



State of Data Ecosystems in Latin America:

Analysis and recommendations
for their development

In partnership with



**Global
Partnership**
for Sustainable
Development Data



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. The use of this document for resale or any other commercial purpose without prior written permission from Cepei is not permitted.

Suggested citation

Centro de Pensamiento Estratégico Internacional. (2022). *State of Data Ecosystems in Latin America: Analysis and recommendations for their development*. Cepei

Main author

Hernán Muñoz, Cepei

Contributing author

Andrés Clavijo

Translation

Priscila McCarthy

Graphic design

Marisol Valencia, Cepei

Credits

© Maps, photographs, and illustrations as specified within

-Creative Commons-

Table of Content

1. INTRODUCTION	4
2. BACKGROUND	6
3. METHODOLOGY	7
4. DATA ECOSYSTEMS FOR THE SDGS	8
5. CHALLENGES FACED	11
6. CHALLENGES TO THE FUTURE	14
7. FINDINGS AND LESSONS LEARNED	16
8. RECOMMENDATIONS FOR THE DEVELOPMENT OF A DATA ECOSYSTEM	20
9. CONCLUSIONS	24
10. BIBLIOGRAPHIC REFERENCES	25

1. INTRODUCTION

The fourth technological revolution has changed how data are consumed, produced, and transferred between individuals, organizations, and countries. Lower costs of technologies that predict or forecast behaviors and events have led data to become a valued and tradable good. The place where these production, consumption, and transfer processes are generated is known as a data ecosystem. Even though there is no standard definition, *data ecosystems* can be said to include innovation, development, and growth factors connected to their level of maturity.

The 2030 Agenda for Sustainable Development poses a series of challenges regarding data needs, which means that ecosystems must be adapted and developed. This is also a historic opportunity to take the often delayed quality leap to have more and better data in countries. The United Nations World Data Forum (UN-WDF) and the Cape Town Global Action Plan (CT-GAP), launched in 2017, are two global data agenda milestones that account for these new challenges. **In 2024, the region will host the UN-WDF in Medellín, Colombia, which will be an ideal opportunity to shed light on this agenda and build a path between regional actors and more sustainable, mature data ecosystems.** On the road to Medellín, regional statistical forums, such as the ECLAC Statistical Conference of the Americas, and data community initiatives, such as Abrelatam Condatos, become relevant environments for this discussion.

Cepei, in collaboration with the Global Partnership for Sustainable Development Data (GPSDD), conducted this study on the state of data ecosystems in Latin America (LA) to provide recommendations for the progress, needs, and future lines of action, to build data capacity in the region. This work aims to obtain a general characterization of data ecosystems for the Sustainable Development Goals (SDGs) in LA countries and identify general data trends and priorities, which are not necessarily part of the traditional statistical agenda. As well as determine where international cooperation and financing efforts should be channeled to meet the demands of the 2030 Agenda. To this end, representatives from national statistical offices (NSOs), international organizations, and civil society organizations were interviewed¹ to gain a deep understanding of how their data ecosystems work, identify the progress made in the last few years, their challenges and opportunities, and their priorities for the future. Lastly, we provide a set of recommendations on the possible lines of action to strengthen the region's data ecosystems, where efforts will focus on four issues:

1. Cepei and the GPSDD thank the consulted representatives of the regional data community in partner institutions, international organizations, and statistical offices for their time, readiness, and the information provided to this study. Additionally, we appreciate the feedback provided by Julio A. Santaella during the peer review of this document.

- Modernizing institutional statistical frameworks
- Improving the mechanisms for inter-institutional coordination and articulation
- Incorporating non-traditional data sources into the production of official statistics
- Innovation in governance and data stewardship frameworks

Key messages

These key messages are a call to action for governments, international cooperation bodies, and actors in the data ecosystems:

- Data ecosystems in the region are mainly disjointed, and the actions necessary to improve their operation are still to be positioned within the priorities of key actors.
- Data ecosystems for the SDGs tend to overlap national statistical systems, with a central role of NSOs that requires further strengthening through modern legal frameworks and resource allocation.
- The priorities of the region's national statistical offices are broadly focused on the traditional stages of capacity-building, such as strengthening the physical, technological, and statistical infrastructure. Innovation within data ecosystems will continue to lag if more urgent matters are not attended to decisively.
- Developing a data ecosystem tends to be a mission sought after by data producers but delayed in the stages of higher stability and resource availability. Thus, conditions that enable innovation must be created.
- Using non-traditional data sources and including data from the private sector is still in the exploratory stages. Sustainability in these initiatives faces challenges that incentives and the appropriate legislation should resolve.
- Intensive use of administrative records and geospatial data is among the strategic priorities of data producers and is the more feasible innovation framework for more mature ecosystems.
- The comprehensive modernization of legal frameworks understood as statistical legislation, and the frameworks for data governance, transfer of data held by the private sector, privacy, and open data are central to consolidating ecosystems that ensure the quality and sustainability of data over time.
- Gaps in data for the Sustainable Development Goals are an opportunity to bridge structural gaps in capacity, but this requires higher visibility of the 2030 Agenda within states and positioning the Agenda in the private sector and civil society.
- Actors are dissimilar among countries, given the maturity condition of the ecosystem and the most urgent data gaps. Countries must progress in identifying specific actors that could eventually become agents for change in data ecosystems.
- There is a high-impact framework for action on international cooperation in building solid foundations that could improve the articulation and development of data ecosystems and fully profit from the data revolution.

2. BACKGROUND

The United Nations 2030 Agenda for Sustainable Development sets forth a transforming vision of the economic, social, and environmental sustainability of subscribing states and is the reference guide for the work of the international community until 2030. Within this framework, data play a central role in following up its 17 Goals and 169 targets. The 2030 Agenda explicitly requests that data capacity be strengthened to support national SDG implementation plans.

National statistical offices are fundamental actors in data ecosystems for the Sustainable Development Goals. Throughout the region's countries, a large portion of the indicators for monitoring the 2030 Agenda is produced by NSOs or based on statistical operations in statistical offices. However, their role goes beyond production. National statistical offices are currently the leading articulators of these ecosystems and the main channel for incorporating methodological innovations and using new data sources.

Given the importance of NSOs in modernizing and strengthening SDG data ecosystems, within the framework of the first United Nations World Data Forum (UN-WDF) in 2017, the Cape Town Global Action Plan for Sustainable Development Data (CT-GAP) was launched.² This provides a roadmap for developing statistical capacity in countries and represents a call for governments and the international community to mobilize resources and strengthen partnerships to improve statistical activities and programmes, innovate national statistical systems, and perfect the dissemination and use of data. Additionally, at the regional level, ECLAC has launched a series of mechanisms and working groups to strengthen and improve statistical capacity in the institutions involved in measuring the Sustainable Development Goals.³

Civil society contributes to capacity-building throughout the region with new methods and tools that enable progress in timeliness, coverage, and quality of data for the Sustainable Development Goals. Cepei and the GPSDD have identified these needs and carried out, within the framework of the Data for Now initiative,⁴ projects to measure poverty using non-traditional sources in Colombia and creating a water information system in Paraguay.

Additionally, the GPSDD Data Value Project⁵ provides a general discussion framework on the principles that should sustain the future of data for development. Especially concerning how power is distributed in data production, exchange, and use, and how data use and governance can either challenge or fuel existing power imbalances.

The data revolution plays an increasingly relevant role in generating information, leading to the need to develop big data sources, methodologies, and applications, mainly driven by the capacity to connect and exchange

2. See unstats.un.org/sdgs/hlg/cape-town-global-action-plan

3. See www.cepal.org/es/organos-subsidiarios/conferencia-estadistica-americas-grupo-coordinacion-estadistica-la-agenda-2030-america-latina-caribe

4. See www.data4sdgs.org/initiatives/data-now

5. See www.data4sdgs.org/datavaluesproject

data through communication networks and the increasingly higher possibility of digitally processing information. This context challenges improving data governance and quality through greater awareness for primary data source holders. Here, national statistical offices play an essential role. This trend lifted off during the COVID-19 pandemic when data producers faced sudden interruptions in their traditional collection sources and were forced to develop new methodologies to use alternative sources.

It is evident that data supply has increased and that the world has made significant progress toward strengthening official information systems. Any potential producer of this information must inevitably face demands for greater relevance (statistics on new phenomena), timeliness (real-time statistics), and coverage, including granularity (new data sources allowing a higher level of detail), as well as reductions in costs and response burden. The use of non-traditional sources, such as data from financial transactions, mobile networks, sensors, or internet platforms, would substantially increase statistical quality in these dimensions.

3. METHODOLOGY

The research questions on which the study was structured were: What is the state of the region's data ecosystems?; What positive experiences can be transmitted? What are the most effective actions to strengthen these ecosystems?

To this end, a literature review on the functioning of LA and other regions' data ecosystems was carried out, especially the conditions that have enabled advanced maturity stages in other countries. This was complemented by interviews and consultations, under the Chatham House Rule,⁶ to representatives from national statistical offices in the region, international organizations, and civil society organizations. Anonymizing the testimonies helped avoid institutional stances and identify weaknesses and challenges during discussions.

Interviews were based on trigger questions and were systematized anonymously throughout the present document, especially in the findings section. The selection of actors and countries was based on their regional relevance and representativeness to cover various levels of capacity, socio-economic characteristics, and different political and territory organizations that account for different enabling environments for data production.

6. The Chatham House Rule is used across the globe to encourage inclusive and open dialogue in meetings. Participants are free to use the information received, but the identities or affiliations of speakers and other participants must not be revealed (See www.chathamhouse.org).

4. DATA ECOSYSTEMS FOR THE SDGS

What do we mean by data ecosystem?

Although there is no standard definition, the place where data production, consumption, and transfer processes are generated is known as a data ecosystem. Data ecosystems have emerged in the business and information technology sectors as collaboration networks for stakeholders sharing and extracting value from integrating multiple data sources. This term has recently become popular to identify any integrated data grouping that satisfies the information needs of a group of stakeholders. Thus, users play a crucial role in generating demand and maintaining data production and publication. Actors such as the private sector, the academic community, civil society, and the media are essential to underpin development, demand new resources, and build trust in data producers. Still, they are also part of the ecosystem as primary or secondary data producers.

National statistical offices are used to function as structured, coherent data islands in a world of raw, unstructured data (Menghinello et al., 2022). This feature can also be attributed to any organization in the public or non-government sector producing indicators continuously, even though production is based on a specific set of data pertaining to the mission or sectoral interest of the entity in these cases question. However, the mission of an NSO is centered around producing diverse data to be consumed by external users and managing quality in statistical processes throughout the national statistical system (NSS).

Thus, data ecosystems for the Sustainable Development Goals should be understood from a multidimensional viewpoint, given the nature of their components. In this sense, the concept of data ecosystem and national statistical systems share common features regarding processing primary, unstructured data to generate information and the adoption of methodologies that ensure their quality and meet users' needs. However, it is clear that a data ecosystem transcends a country's official statistics and requires full articulation between data, statistics, and informing policymakers.

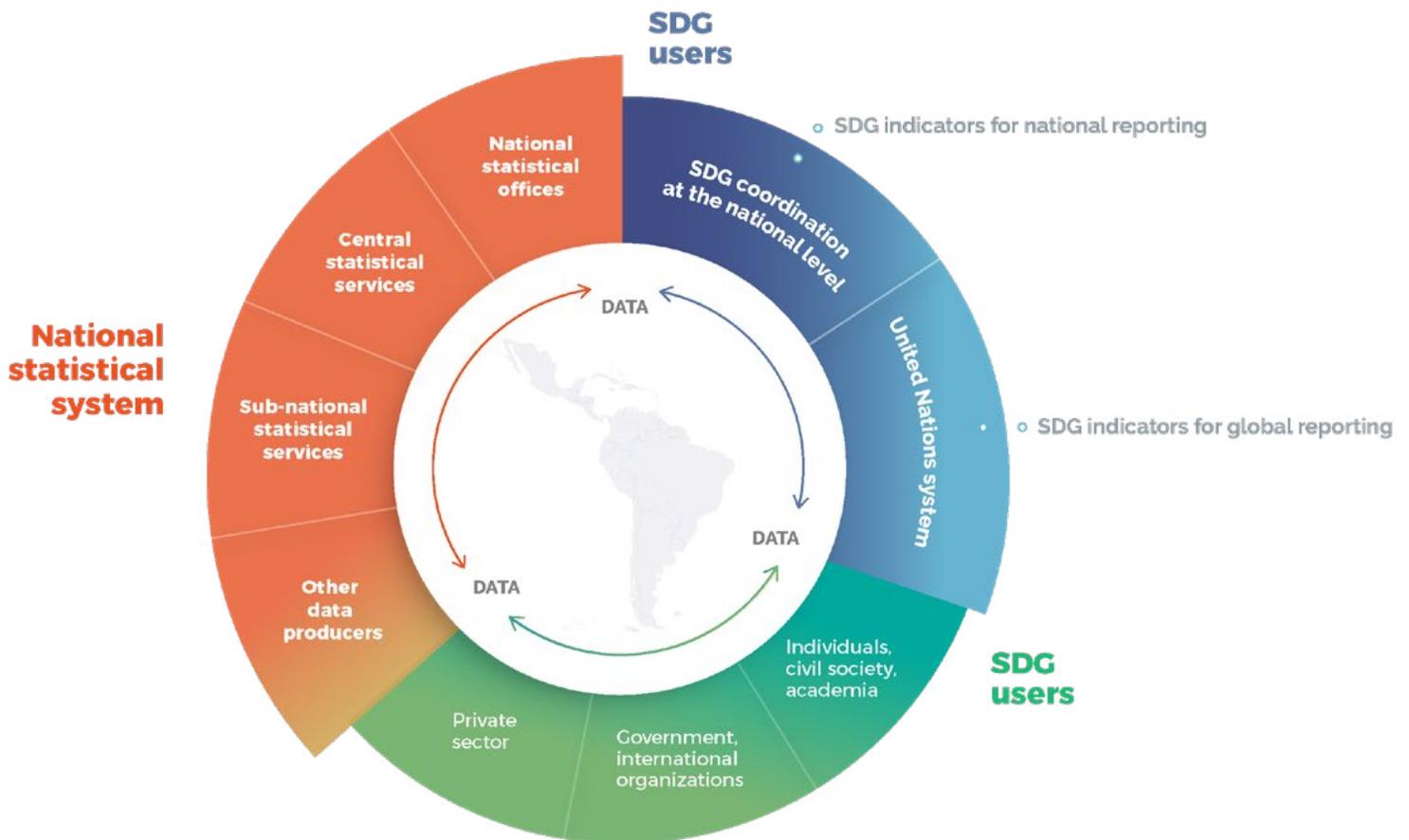
Generating reliable information within data ecosystems requires an enabling environment, understood as all resources, processes, systems, laws, and information on the entities participating in operations to generate information within the system. Thus, it is necessary to have leadership over quality management and a periodic assessment of the ecosystem's state and identify the areas that need development.

Given the features and mandates generally stipulated within states, national statistical offices are the ultimate actor to play this role. Nonetheless, it is essential **to distinguish NSO's function of producing indicators to monitor the 2030 Agenda and governments' general role in producing data for non-statistical purposes and targeted policies.** The first is mainly of interest for SDG data ecosystems.

The features of data ecosystems in the region

In the LA region, data ecosystems for the Sustainable Development Goals can be classified based on the quantity and diversity of components into large, medium, and small. Sizes vary from 25 to 70 organizations collecting or intensely using statistical and geospatial data, inter-related, as shown in Graph 1. In terms of data production, **data ecosystems for the SDGs in the region largely coincide with the national statistical systems**, with a tendency towards expanding to other information sources that need further consolidating. Thus, issues faced by NSSs are common to data ecosystems. NSOs have a central, leading role in the direct and indirect production of a good portion of indicators, regulatory coordination, and quality management.

Figure 1. National Statistical System



Source: Cepei, 2022

Why do some data ecosystems progress more than others?

As data ecosystems evolve and become increasingly complex and inter-related, dialogue and partnerships between stakeholders and the stewardship function of national statistical offices receive more attention at the global level. In LA, however, schemes adopted by countries for SDGs political governance have been defined in the configuration and development of their data ecosystems. In many cases, NSOs have been assigned a secondary role, which may cause their lack of ownership of the 2030 Agenda.

The relative initial indifference of the region's national statistical offices for the SDGs had medium-term consequences evident in the lack of articulation of their data ecosystems to generate synergies that allow them to have more and better data to monitor the 2030 Agenda. A comment received during the interviews was that "NSOs are more involved in the SDG agenda from an international relations perspective than in the coordination of a data infrastructure from a technical standpoint." However, the data ecosystems of some countries, such as Colombia, have made defining progress in this regard (See Box 1).

Box 1

Colombia: the NSS 2.0 and the National Statistical Infrastructure Plan

During the past few years, a virtuous trident made up of the National Planning Department (DNP), the Ministry of Information and Communications Technology (MinTIC), and the National Administrative Department of Statistics (DANE) has been working on the formal definition of the data ecosystem and guidelines for coordination between different actors. Although challenges remain in terms of inter-institutional articulation, there is a battery of rules and plans to ensure data quality, openness, and use, especially in the NSS 2.0 and the National Data Infrastructure Plan. The creation of the NSS 2.0 includes a Technical Advisory Council of the national statistical system (CASEN) to advise and evaluate the development of the NSS and statistical production at the national level by providing recommendations and concrete proposals for statistical methodologies and technological tools aimed at improving the production of official statistics..

Additionally, the CASEN relates to the Administrative Data Committee (CAD), in charge of promoting agreements to access data in safe, responsible, and ethical environments for members of the NSS 2.0. This articulation structure has allowed DANE to progress in managing a system of information and save time in discussing and identifying information gaps throughout the country. In addition, the MinTIC, the DNP, and the Administrative Department of the Presidency of the Republic (DAPRE) launched, in late 2021, the National Data Infrastructure Plan as a national strategy to make progress, define, and

implement the Colombian data infrastructure. The Plan is based on seven principles: 1) data quality; 2) easy retrieval, accessibility, interoperability, and reutilization; 3) data security and protection; 4) privacy by design and by default; 5) public trust and ethical management of data; 6) standardization and interoperability; and 7) strategic sectorization. It also includes articulation with the National Space Data Infrastructure (INDE), promoted by the Colombian Geographical Institute.

The lack of ties with the 2030 Agenda also coincides with the existence of other priorities within the NSOs, on fulfilling their regular operations and on traditional longstanding issues of public management, which include elements such as:

- Low budget
- Bureaucratic and administrative limitations
- Weak institutionalality
- Political instability
- Limited human resources and retention power
- Insufficient technological capacity to harness the data revolution
- Excessive burden and resistance from respondents to providing information

The Sustainable Development Goals and, mainly, instances such as the UN World Data Forum and the Cape Town Global Action Plan emerged to promote more robust data ecosystems in countries. Nevertheless, a quick comparative analysis of data capacity throughout the region, before and after **the 2030 Agenda, concludes that the SDGs have not had the expected catalytic effect.** In some cases, there is even a reversal of the enabling environments in these ecosystems, especially following significant political and economic instability after 2015.

5. CHALLENGES FACED

Aspirational agenda due to lack of resources and priorities

Several of the consulted representatives from the national statistical offices in the region stated that they have faced criticism and demands whenever they have focused on fulfilling the basic, traditional statistical operations. Due to a lack of innovation towards new methodologies, especially a more intensive use of records and other non-traditional sources. However, an often shared opinion is that systems in general, and NSOs in particular, are not yet ready to take this qualitative leap, mainly due to the lack of financial, human, and technological resources and conducive institutional frameworks.

Difficulties to access new data sources

For example, countries such as Peru and Argentina have specific agencies to integrate all records. However, fiscal secrecy and legislation on protecting individual data have not been supplemented by others that allow their use for statistical purposes. Neither has a mediator role enabled NSOs to facilitate the inter-operability of databases. This situation is repeated throughout many of the region’s countries with no specific regulation to allow the lifting of secrecy for statistical purposes. This means they must access them through bilateral agreements with the producing entities in a limited, interrupted manner, sustained only by the parties’ willingness

Obsolete and dispersed institutional frameworks

The lack of modernization of institutional frameworks for the governance of statistical systems enables an understanding of limitations in developing LA data ecosystems. The obsolescence of statistical laws is a particular feature in this region. A recent study from the World Bank⁷ positions Latin America as the region with the most outdated statistical laws: 90 % of countries have laws older than ten years, a percentage much higher than the rest of the regions. However, some countries, such as Costa Rica and Paraguay, have achieved regulatory improvements in the last few years (See Box 2).

Graph 2. Statistical laws in Latin America



Source: Cepei, 2022

7. See: documents1.worldbank.org/curated/en/826351643712794722/pdf/Survey-on-the-Implementation-of-the-Cape-Town-Global-Action-Plan-for-Sustainable-Development-Data.pdf

Box 2***Paraguay: The missing NSO***

Until 2020, Paraguay had the oldest statistical law in the region (1942) and was one of the only two countries, together with El Salvador, without a statistical office (it was a general directorate under the organization structure of the Paraguayan state). Although the functioning of the new law has yet to be regulated, Paraguay has been gaining traction through innovative initiatives, setting a clear example of how much progress can be made when decisions are based on a sound foundation. With its new statistical law passed, Paraguay achieved a modern legal framework and the creation of its National Statistical Institute. They have also conducted decisive authority reinforcement in coordinating their data ecosystem. A tool that contributed to this process is the National Statistical Development Strategy, which expands the spheres and uses of statistics, facilitates collaboration between entities in the public sector, creates a Statistical Innovation Lab, and builds a Geospatial Data Infrastructure (IDE).

Other countries, such as Trinidad and Tobago, have enacted legislation that governs various aspects of the data ecosystem. These pieces of legislation cover most of the critical legal elements required for a modern and functional data ecosystem, including protection of human rights, privacy, personal information, intellectual property, private property, financial information, freedom of information, data dissemination, statistical data collection, e-commerce, and computer misuse (UNDP, 2017).

The risk of having an overly dispersed legal body is the eventual loss of coherence and confusion between laws, with no clear leadership of the data ecosystem and possible overlaps in functions. The statistical law should contain these prescriptions and articulate with laws representing essential agendas given their specificity and relevance.

Political instability

The political instability in which data producers are immersed, with no legal framework providing a mandate, is an interruption factor in data ecosystems' development and articulation processes. Peru and Mexico are examples of how existing established mandates for the leadership of a statistical office can enable policy continuity and maturity. Conversely, countries such as Ecuador and Brazil have been subjected to frequent changes in the administration of their national statistical offices in the last five-year period, which causes interruptions in the maturity processes of their working programs and data ecosystems. The interviews confirm that innovations in the use of non-traditional sources and public-private associations have followed the fluctuations of the NSOs, which explains the tendency to keep innovation projects in the exploratory stages.

6. CHALLENGES TO THE FUTURE

Changes in government

Changes in government in the region raises questions about the development of data ecosystems. Considering that opposing parties have won the last fourteen presidential elections in LA, this trend predicts an increasing instability in the leadership of NSOs and the main data-producing public bodies, which adds to the difficulty of retaining human capital and technical capabilities. Thus, it is hard to envision a different path from the one observed.

Graph 3. Electoral landscape in the region: Changes in government for an eventual change in NSO leadership



Source: Cepei, 2022

In the region’s current context, **improvements in data quality are based on the personal leadership of producing institutions.** Instead of sound legal and institutional frameworks establishing stable, adequate principles, procedures, and plans that guarantee proper operation of national statistical offices and efficient coordination of statistical activities throughout the national statistical system.

Box 3***Future OECD countries: Good practices to set a course***

The Organisation for Economic Co-operation and Development (OECD) has successfully leveraged reforms to national statistical systems at the global level, through their recommendation on good practice, which countries seeking accession must implement. In the region, Mexico, Chile, Colombia, and Costa Rica are currently members and the OECD imprint is evidenced in sound statistical systems. Among the current OECD requirements, we may note having a modern statistical law that includes the 12 good practice recommendations on how to improve coordination of data systems and the specific strengthening of the national statistical office's professional independence. Within this framework, Costa Rica began its legislation modernization in 2019. Peru and Brazil have started their OECD accession processes, and Argentina is awaiting confirmation. It is probable that these countries will aim their actions in the near future towards a path involving new practices in the progress of the region's data ecosystems.

Disinformation and effective communication

Disinformation and fake news treatment also involve a significant challenge in consolidating the interaction of the data ecosystem. Avoiding the bad press that can affect the transfer and integration of databases from different sources is fundamental. When there is any leak in any entity of the public or private sector, suspicions of violations of personal data rapidly spread to damage the reputation of the whole system, which affects collaboration initiatives. An example is the recent data leak in Argentina's National Registry of Persons, which indirectly impacted the population census, especially on the question about the national identity document number, indispensable to progress towards a system based on administrative records. Generating trust and effective communication are fundamental tools to consolidate ecosystems.

Data stewardship

Another challenge for data ecosystems is related to the data steward role. Data stewardship aims to make data available to all users openly and efficiently (Saura et al., 2022). With this purpose in mind, national statistical offices must play a relevant role in data governance structures, which they seem not to be playing. Thus, it is essential to distinguish between NSOs' roles as data stewards in producing official statistics and the role of government data stewards for non-statistical purposes. It is unclear how different functions are assigned within governments, and regulation in force often creates overlapping and conflicting positions.

Access to new data sources raises governance issues regarding quality, accuracy, ownership, and reputation (Krizman & Tissot, 2022). A relevant aspect is the alignment of alternative information sources and data producers outside the national statistical system with the Fundamental Principles of Official Statistics and other quality protocols followed by NSOs. Concern runs on two tracks: What incentives can be offered to the private sector to align the quality of their data with official statistics? And, if data from alternative sources are used, how could this affect the reputation of official statistics due to evident issues on the sustainability and quality of information, data leaks, or company reputation events unrelated to data production?

7. FINDINGS AND LESSONS LEARNED

The main findings to strengthen data ecosystems in the region stem from interviews with diverse representatives are summarized as follows:

- **The general trend, with some exceptions such as Colombia and Mexico, is “de facto” ecosystems in early articulation and regulatory stages.** This means that, in general, inter-relation schemes are voluntary and volatile. Likewise, data production and use are articulated, non-systematic, and spontaneous. During the pandemic, information needs were solved by applying indirect or ad hoc methods. This means that when the ecosystem requires data, it creates them somehow. In this sense, data gaps catalyze more articulation and collaboration within the ecosystem.
- **Actors are dissimilar among countries, depending on the maturity condition of the ecosystem and the most urgent data gaps.** In some cases, actors such as central banks maintain vital statistical responsibilities and have more significant resources than national statistical offices. In other instances, geographical institutes may be actors for change, closing gaps in environmental aspects. Planning and ICT offices appear as essential partners in leveraging data policy. Civil society organizations and the private sector do not occupy a significant place, although there have been successful exploratory experiences.
- **The 2030 Agenda requires greater prominence among the priorities of the region’s data ecosystem actors.** A possible hypothesis is related to the relegated role of data producers in SDG governance schemes. Many countries have assigned the implementation of the Sustainable Development Goals to a coordinating office with the global mandate of the 2030 Agenda, included in the data item. NSOs are reduced to a technical counseling function for national adaptation of indicators. Data policy and bridging gaps for SDG reporting are being done by an executive office that does not produce—or even intensively use— data and is closer to policymaking and plan design.

- **The priorities of the region's NSOs are focused on traditional stages of capacity-building.** National statistical offices are in a critical situation regarding innovation and modernization in their production since they are often unable to keep the commitments made due to a lack of financial and human resources and a lack of a clear legal mandate to empower them politically. Current efforts are centered on meeting the budgetary and political commitments needed for the population census round and other large-scale statistical operations. Both census programs and the strengthening of traditional surveys are among the funding and demand priorities of NSOs to entities such as the Inter-American Development Bank. Thus, articulating the data ecosystem seems to be a widely sought-after mission but is delayed by stages to achieve greater stability and development.
- **Data regulation is insufficient and obsolete,** especially statistical laws, which result in weak institutional frameworks for data use and production and unstable development processes for NSOs and data ecosystems.

Box 4

The elusive goal of a statistical law

The situations of the national data ecosystems of countries such as the **Dominican Republic, Argentina, Colombia, and Ecuador** differ in many aspects but share a similarity. They have made several unsuccessful attempts to modernize their statistical laws and overcome existing weaknesses to develop their data ecosystems.

The **Dominican Republic** is currently discussing its statistical law, which has not worked in the past, mainly due to internal disagreements between the NSO and the Central Bank (Dargent et al., 2018). In **Argentina**, the executive power presented a bill in 2019 that lost momentum due to internal disputes between the National Institute of Statistics and Censuses of the Republic of Argentina (INDEC) and the Ministry of Economy, delaying the legislative debate by two years. It was also interrupted by a change in government and the COVID-19 pandemic (Dupont & Muñoz, 2021). In the case of **Colombia**, the statistical legal framework is scattered among various laws with different objectives. In addition, in 2014, due to the possibility of joining the OECD, the statistical law was heavily promoted, although efforts were later dampened due to internal struggles. In 2022, the project was resumed and is currently being discussed in Congress. These cases are not alien to Ecuador, whose statistical system has gone through many crises in the last few years. Recently, through the impetus of the new government, the National Institute of Statistics and Censuses of **Ecuador** (INEC) has worked on a provisional adaptation of the regulatory framework for a modern statistical law to provide INEC with stability in its leadership and planning, although this goal has proven elusive.

Data capacity building in these countries has been marked by personal leadership, with successful results. Even so, individual efforts are not sustainable over time without the “quality leap” of a sound institutional framework.

- **Using non-traditional data sources and partnerships with the private sector are still in exploratory stages.** There is stagnation in the region regarding discontinued or not sustainable projects over time. The lack of incentives for the private sector is directly related to this fact. For example, some national statistical offices have developed partnerships to share data with telephone companies in their countries. Only some have effectively participated, resulting in partial data for restricted periods.
- **Concerning harnessing big data, three limitations operate as barriers:**
 1. Trust in the quality and sustainability of data from these types of sources.
 2. Inadequate technological infrastructure to process large data flows.
 3. Difficulties in recruiting qualified human resources that could regularly advance these operations. This last point is essential to understanding the limitations of the region’s data ecosystems, given that the number of people available that support statistical or data science issues are limited in the public sector, with varying levels of education, training, and experience. Moreover, retaining qualified resources in the public sector is difficult due to the intense competition presented by private sector remuneration and limitations in career plans, qualifications, and other incentives. Even within NSOs, a limited number of analysts support the processes expected for a functional data ecosystem.
- **NSOs find the intensive use of administrative records relevant and include the progress towards more extensive systematization and access to these records within their priorities.** Also noted was the potential offered by spatial data in the production and dissemination of indicators and strengthening the statistical infrastructure to supplement experiences such as address registries and the delimitation of parcels for agricultural censuses. In this regard, the use of administrative records and the capacity to share and process information within the public sector is one of the main challenges in the following years.
- **Even though the region’s NSOs claim to understand the function of a data steward, they also find that many of the mentioned weaknesses could become impeding barriers.** Although national statistical offices can create value by enabling data reuse, identifying collaboration opportunities, and promoting laws and standards, a proactive and effective response requires mandates and resources that are not currently within reach of most regional NSOs.

- **SDG data gaps are an opportunity to solve structural issues and consolidate ecosystems.** For example, the available data accounting for the environmental dimension, including water information systems, and other phenomena such as migration, natural disasters, and health shocks such as COVID-19 are the main drivers to enhance the capacity of integrating non-traditional data sources.
- **The dynamics surrounding data at the sub-national level are a priority, given that they present more outstanding weaknesses regarding technical capacity, resources, and the intensive use of data for public policies.** Brazil and Mexico are two case studies in this regard: Federal governments with highly centralized data policies and engaging governance experiences, paired with a sound operating structure in their statistical offices. They represent exceptional cases, even at the global level, since they include, in a single agency, both the statistical operations and the production of geographical data: Brazilian Institute of Geography and Statistics (IBGE), National Institute of Statistics and Geography (INEGI, Mexico) (See Box 5.). This entails many advantages in terms of coordinating geostatistical data.

Box 5

IBGE and INEGI: Different scales, replicable lessons

IBGE and INEGI are two of the largest NSOs in the world, with over 10,000 employees. Given the vast territorial spectrum they must cover, added to the centralization of competencies in geo-statistical and territory terms, they account for two advanced data ecosystems concerning their foundational pillars. However, their experiences are not necessarily replicable in the rest of the region. The characteristics of both countries are unique in terms of population, territory size, and the nature of their data ecosystems, strongly characterized by their NSOs. However, other experiences could be transferred to the rest of the region: i) the Mexican regulation concerning the division of thematic domains into data sub-systems (See Box 1.), and ii) the use of human resources qualified in statistical issues and of government data in Brazil.

i) Sub-systems of information in Mexico

The Mexican National Statistical and Geographical System have national information sub-systems: the National Demographic and Social Sub-system (SNIDS), the National Sub-system of Economic Information (SNIE), the National Sub-system of Geographical Information, Environment, Territory and Urban Planning (SNIGMAOTU), and the National Sub-system of Government, Public Safety and Provision of Justice Information (SNIGSPIJ). Each sub-system has an executive and special technical committee that groups the corresponding agencies according to the theme, gathering together the state units interested in each topic. They must have an information infrastructure, key indicators, and sources to obtain information for the indicators. This experience provides articulation between stakeholders, the possibility

of early detection and bridging gaps with different types of sources, and the potential for innovation in developing any data ecosystem.

ii) The IBGE school of government statisticians

Brazil has a National Statistical Sciences School (ENCE), a federal higher-education institution part of IBGE in Rio de Janeiro, Brazil. The ENCE was founded in 1953 and is responsible for training IBGE officials. It offers undergraduate and specialization courses and develops research activities. This experience, due to its trajectory and usefulness, represents a good practice example for the region regarding education and retaining qualified human resources, one of the significant weaknesses among the region's NSOs.

8. RECOMMENDATIONS FOR THE DEVELOPMENT OF A DATA ECOSYSTEM

Following the findings on the situation of data ecosystems for the Sustainable Development Goals in Latin America, the following are some recommendations that might help guide efforts towards their strengthening.

1) *Modernizing institutional statistical frameworks*

- The need to strengthen institutional frameworks, understood mainly as statistical laws, that also contemplate a comprehensive data infrastructure, privacy protection, and opening and transferring data to the private sector, among others, is a recurring barrier to consolidating data ecosystems beyond national statistical systems. This call to action aims to generate the necessary conditions for developing data ecosystems on solid grounds.
- The obsolescence of statistical laws is a particular feature in the region that must be addressed—especially on the professional independence of data producers, coordination within the system, confidentiality, the mandate and responsibilities for data collection, unrestricted access to administrative records for statistical purposes, and statistical quality and data dissemination frameworks, among others. Moreover, producing a regulatory framework for re-utilizing data held by the private sector is a fundamental aspect of the future of official statistics.
- Data capacity based on proactive leadership is sustainable as long as it is built on solid pillars that translate into state policies. One of the main challenges in the region is consolidating regulatory progress for access and privacy of personal data and rules that enable incentives in public-private initiatives, as well as the definition of the roles taken on by each actor within national data ecosystems.
- Another critical issue is preventing rules, strategies, and plans from being only on paper and not adopted in practice. A related recommendation is to design viable, realistic frameworks adapted to national needs so that the implementation can be paired with the regulation.

- There is also a need to strengthen the availability of physical, human, and financial resources for actors in the region's data ecosystems. And mainly to harness the capacities acquired by the private sector in this regard. Resources should be provided at the national level in a predictable, programmatic manner and, whenever possible, established by law. This will allow for coordinated planning surrounding the support required by the data ecosystem and will avoid individual competition for resources between ministries and organizations.

2) Improving the mechanisms and instances for inter-institutional coordination and articulation

- The 2030 Agenda can play an essential role in improving the coordination and articulation of data ecosystems. However, this requires higher visibility of the Agenda within states, the private sector, and civil society.
- Data ecosystems in the region must achieve greater articulation and leadership from national statistical offices. A first step should consider the experience, regulation, and lessons in coordinating the national statistical systems. The most effective coordination tools of NSSs can become pioneer elements in articulating and expanding data ecosystems. Regarding this issue, quality certifications have proven to be exceptionally effective. Statistical yearbooks were the ideal coordination tool for national statistical systems in the past. Whatever was left out of the yearbook was understood as not meeting the quality standards to be considered official statistics. This means there were implicit rules on periodicity, timeliness, relevance, and punctuality in the production of information. Unifying portals for official statistics may play a vital role in this regard, as well as producing inventories of statistical operations.
- Another good practice that could improve cooperation between different data providers is statistical advisory committees, mainly aimed at strengthening communication and relations between various stakeholders and coordinating between parties to reduce respondent burden and underpin the importance and visibility of data. Planning and development ministries and modernization/ICT ministries could play a fundamental role in leveraging coordination initiatives.
- We suggest the identification of critical actors in each country that could become agents for change within data ecosystems and help guide actions to underpin their role.

3) Incorporating non-traditional data sources into the production of official statistics

- "Traditional" statistical surveys and censuses could be supplemented with new sources of information, either private business datasets or public registries that were not originally created for statistical purposes (non-traditional data sources). Official surveys and censuses are irreplaceable points of reference that will allow the assessment of whether new big data sources are precise, correct, and representative of the situation under analysis. Contrasting exercises between data from non-traditional sources

and data from traditional sources are a great first step.⁸ The exercise mentioned above is an opportunity for the region, given the workload of NSSs. They deny having sufficient time or resources to explore alternative sources without compromising the core of their conventional operations.

- Efforts towards using non-traditional data should be focused on readily available data, such as satellite images and price sweep or scanning data, which can replace costly operations in agriculture and price statistics. This demonstrates the need to reinforce the basic statistical production programs before investing in new methods with uncertain coverage and quality standards.
- Another recommendation for action aimed at incorporating non-traditional sources in the production of official statistics pertains to taking concrete measures to progress regulatory frameworks towards schemes that establish mutual supporting relations between the public and private sectors. For example, there are incentive strategies for data transfers, offering compensation in terms of knowledge transfer to national statistical offices in the form of methodologies, processes, improvements in classifications, integration, and data quality assessment, and even anonymization and integration of different databases.
- It is helpful to distinguish between partnerships that seek to identify the potential of a data source for future statistics and associations for regular statistics. Regular statistics require sustainable collaboration. Voluntary collaboration must be backed by enforceable responsibilities for companies and allow the reutilization of private data. In this sense, it is necessary to adapt the regulatory framework behind this practice beyond voluntary agreements.
- The region's countries should lean on other countries' ongoing experiences. For example, Eurostat (European Statistical System - ESS) has initialized a Group on using privately-held data for official statistics, which has identified ten principles for using privately-held data for official statistics.

4) *Innovation in governance and data stewardship frameworks*

- Although in the future, national statistical systems may face a decline in their traditional “data collector” functions, they need to play an essential role as data quality stewards. Data stewards must pursue the systematic, sustainable, and responsible use of data through trans-sectoral collaboration; audit data and ensure their ethical use; disseminate data culture in society, and promote information rules. They must set a minimum requirement for all information released from public spaces and infrastructure to combat disinformation, fake and incorrect information, rumors, and algorithmic bias.
- The institutional and data-quality frameworks traditionally adopted by NSOs in the production of official statistics must evolve to successfully exploit the richness of the information included in data ecosystems. In

8. For example, during the project mentioned in Section 2, within the framework of the Data for Now initiative in Colombia, poverty was measured using alternative sources, which were later contrasted with the traditional household survey method. See: unstats.un.org/capacity-development/data-for-now/data-for-now-in-Colombia

this sense, the frameworks must be holistic, i.e., include all organizations involved, as well as the corresponding principles, policies/procedures, structures, functions, and responsibilities, and be an integral part of their strategic plans.⁹ It is crucial that a clear set of requirements and safeguards for public or private data holders is coded into the legal framework.

- Data producers can still play a vital role in promoting open data initiatives and publishing a broader range of statistics providing additional value for all partners and users.

9. The Value Data Project, coordinated by the GPSDD, provides a collective view for the future of fair data with agency, responsibility, and action as the main features.

See: www.data4sdgs.org/news/introducing-the-data-values-project

9. CONCLUSIONS

This document seeks to reflect an overview of the situation of data ecosystems and shed light on critical conditions obstructing data capacity development in LA.

It follows from these findings and recommendations that there is ample room for action from interested parties who seek to promote and underpin data capacity in the region. Designing strategic actions aligned with these recommendations is the most effective way to channel efforts toward developing data ecosystems for the Sustainable Development Goals.

The limitations of national statistical offices in fulfilling their articulating role within their data ecosystems is a structural aspect primarily due to traditional issues that still affect the region's statistical offices. NSOs are captives of their operative agendas and have limitations in institutional terms and resources to make progress even in coordinating their own national statistical systems.

In this context, although the SDGs should have been a window of opportunity to strengthen NSOs, in some cases, they have become a challenge that is as hard or even harder to face. Before this, many NSOs have veered towards other priorities. Additionally, the lack of direct responsibility of NSOs over the SDG data agenda has resulted in a scheme of rather voluntary, dispersed efforts and has made them concentrate on international relations areas rather than their technical capacity.

This document is a call for action regarding redoubling efforts toward developing SDG data ecosystems in line with the Cape Town Global Action Plan. It calls on the region to strengthen institutional frameworks, implement improved coordination mechanisms, incorporate non-traditional data sources, and advance towards modern governance and data quality assurance schemes.

10. BIBLIOGRAPHIC REFERENCES

- Aaser, M., Kanagasabai, K., Roth, M., & Tavakoli, A. (2020, November 23). *Four ways to accelerate creating data ecosystems*. McKinsey & Company. www.mckinsey.com.br/business-functions/quantumblack/our-insights/four-ways-to-accelerate-the-creation-of-data-ecosystems
- Abdulla, A., Janiszewska, E., & Podlesny, J. (2021, March 8). *Data ecosystems made simple*. McKinsey & Company www.mckinsey.com/business-functions/mckinsey-digital/our-insights/tech-forward/data-ecosystems-made-simple
- Azkan, C., Möller, F., Meisel, L., & Otto, B. (2020). *Service Dominant Logic Perspective on Data Ecosystems - A Case Study based Morphology* [Conference presentation]. 28th European Conference on Information Systems, Marrakesh, Morocco.
- Cañón, V., Clavijo, A., Godoy, L., Letouzé, E., Pestre, G. y Ricard, J. (2017). *Definición de la estrategia de Big Data para el Estado Colombiano y para el desarrollo de la industria de Big Data en Colombia*. Data-Pop Alliance. datapopalliance.org/wp-content/uploads/2020/11/Documento2_BuenasPracticasyDefiniciones_DPAMIT.pdf
- Clark, X., Zaror, D., & Mejía, J. (2020). *Statistical legal frameworks in Latin America. Realities, best practices and recommendations*. IDB.
- Dargent, E., Lotta, G., Mejía, J.A., & Moncada, G. (2018). *Who wants to know? The Political Economy of Statistical Capacity in Latin America*. IDB.
- De Beer, B., & Tissot, B. (2020). *Implications of Covid-19 for official statistics: a central banking perspective*. (IFC Working Papers, n.º 20).
- Dupont, J., & Muñoz, H. (2021, October). *Rebuilding of the National Statistical System of Argentina. Some lessons learned*. (UNECE Expert Group on Modernizing Statistical Legislation Frameworks, Geneva, Working Paper n.º 5). unece.org/sites/default/files/2021-10/WP5_DupontMunoz_Rebuilding_NSS_Argentina_LessonsLearned_2021_EM.pdf
- European Statistical System (ESS). (2017). *European Statistics Code of Practice*. European Union. ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf/e7f85f07-91db-4312-8118f729c75878c7?t=1528447068000
- European Statistical System (ESS). (2020). *European Statistical System handbook for quality and metadata reports, 2020 edition*. European Union.

- European Statistical System (ESS). (2021). *European Statistical System (ESS) position paper on the future Data Act proposal*. [ESS.ec.europa.eu/eurostat/documents/13019146/13405116/main+ESS+position+paper+on+future+Data+Act+proposal.pdf/37f3b5c7-abfd-5a05-6be2-fdc4b87ee7d2?t=1631695372906](https://ec.europa.eu/eurostat/documents/13019146/13405116/main+ESS+position+paper+on+future+Data+Act+proposal.pdf/37f3b5c7-abfd-5a05-6be2-fdc4b87ee7d2?t=1631695372906)
- European Statistical System Committee. (2020). *Actions Enabling the Use of Privately Held Data for Official Statistics*. [Conference presentation]. 43rd meeting of the ESS Committee, document ESSC 2020/43/8/EN.
- International Statistical Institute (ISI). (2010). *Declaration on professional ethics*. ISI. www.isi-web.org/files/docs/declaration-on-professional-ethics_2010.pdf
- Krizman, I., & Tissot, B. (2022). *Data governance and quality: involving data sources' owners* [Conference presentation]. 10th European Conference on Quality in Statistics, Vilnius, Lithuania.
- Menghinello, S., Ceccarelli, C., & Vitaletti, S. (2022). *Data ecosystems and NSIs: toward an evolution of institutional and quality frameworks in the production of official statistics?* [Conference presentation]. 10th European Conference on Quality in Statistics, Rome, Italy.
- OECD. (2020). *Report on the Implementation of the OECD Recommendation on Good Statistical Practice*. OECD. legalinstruments.oecd.org/api/download/?uri=/public/0e2576c5-f161-4d3f-8d4a-1c37a0b2cc7e.pdf&name=Implementation-of-the-OECD-Recommendation-on-Good-Statistical-Practice.pdf
- OECD. (2019). *The path to becoming a data-driven public sector*. OECD Digital Government Studies, OECD Publishing.
- Ricciato, F., Wirthmann, A., Giannakouris, K., Reis, F., & Skaliotis, M. (2019). *Trusted Smart Statistics: How New Data Will Change Official Statistics*. *Statistical Journal of the IAOS* 35 (2019), 589-603. ec.europa.eu/eurostat/cros/system/files/sji190584.pdf
- Saura, A., Gómez, Y., & Cánovas, A. (2022). *Data stewardship: To be, or not to be, that is the question* [Conference presentation]. 10th European Conference on Quality in Statistics, Madrid, Spain.
- Schweinfest, S., & Jansen, R. (2021). *Data Science and Official Statistics: Toward a New Data Culture*. *Harvard Data Science Review*, Issue 3.4.
- The World Bank, Paris21 y UN Department of Economic and Social Affairs (UNSD). (2022). *Survey on the Implementation of the Cape Town Global Action Plan for Sustainable Development Data*, Washington DC. *International Bank for Reconstruction and Development / The World Bank*. documents1.worldbank.org/curated/en/826351643712794722/pdf/Survey-on-the-Implementation-of-the-Cape-Town-Global-Action-Plan-for-Sustainable-Development-Data.pdf

- United Nations Development Programme (UNDP). (2017). *Data ecosystems for sustainable development. An assessment of six pilot countries*. UNDP. www.undp.org/publications/data-ecosystems-sustainable-development
- United Nations Economic Commission for Europe (UNECE). (2016). *Generic Law on Official Statistics*. UNECE. unece.org/DAM/stats/publications/2016/ECECESSTAT20163_E.pdf
- United Nations Economic Commission for Europe (UNECE). (2019). *The role of national statistical systems in the new data ecosystem* [Conference presentation]. Conference on European Statisticians, Paris, France.
- United Nations Statistical Division (UNSD). (2014, January). *Fundamental Principles of Official Statistics (A/RES/68/261)*. UNSD. unstats.un.org/unsd/dnss/gp/fundprinciples.aspx
- United Nations Statistical Division (UNSD). (2019). *A review of open data practices in official statistics and their correspondence to the Fundamental Principles of Official Statistics*. UNSD. unstats.un.org/unsd/statcom/50th-session/documents/BG-Item3c-Open-Data-guidance-and-mapping-to-FPOS-E.pdf
- US National Science and Technology Council. (2022). *Protecting the Integrity of Government Science, A Report by the Scientific Integrity Fast-Track Action Committee*. US NSTC. www.whitehouse.gov/wp-content/uploads/2022/01/01-22-Protecting_the_Integrity_of_Government_Science.pdf
- Van den Homberg, M. (2018). *Characterizing Data Ecosystems to Support Official Statistics with Open Mapping Data for Reporting on Sustainable Development Goals*. *International Journal of Geo-Information*, 7(12), 456.



..... www.Cepei.org

 [@infoCEPEI](https://twitter.com/infoCEPEI)  [@infoCEPEI](https://www.facebook.com/infoCEPEI)  [Info CEPEI](https://www.youtube.com/InfoCEPEI)  [InfoCEPEI](https://www.linkedin.com/company/infoCEPEI)

In partnership with



**Global
Partnership**
for Sustainable
Development Data