, Cristian Andres Copaja Espina, Verónica Padilla Gonzalez, Macarena Trivino Melio.

Annex II - Interview guide

- A) Introduction to the project...
 - Description of the partnership - How did it start?
 - Description of the thematic area – What is the state of the art?
 - Description of the data action – Describe the main features of the data initiative and how it contributes to the public good.
- B) Impact on the SDG Agenda...
 - How relevant is the Sustainable Development Agenda in your organization’s vision?
 - Which SDGs goals are most impacted by this initiative?
 - How do you think your initiative can scale up its impact on the SDGs?
 - Why do you think this case study is important for showing the Private Sector’s role in the SDGs?
- C) Findings and lessons...
 - Why is this particular data action important – what gap is it filling?
 - What are the main challenges and opportunities for this particular type of data action? E.g., privacy concerns on the data, government bureaucracy, copyrights of the data satellite imagery, and technology gaps.
 - Are there any general thematic guidance documents, policies, corporate standards, etc that are relevant to the particular data actions?
 - What were your incentives/strategic vision for getting involved in this initiative?
 - What worked? What didn't? Why? What is replicable and what is not?
- D) Enabling environment...
 - What does the local enabling environment look like?
 - Are there domestic laws, policies, favorable political statements, etc. that support this particular data action/case study?
 - What characteristics of the local environment support or constrain a program like this?
- E) Partnerships...
 - How did you look for partners? How was your engagement strategy?
 - What are the lessons learned that might be useful for similar initiatives?
 - Highlight specific points as relevant to the case study itself



SDG ACCELERATION ROADMAP

UNLEASHING THE POWER OF
PRIVATE-SECTOR DATA IN THE GLOBAL SOUTH

✉ Please get in touch with us at
direccion@cepei.org

www.cepei.org

✉ @infoCEPEI

f @infoCEPEI

📺 @infoCEPEI

🌐 @infoCEPEI