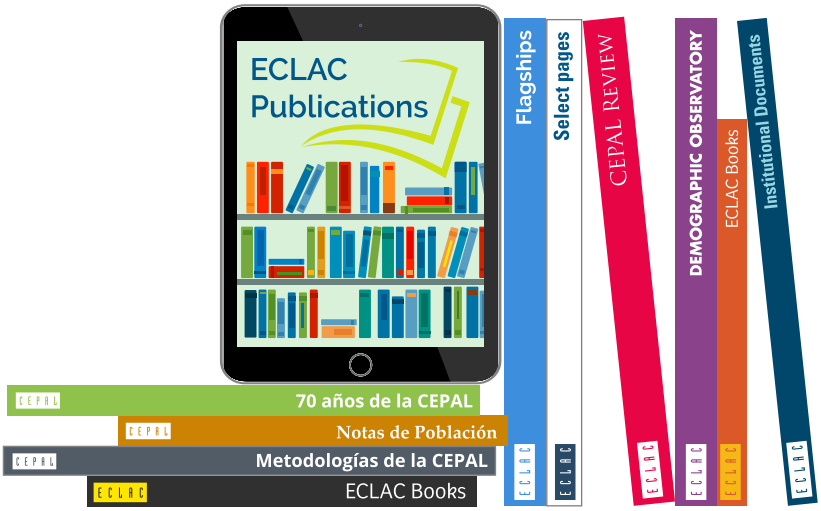
Veronika Vilgis
Valeria Jordán
Alejandro Patiño

Measuring the Internet economy in Latin America

The cases of Brazil, Chile, Colombia and Mexico

UNITED NATIONS
ECLAC
75 years
Working for a productive, inclusive and sustainable future
BigDATA
Digital economy for Latin America and the Caribbean
DEVELOPMENT in transition
European Union regional facility for Latin America and the Caribbean

Thank you for your interest in this ECLAC publication



Please register if you would like to receive information on our editorial products and activities. When you register, you may specify your particular areas of interest and you will gain access to our products in other formats.

Register

www.cepal.org/en/publications

www.instagram.com/publicacionesdelacepal

www.facebook.com/publicacionesdelacepal

www.issuu.com/publicacionescepal/stacks

www.cepal.org/es/publicaciones/apps

UNITED NATIONS

ECLAC

Measuring the Internet economy in Latin America

The cases of Brazil, Chile, Colombia and Mexico

Veronika Vilgis
Valeria Jordán
Alejandro Patiño

UNITED NATIONS
ECLAC

BigDATA
Digital economy for Latin America and the Caribbean

DEVELOPMENT in transition
European Union regional facility for Latin America and the Caribbean

This report was prepared by Veronika Vilgis, a consultant at the Division of Production, Productivity and Management of the Economic Commission for Latin America and the Caribbean (ECLAC), and Valeria Jordán and Alejandro Patiño, officers at the same Division. The document also includes contributions from consultants Andrés Garrido and Sebastián Santana. The authors wish to thank Laura Póveda and Valeria Lobos of ECLAC for their assistance.

The following organizations participated in the preparation of this study: Brazilian Institute of Geography and Statistics (IBGE), Regional Centre for Studies on the Development of the Information Society of Brazil (CETIC.br), National Institute of Statistics of Chile (INE), National Administrative Department of Statistics of Colombia (DANE), National Institute of Statistics and Geography of Mexico (INEGI) and Network Information Center of Chile (NIC Chile) of the University of Chile.

This publication was produced as part of the United Nations Development Account project entitled "Big data to measure and promote the digital economy in Latin America and the Caribbean". Support was also received from the Regional Facility for Development in Transition for Latin America and the Caribbean, in the context of the strategic alliance between ECLAC and the European Union, and from the Regional Observatory for Digital Development.

United Nations publication
LC/TS.2023/51
Distribución: L

Printed at United Nations, Santiago
S.2300976[E]

This publication should be cited as: V. Vilgis, V. Jordán and A. Patiño, "Measuring the Internet economy in Latin America: the cases of Brazil, Chile, Colombia and Mexico", *Project Documents* (LC/TS.2023/51), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC), 2024.

CONTENT

Table

Figures

Map

Diagrams

INTRODUCTION

Digital technologies play a crucial role in diversifying and advancing productive sectors, as well as promoting social inclusion. In Latin America and the Caribbean, the adoption of these technologies is essential to address challenges such as low economic growth, limited investment, increasing poverty rates, and the need to bridge productivity gaps with more advanced countries. To foster innovation and technological change, it is imperative to develop evidence-based policies.

As the digitalization of economic and social activities continues to progress, the availability of data grows exponentially, presenting new opportunities for information and driving data science and innovation. This wealth of information can provide valuable insights into the dynamics of the digital economy and inform the design of policies that facilitate these transformations.

The COVID-19 pandemic has accelerated the process of digitalization, leading to a surge in teleworking, distance education, and online activities. In particular, the proliferation of online stores and e-commerce solutions has become more prominent during this period. However, traditional statistical sources have struggled to provide timely information on these rapid changes. Therefore, exploring alternative measurement methods is crucial to gain a better understanding of emerging phenomena.

This study aims to address this need by combining data extracted from the web with official statistical sources, specifically the business registers of National Statistical Offices (NSOs). The goal is to analyze and quantify online business activities, going beyond traditional industry classifications. This exercise promotes data innovation and the exploration of big data techniques and tools, providing methodological and conceptual insights that can enhance our understanding of these phenomena and serve as a foundation for future research.

Through this work, a novel classification of companies based on their internet usage has been developed, forming what is known as the Internet economy. This classification was applied to four Latin American countries —Brazil, Colombia, Chile, and Mexico— between 2019 and 2021. The first part of the paper examines the importance and challenges of measuring the digital economy, while also presenting relevant conceptual definitions. Subsequently, an analysis of business websites in the four countries is presented. The report concludes with a review of the results obtained by combining web data with business data from registers, along with methodological considerations. It also reflects the main findings of the study and the challenges encountered.

By improving our understanding of the digital economy through innovative methodologies and data analysis, we can generate valuable insights to inform policies and drive further progress in Latin America and the Caribbean.

THE IMPORTANCE OF DIGITAL TRANSFORMATION FOR SUSTAINABLE DEVELOPMENT AND ITS MEASUREMENT

I

A. Why measure the adoption of digital technologies?

The digital revolution is ushering in a new era of industry driven by digital technologies. In this era, the traditional economy overlaps and merges with its organizational, productive and governance schemes with the digital economy and its own innovations, business models, production and business organization (ECLAC, 2018; ECLAC, 2022).

This transformation is having a profound impact on societies and economies, particularly in regions like Latin America and the Caribbean that are grappling with structural gaps. The adoption of digital technologies plays a pivotal role in these countries' development, offering opportunities for enhanced well-being and more productive and competitive economies. However, the journey towards digital transformation is not a straightforward one. It requires a strategic approach encompassing various policy areas such as infrastructure development, skills promotion, regulatory frameworks, and innovation support. To effectively navigate this landscape, it is crucial to have reliable and comparable statistical information that can guide evidence-based policies and regulations.

The international statistical community, along with national statistical offices and international organizations, has made significant strides in developing standardized frameworks and benchmarks to measure the adoption of information and communication technologies (ICTs). These efforts have primarily focused on indicators related to internet connectivity for individuals, households, and businesses. However, measurement methods have struggled to keep up with the rapid evolution of digital technologies and their pervasive impact on society and the economy. As a result, there is a need to update metrics and indicators to capture the adoption of emerging digital technologies, gather information in a more detailed and timely manner, and account for the diverse range of actors, uses, and business models that have emerged in this new digital landscape.

B. The concept of the digital economy and its different dimensions

The digital economy is a complex and multifaceted concept that lacks a single, universally agreed upon definition. This is primarily because of its versatile nature and the various dimensions it encompasses. However, many organizations associate this term with the economy that supports the processes and capacity for value creation in digital technologies. It is also used to describe how these technologies are revolutionizing consumption and production patterns.

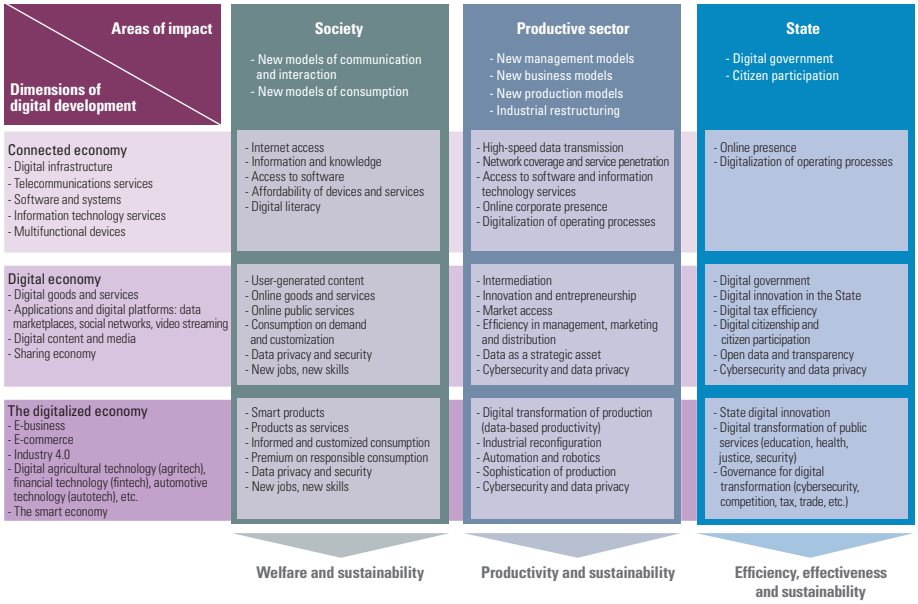
According to the European Commission (2014) the digital economy is an "economy based on digital technologies (sometimes referred to as the Internet economy)". On the other hand, in its most recent definition the Organization for Economic Cooperation and Development (OECD, 2020) defines this term as "any economic activity that depends on the use of digital inputs or whose performance is significantly improved through their use, including digital technologies, digital infrastructure, digital services and data.

Considering all producers and consumers, including the government, who use these digital inputs in their economic activities." The United Nations Conference on Trade and Development (UNCTAD, 2022) takes up the OECD concept and emphasizes the importance of having a flexible approach that allows analyzing different components of the digital economy such as products (ICT goods and services, digital content, services provided digitally), transactions (digital purchases or deliveries) and production (degree of dependence on digital inputs).

On the other hand, ECLAC (2021) proposes a three-dimensional analytical framework to study the potential of digital technologies to increase people's well-being, business productivity, state efficiency and effectiveness, and environmental sustainability. These dimensions are the connected economy, the digital economy, and the digitized economy (see diagram 1):

- Connected economy: corresponds to the digital infrastructure (broadband networks, traffic exchange points, data centers, etc.), the massification of various devices (computers, laptops, tablets and smartphones) for the use of the Internet, and the use of sensors that enable the connection not only of people, but also of machines and objects.
- Digital economy: consists of the part of economic production derived from business models based on digital technologies for the supply of goods and services. This dimension includes the models of digital platforms and intermediaries that change the value proposition in various economic sectors, through the intensive use of data, the optimization of operational processes, market segmentation and the personalization of products and services.
- Digitized economy: refers to the transformation of business and production models of companies in traditional sectors into intelligent connected systems, thanks to the adoption of advanced technologies, such as fifth-generation mobile networks (5G), Internet of Things, cloud computing, artificial intelligence, virtual and augmented reality, Big data analysis and cognitive robotics. The objective is to increase the competitiveness, productivity and sustainability of operations.

Diagram 1
Dimensions and elements of digital development



Source: Economic Commission for Latin America and the Caribbean (ECLAC), *Foreign Direct Investment in Latin America and the Caribbean*, 2021 (LC/PUB.2021/8-P), Santiago, 2021.

C. Challenges in measuring digital technologies and the importance of statistical innovation

Traditionally, the statistical measurement of the digital economy has been approached from the production (supply) or consumption (demand) side of ICTs, as well as from the deployment of telecommunications infrastructure and services and trade in ICT products. For example, on the supply side, indicators can be generated on the industries and companies that provide ICT infrastructure, goods and services, while on the demand side, the levels of adoption and use of ICT by companies and individuals can be analyzed. Part of the design of such statistics relates to the definition of the ICT sector, which mainly covers ICT manufacturing, trade and services. From this design, estimates can be made on the contribution of ICTs to growth and employment (UNCTAD, 2022).

While there is a definition of the ICT sector,[1] this dimension fails to capture all facets and perspectives of the digital economy[2] today. On the other hand, there is no agreed statistical definition on this concept, so national statistical offices at the international level use different approaches. Faced with this gap, the G20 Roadmap on a common framework for measuring the digital economy proposes a tiered classification. The first level is a "basic measurement" of the digital economy where only economic activities derived from suppliers of ICT goods and services are included. The second level incorporates a "reduced measurement" where the products of ICT goods and services are added to the economic activity derived from companies that depend on digital inputs. The third level is considered a "broad measurement", since to the first two dimensions, the economic activity of companies that significantly improve their performance through digital inputs is added. Finally, this classification also refers to the digital society to include in the analysis the rest of the effects of digital technologies on society and the economy (OECD, 2020).

The classification provided in the G20 Roadmap allows for a more precise analysis framework of the digital economy, however, the same document warns that there are still ambiguities when defining the level of dependence on digital inputs of companies (basic, reduced and / or broad), for which it suggests continuing to work on additional guidelines or criteria to help with the delimitation by levels.

In this context, this study aims to make a contribution by addressing the measurement of the digital economy from an innovative perspective that allows classifying companies according to the level of dependence on the Internet in their business model, using big data techniques to extract information from the web and its combination with other traditional statistics, and thus be able to dimension, what is called the "Internet economy". This approach allows analyzing key segments of companies that base their entire business model on the Internet, such as e-commerce platforms, professional services and applications, but also to measure the level of Internet use of other companies in the economy.

1. *Big Data*: concept and methodological challenges

Rapid technological progress driven by the digital revolution has led to an exponential increase in the ability to collect, store, process and analyze data. Activities carried out online leave a

1 (OECD, 2011) using the International Standard Industrial Classification (ISIC) Rev. 4.

2 This classification excludes companies that make heavy use of digital technologies such as platforms, e-commerce companies or digital services. There are also other definitions such as information industries that in addition to the ICT sector incorporate audiovisual content sectors.

digital trail that produces large amounts of data that have become an intangible good and that allow the development of new disruptive technologies such as artificial intelligence.

The term "Big Data" has been coined to refer to the large volume of information produced by (our) digital fingerprint. However, size alone does not define *Big Data*, but other aspects, such as the variety and speed at which data is produced and processed also play a role (Layney, 2001). Variety involves the expansion of traditional (tabular) structured data to unstructured data, such as text, images, video and audio information. Speed refers to how quickly data is produced and analyzed, often enabling real-time processing of various transactions. Since the introduction of these three dimensions or three "V", other features have also been added to the term *Big Data*,, including veracity (data quality), variability (changes in the rate at which data is generated), and value (data in its original form often has a low value relative to its volume).

However, the above descriptions only focus on what is meant by *Big Data*, and they don't consider what *Big Data*, does. For this reason, it is best to refer to *Big Data*, as new techniques of data analysis, and in this way consider the analytical methods that arise from this trend. Big data analytics requires specialized tools and programs to store and analyze large volumes of unstructured and highly variable data (Ward and Barker, 2013).

Big Data, *analytics* offers a unique opportunity to measure aspects of the digital economy that traditional measurement approaches cannot capture or cannot provide within a timely time frame to analyze the rapid changes occurring in this new paradigm. As it is a new and constantly evolving phenomenon, the monitoring of the digital economy demands not only new metrics that account for the dynamics of the different elements that compose it, but also to innovate in measurement techniques. Thus, as socio-economic activities leave more and more digital traces, analytics and data science offer new opportunities.

The challenge no longer lies in the availability of data, but in the ability to use these new tools, which depends on enabling elements such as technological infrastructure, technical skills and the ability to extract meaning from data. On the other hand, there are also certain aspects that deserve attention when using these tools. The data produced by digital activities is organic (transactional and unstructured) and is not the product of a specific statistical design (e.g. surveys), nor does it arise from random samples, which makes it difficult to establish general patterns to apply to the entire population. On the other hand, sometimes this data comes from specific services such as platforms[3] or the Internet, and therefore maintain a bias by not being representative, since not all the population has access to these services. Additionally, experience has shown that the cleaning and harmonization work required to use *Big Data* is significant and requires certain parameters and guidelines, for which standardized methodologies often do not exist. In this way, everything could indicate that working with such tools will not completely replace traditional statistics, but in any case complement them (ECLAC, 2020).

Another crucial caveat regarding the analysis of *Big Data* are the safeguards that must be taken regarding privacy and personal data. These tools, especially the link between data science and the use of social networks, allow people to be identified more easily and even develop systems and techniques to predict individual behavior (Davahli et al., 2020; Markovikj et al., 2013). As such, the ethical collection and use of data is an imperative when working with *Big Data*. To this end, there are manuals that define principles to guide the identification and resolution of ethical problems arising from research involving this type of information sources and technologies, such as the Menlo report (Bailey et al., 2012).

[3] Ex. Twitter or Uber.

INTERNET CHARACTERIZATION OF BRAZIL, CHILE, COLOMBIA AND MEXICO

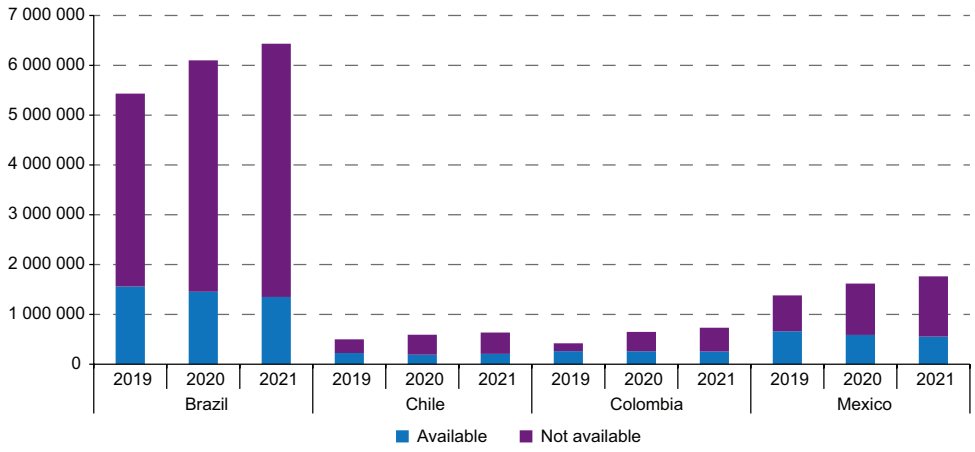
II

A. Overview of the website landscape

To assess the online presence and business activities in Brazil, Chile, Colombia, and Mexico, this exercise begins by analyzing the websites in these countries. To generate data that outlines the internet landscape of each country, web crawling techniques are employed with a tool capable of indexing over 600 million domain names in 50 countries.[4] The collected information is then used to create a comprehensive database, including variables that provide insights into the website's content, technical characteristics, and commercial data. The country of origin for each website is determined by examining various factors, such as the hosting country, language, domain name, and contact details listed on the website (e.g., postal codes, telephone numbers, addresses, etc.). It's worth noting that the database encompasses both country-level domain names (.cl, .br, .co, .mx) and generic domain names (e.g., .com, .org, .net).

In September 2021, a total of 434,750 websites were identified in Brazil, 636,405 in Chile, 732,918 in Colombia, and 1,763,504 in Mexico. However, after conducting a cleanup, it was found that some sites were "unresponsive" due to denial of access or could not be found on the server ("host not found"). As a result, only about 25% of the sites remained available. This translates to a final count of 1,350,163 websites in Brazil, 208,077 in Chile, 254,441 in Colombia, and 554,494 in Mexico. These websites are accessible to the public and offer various types of content (refer to figure 1). Worth mentioning is the fact that between 2019 and 2021, there was a decrease in the number of available sites in all four countries, with Mexico experiencing a decline of 16%, Brazil 13%, Chile 7%, and Colombia 3%.

Figure 1
Websites available and not available in Brazil, Chile, Colombia and Mexico as of September 2019, 2020, and 2021
(Number of websites)

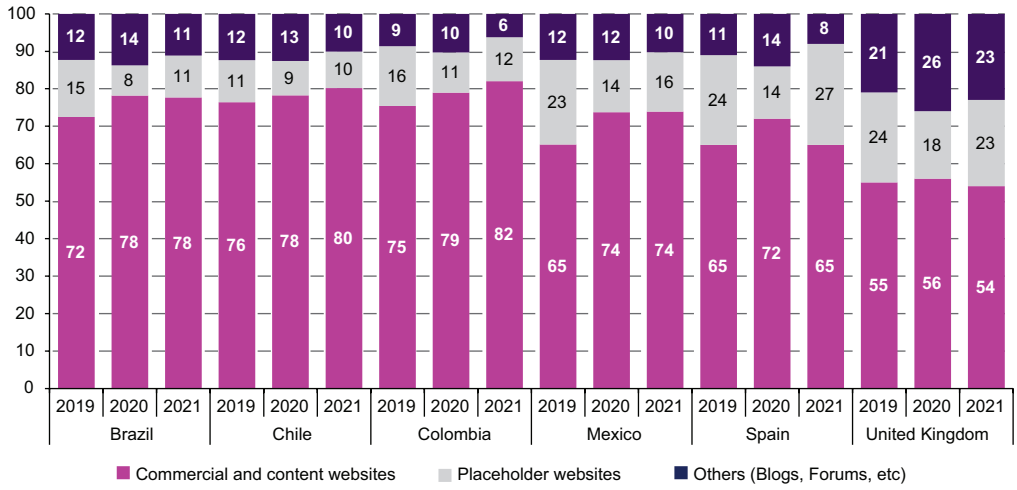


Source: ECLAC with data from Dataprovider.com.

[4] This tool was provided by the company Dataprovider.com, more information at: Dataprovider.com.

Figure 2 illustrates the distribution of websites by type for the four countries studied, as well as Spain and the United Kingdom. The analysis identifies three distinct categories: (i) commercial and content websites, (ii) placeholder websites, and (iii) others (blogs, forums, etc.). The first category encompasses business websites, online stores, and content sites. Business websites present corporate information about products or services and may include online shopping capabilities. Online store sites exclusively serve as electronic marketing platforms for companies. On the other hand, content websites encompass a wide range of creative elements, such as text, images, audio and video files, data, and electronic services (e.g., time reservation, calculators). These websites may also have commercial intent, catering to suppliers and independent professionals, for instance.

Figure 2
Distribution of available websites by type as of September 2019, 2020 and 2021
(Percentages)



Source: ECLAC with data from Dataprovider.com.

As of September 2021, across all four countries, the largest proportion of available websites were commercial and content, accounting for 82% in Colombia, 80% in Chile, 78% in Brazil and 74% in Mexico. These values were higher than those registered in Spain and the United Kingdom, where this type of sites represented 65% and 54%, respectively on the other hand, placeholder sites, which serve to reserve a website that at some point will be completed with some type of content, represented between 10% (Chile) and 16% (Mexico). This category contained about a quarter of the total available sites from the European countries in the sample.

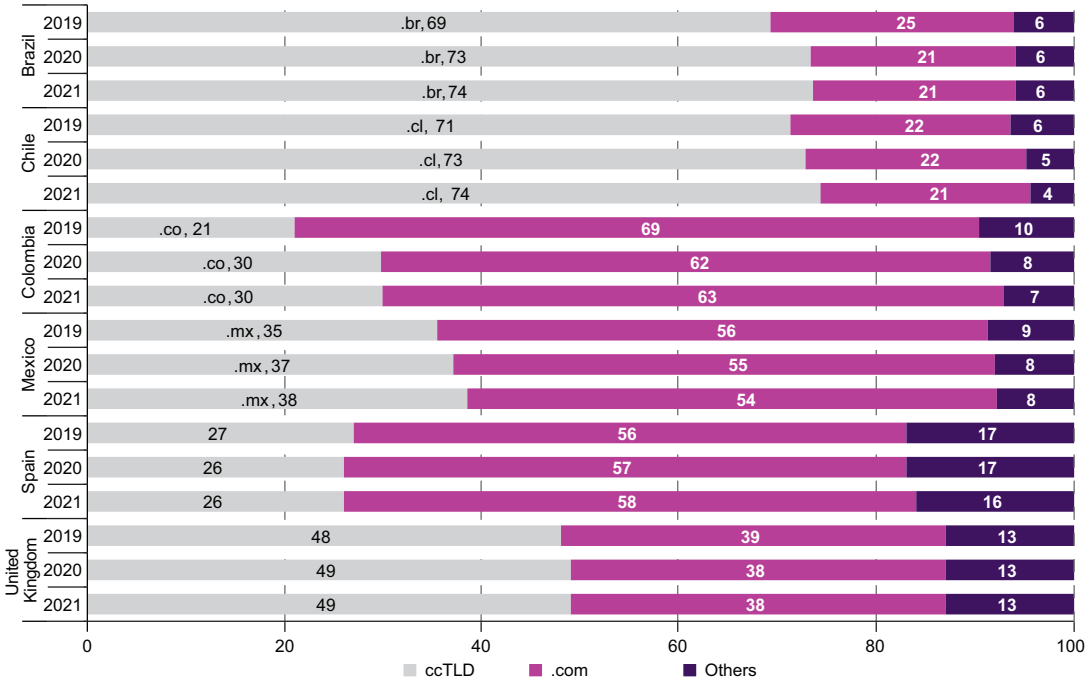
During the period between September 2019 and September 2021, a temporal analysis reveals an interesting trend in the countries of the region. The relative weight of commercial and content sites experienced a significant increase, with Mexico seeing a rise of 14%, Colombia at 9%, Brazil at 7%, and Chile at 5%. Surprisingly, Spain and the United Kingdom experienced little to no change in this regard.

This data suggests a notable surge in online business presence during the months of the pandemic in the countries of the region. It can be attributed to the necessity of maintaining marketing activities, especially in Chile and Colombia where strict and prolonged quarantines were implemented.

Moreover, online store sites doubled their participation in the countries of the region during the same time period. Conversely, the European countries in the sample did not experience any significant change in this aspect.

Figure 3 illustrates the distribution of websites in each country based on the type of domain name used, specifically whether a country code top-level domain (ccTLD) or a generic domain like ".com" (gTLD) is utilized. Notably, Mexico and Colombia exhibit a high prevalence of generic domains, with figures close to 60%. Conversely, in Brazil and Chile, a significant majority of websites, 74% to be exact, are associated with that country's ccTLD in 2021. Spain, compared to its European counterpart, has a relatively low proportion (26%) of sites using country code, while the proportion of ".com" websites reached 58% in the same year. In the United Kingdom, approximately half of the domains end with ".uk". It is worth mentioning that both European countries showcase a greater diversity in gTLDs, with notable participation from endings such as ".org" and ".net" in addition to ".com". These differences can be attributed to various factors, including domain registration strategies and the commercial positioning strategies employed by companies that utilize these domains.

Figure 3
Distribution of available websites in each country by domain name as of September 2019, 2020 and 2021
(Percentages)

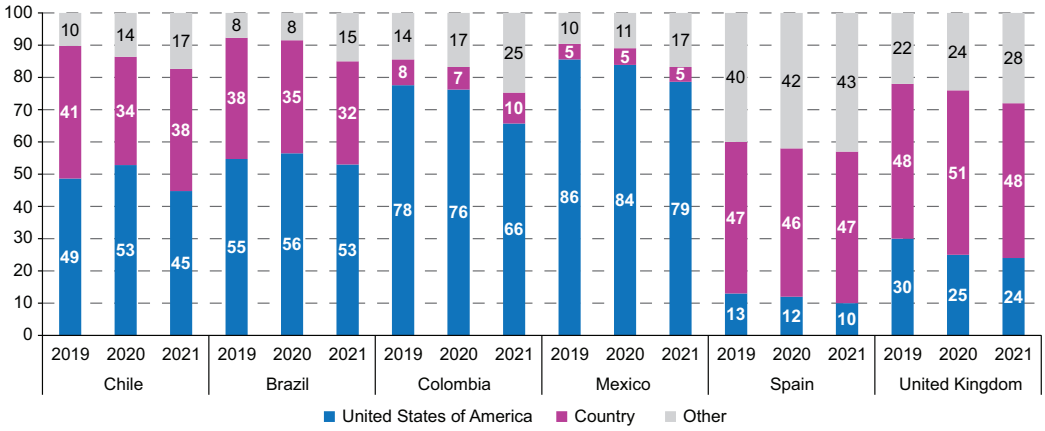


Source: ECLAC with data from Dataprovider.com.

When it comes to website hosting, there is a clear trend among the four countries analyzed - a preference for servers located in the United States of America, particularly in the cases of Mexico and Colombia (see figure 4). As of September 2021, more than 65% of websites in these countries were hosted in the USA. However, there has been a notable decrease since 2019, when the proportions were closer to 80% in Colombia and 90% in Mexico.

In contrast, Spain and the United Kingdom stand out for hosting half of their websites locally, which is a significantly higher proportion. The location of the server in web hosting plays a crucial role in achieving faster loading times for websites. This directly translates into a better browsing experience for customers, leading to fewer abandoned visits, increased chances of converting visitors into customers, and improved rankings on web search engines. This is because crawling robots often associate the hosting location with the geographical location of the company.

Figure 4
Distribution of available websites by hosting location as of September 2019, 2020 and 2021
(Percentages)

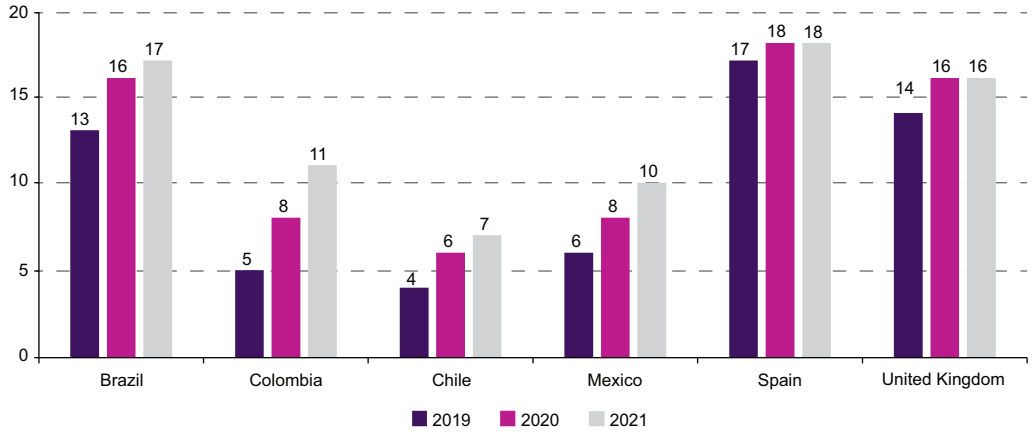


Source: ECLAC with data from Dataprovider.com.

IPv6, the latest version of the Internet Protocol (IP), serves as the foundation for the operation of the Internet. It was developed and deployed to address the limitations of its predecessor, IPv4. The primary driving force behind the design and implementation of IPv6 was the need to expand the available address space on the Internet. This was prompted by various factors, including the emergence of trends like the Internet of Things and an increasing requirement to connect a larger number of devices to the network. Analysis of technical data from various websites reveals a noticeable uptick in the adoption of the IPv6 protocol across the countries examined. Notably, Brazil leads the way in September 2021, with 17% of websites boasting an IPv6 address. Colombia follows closely behind with 11%, followed by Mexico with 10% and Chile with 7%. Brazil's adoption rates are comparable to those of Spain (18%) and the United Kingdom (16%).

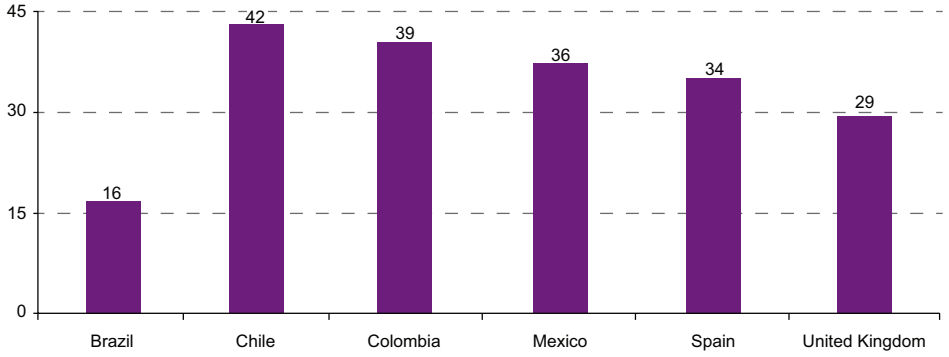
Data protection and privacy are becoming more and more crucial in today's digital landscape, as they directly impact the level of trust that users have in digital solutions. Out of the four countries analyzed, Chile stands out with the highest proportion of privacy-sensitive websites. These websites have the capability to store personal information, allowing users to create accounts, log in, fill out contact forms, and make payments (refer to figure 6).

Figure 5
Available websites with a IPV6 protocol, as of September of each year
(Percentage of total available websites)



Source: ECLAC with data from Dataprovider.com.

Figure 6
Proportion of websites labeled as "privacy-sensitive", September 2021
(Percentages)



Source: ECLAC with data from Dataprovider.com.

B. The changing Internet landscape: effects of the COVID-19 pandemic

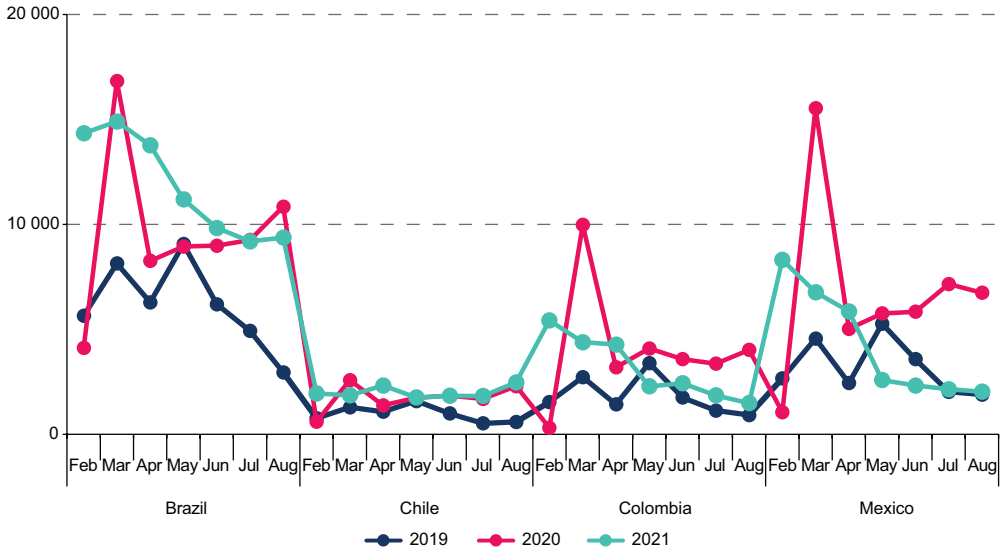
The COVID-19 pandemic has undeniably propelled individuals and businesses to swiftly embrace new digital tools and shift their operations online. With the implementation of quarantines and social distancing measures mandated by governments, numerous shops, restaurants, and services were compelled to transition to the virtual realm. This sudden shift posed a challenge for many businesses that previously relied solely on static websites or social media profiles, as they were now confronted with the urgent need to establish a dynamic online presence for sales, marketing and customer relations. Considering these circumstances, this study holds particular significance in examining the transformative effects brought about by the COVID-19 pandemic.

Upon analyzing the data on new business websites during the initial months of confinement in comparison to the months preceding the pandemic, a significant surge in all four countries under examination can be observed. This surge indicates a notable increase in entrepreneurial activity. The highest growth occurred in March 2020, which aligns with the implementation of the first social distancing measures. Following this boom, the number of new business websites experienced an immediate decline but then returned to a sustained growth in the subsequent months of that year.

Between February and August 2020, the average monthly growth of websites was remarkable: 48% in Brazil, 57% in Chile, 511% in Colombia, and 224% in Mexico. These growth rates are particularly striking when compared to the same months in 2019. In that year, Brazil witnessed a 4% decrease in available websites, while Chile, Colombia, and Mexico saw increases ranging from 4% to 11%.

In 2021, the growth trend of websites slowed down and even reversed in most countries, except for Chile. Despite facing severe mobility restrictions, Chile continued to experience an average monthly growth of 6% in new websites. Between March and August 2021, Brazil and Chile accumulated more new websites than in the same period in 2020. On the other hand, Colombia and Mexico recorded a lower number of new websites in 2021 compared to the previous year, but still higher than in 2019.

Figure 7
New business websites found between February and August 2019, 2020 and 2021
(Number of websites)



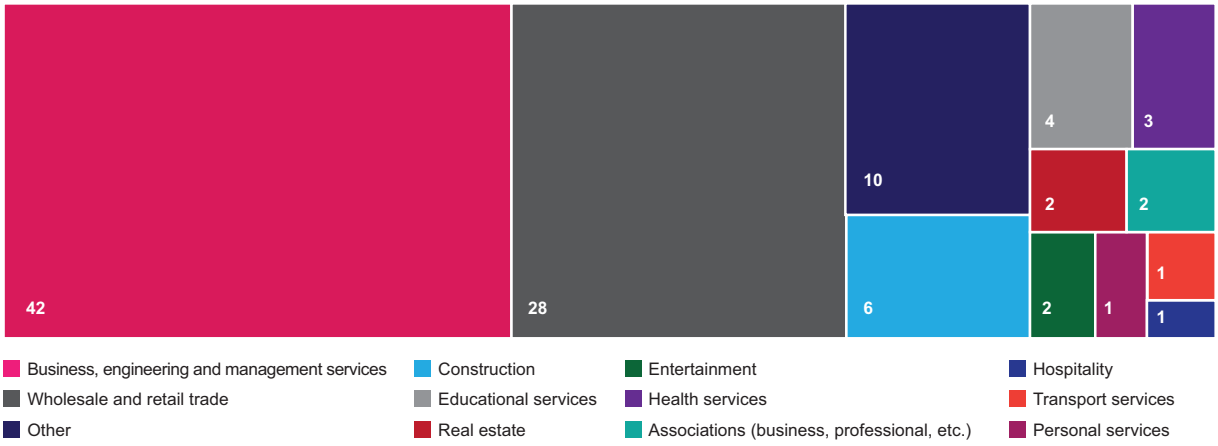
Source: ECLAC with data from Dataprovider.com.

It is worth mentioning that aside from the rise of new business websites, there has also been a notable shift in the nature of online presence between January and June 2020. During the first half of the year, Brazil, Chile, Colombia, and Mexico witnessed a transition of 20% of existing websites from a static site into a transactional site. This indicates a significant change in the online landscape (ECLAC, CAF, and Digital Policy Law, 2020).

During the analysis of commercial websites that emerged between March 2020 and September 2021, it became evident that certain industry segments experienced significant growth. The top sectors were business, engineering, and management

services, which accounted for an impressive 42% of all new business sites. Following closely behind was the wholesale and retail trade sector, making up 28% of the total. The number of websites related to construction, education, and health services also increased notably, further highlighting the prevailing trends in the commercial landscape during this period (figure 8).

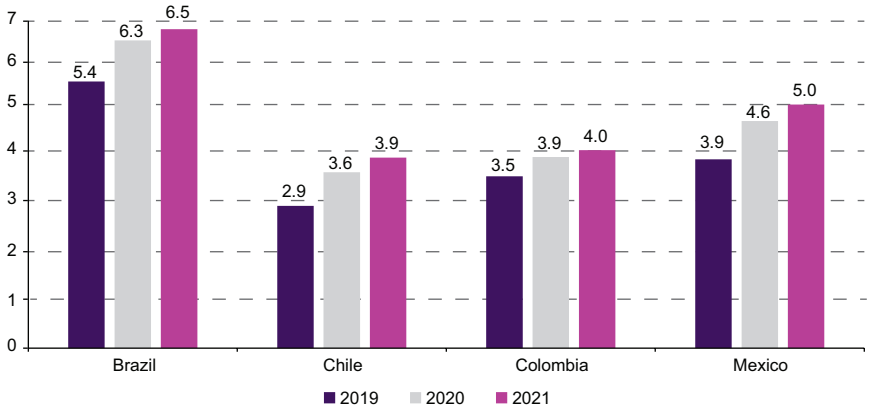
Figure 8
Distribution of new business websites by economic activity category in Brazil, Colombia, Mexico and Chile, between March 2020 and October 2021
(Percentages)



Source: ECLAC with data from Dataprovider.com.

Furthermore, there has been a prominent rise in the number of websites in all four countries that now offer at least one payment option. This indicates that companies have taken proactive measures to enhance their online transactional services. This trend aligns perfectly with the hypothesis that COVID-19 has accelerated the process of digitalization (see figure 9).

Figure 9
Percentage of business websites with at least one payment method available in 2019, 2020, and 2021
(Percentages)



Source: ECLAC with data from Dataprovider.com.

A. Business classification according to online activity and definition of the Internet economy

As mentioned previously, this research aims to analyze companies based on their online presence in order to measure the size of the Internet economy. We accomplish this by utilizing Big Data and employing web scraping techniques to gather public information available on the Internet. Furthermore, we combine this data with business records obtained from national statistical offices. By merging web information with official records, we can gain a more accurate understanding of the activities conducted by formal companies on the Internet. This approach will enable us to identify different groups of companies based on their online behavior.

This study was inspired by the proof of concept conducted by the National Statistical Office of the Netherlands (CBS) on "Measuring the Internet Economy in the Netherlands: A Big Data Analysis" (Oostrom et al., 2016). The objective of their study was to categorize firms based on their use of the Internet, distinguishing those that heavily incorporate this technology into their business models. By supplementing information from business administrative records with data collected through web scraping techniques, they were able to estimate the contribution of these firms to the overall economy. Another similar study, "Measuring the UK's Digital Economy with Big Data" (Nathan et al., 2016), was conducted in the United Kingdom by Growth Intelligence and the National Institute of Economic and Social Research. This study also combined various sources of information, including official statistics, administrative records, and web data, to delve deeper into the analysis of the digital economy.

Those studies found similar conclusions, which can be summarized as: (i) the digital economy (or failing that the Internet economy) is substantially larger than estimated by other traditional classifications; (ii) Internet-intensive businesses have higher average revenues than other businesses; (iii) the digital economy has a substantially higher share of employment than other conventional estimates suggest.

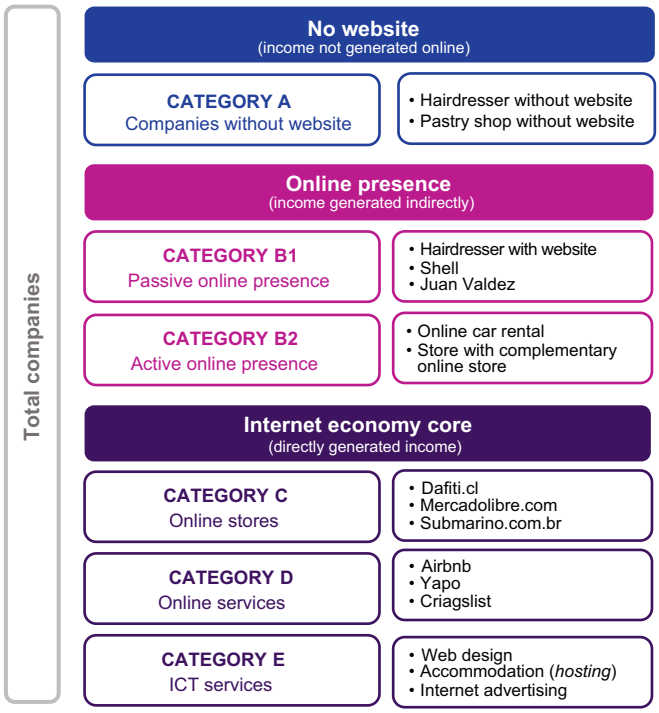
By drawing inspiration from these studies, our own research sought to build upon their methodologies and findings to gain further insights into the digital economy in Latin American countries.

For the purposes of this exercise, the Internet economy will be defined as the set of companies that use the Internet in their business activities and as part of their business model. From this approach, subcategories can be organized according to the level of income generated on the Internet.[5] In this characterization, three main categories are distinguished. The first contains companies that do not generate any online revenue due to their absence on the web (Category A), the second category refers to companies that generate online revenue, but indirectly since their business model is not based exclusively on these marketing channels, (Category B), and finally, companies that base their business models on the web and therefore generate their income mainly from their online activities representing the core of the "Internet economy" (Categories C, D and E). In category B, we also distinguish categories B1

[5] Here we adopt the definition of the Internet economy developed by the Dutch National Statistics Office (CBS).

and B2. The first concerns companies that have a purely passive online presence, i.e. they have a website that contains information about the company and its products or services, but does not allow interaction with it. On the other hand, businesses included in Category B2 allow for some kind of online interaction, for example, subscribing to a newsletter, registering for an account, making a payment, or making a reservation. The core of the Internet economy comprises companies identified as online stores or e-commerce platforms that sell products exclusively through their websites (Category C), companies that offer an online service (Category D), and finally, companies that provide IT services and products. It should be noted that companies in categories C and D include business models that would not exist without the Internet (see diagram 2).

Diagram 2
Business classification based on your online activity



Source: Adapted from Oostrom L. et al. (2016).

To collect information from the Internet, web crawling techniques were used with a tool that allows indexing more than 600 million domain names in 50 countries and that was provided by Dataprovider.com.[6] In the current study, it is crucial to acknowledge that the data utilized is limited to publicly available online sources. This means that transactional information from specific e-commerce platforms like MercadoLibre.com and social networks such as Facebook and Instagram is not included. This limitation is significant because many micro, small, and medium enterprises rely on these platforms to conduct online sales and promote their products and services.

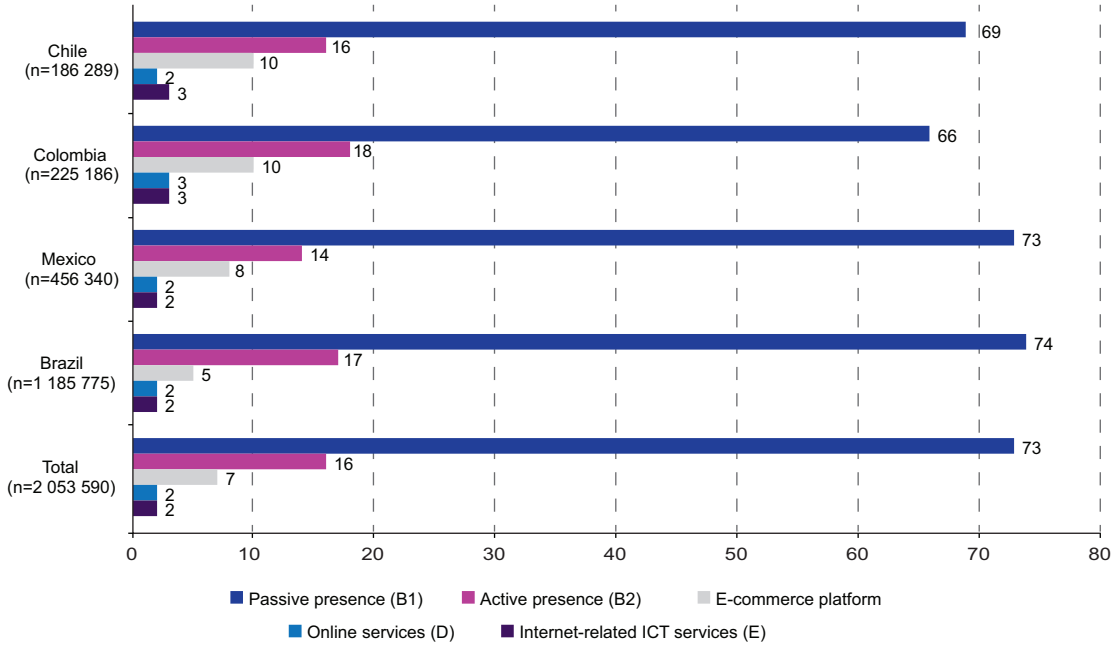
6 A Netherlands-based company specializing in data science.

B. Composition of the Internet economy from web data

Before examining the results of combining website data with information from national business registers, information captured on the web for more than 2 million websites in the four countries will first be analyzed. For this reason, Category A was not included in this part of the analysis.

Figure 10 illustrates the distribution of business websites based on their online activity in four countries as of September 2021. Examining the breakdown of the five categories (B1 – E), it is evident that across all four countries, the majority of company websites, slightly over 70%, had a passive online presence (B1). These websites solely serve as sources of information and do not offer any means for online interaction with customers. Websites classified as having an active online presence (B2) accounted for between 14% (Mexico) and 18% (Colombia) of all websites in their respective countries. Companies owning these websites provide customers with some form of online interaction. Lastly, the Internet core (C-E) constituted between 9% (Brazil) and 16% (Colombia) of all commercial websites, with approximately half of them functioning as online stores. Notably, the distribution among the analyzed countries is quite similar, albeit with minor variations.

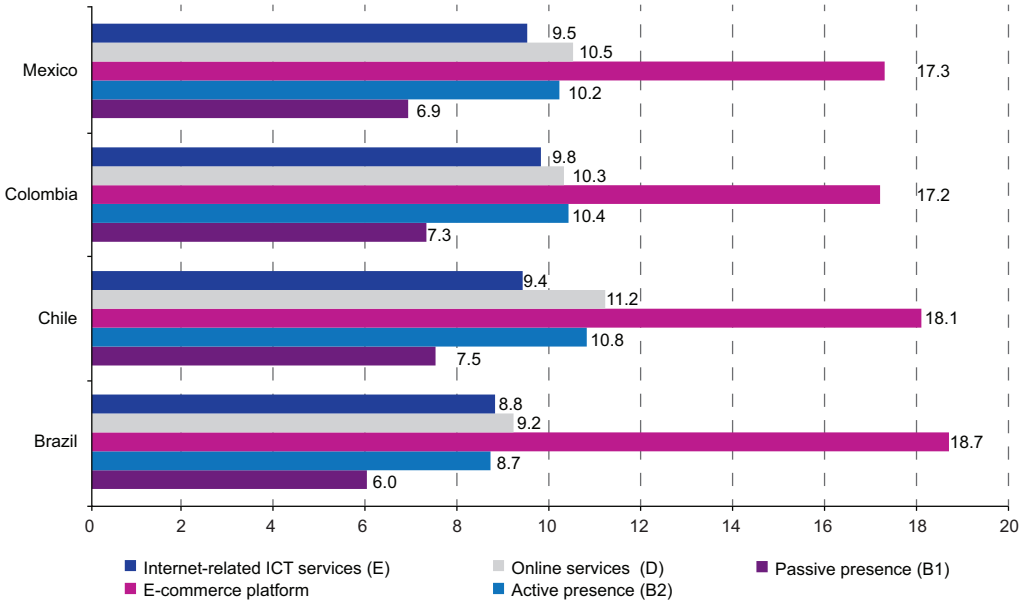
Figure 10
Distribution of business websites by type of online activity by country, September 2021
(Percentages of total business websites)



Source: ECLAC based on Dataprovider.com database.

In all four countries, category B1 (passive online presence) stands out as the largest category. However, when we consider economic activity across all categories, it is e-commerce platforms (category C) that have the highest economic footprint (refer to figure 11). The economic footprint is a metric calculated by Dataprovider.com, providing an estimate of a website's turnover. It is important to note that the economic footprint is on a logarithmic scale, meaning that larger numbers can be hundreds or even thousands of times greater than smaller numbers.

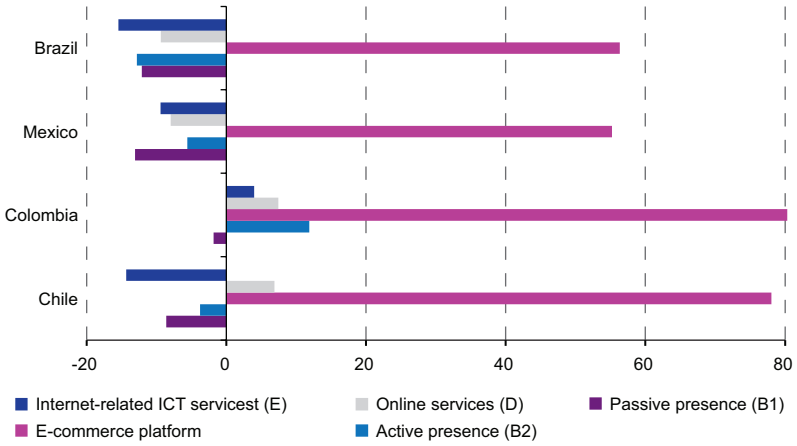
Figure 11
Economic footprint of business websites by type of online activity, September 2021
(Index)



Source: ECLAC based on Dataprovider.com database.

The increase in the number of e-commerce websites between March 2020 and September 2021 is a notable phenomenon, especially when compared to the trends observed in other categories. During this period, after experiencing a surge in growth due to the onset of the pandemic in March 2020, the total number of business websites decreased in Chile (-3%), Mexico (-9%), and Brazil (-10%). Colombia was the only country that saw a modest increase of 6%. Interestingly, the only category that recorded growth across all four countries was e-commerce sites, with remarkable growth rates of approximately 80% in Chile and Colombia, and 55% in Brazil and Mexico (refer to figure 12). Consequently, the weight of the e-commerce category doubled in the total count of business websites across these four countries by September 2021.

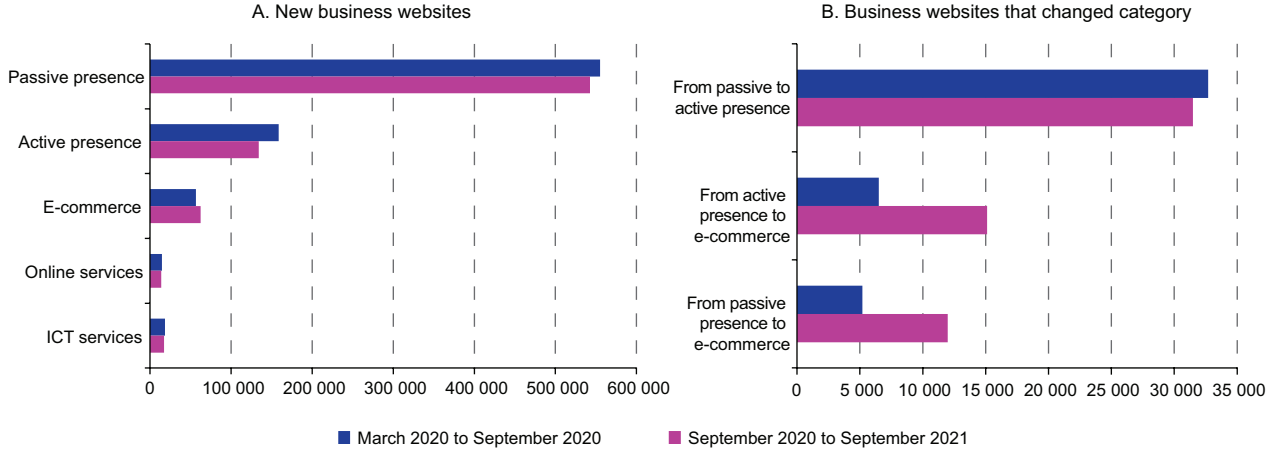
Figure 12
Change in the number of business websites by type of online activity between March 2020 and September 2021
(Percentages)



Source: ECLAC based on Dataprovider.com database.

In our follow-up analysis, we specifically examined websites that transitioned from having a passive online presence to a more active one. We also looked at the number of new websites that emerged in each category. Our findings revealed that the majority of new business websites fell into category B1, which represents a passive online presence. This was followed by new sites with an active online presence (B2) and e-commerce (category C) (refer to figure 13A). Among existing websites, the most significant change was the shift from a passive online presence to an active one. This means that these companies added functionality to their websites to provide customers with opportunities for online interaction. It is also worth noting that a considerable portion of business websites moved from categories B1 and B2 to category C, indicating that these companies began selling products online. This suggests that there are no structural barriers preventing companies from adopting a more sophisticated web presence, which signifies a transformation in their business model.

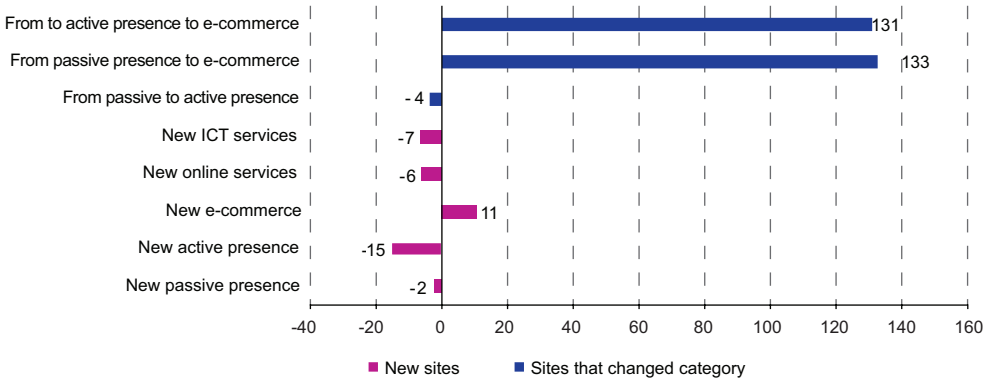
Figure 13
New and upgraded business websites between March and September 2020, and between September 2020 and September 2021 in all four countries



Source: ECLAC based on Dataprovider.com database.

When analyzing the changes in website categories and the emergence of new sites between March and September 2020 versus September 2020 and September 2021, a consistent shift towards the online store format is evident. This trend persists despite a decrease in the number of new websites being created (see figure 14).

Figure 14
Change in the number of business websites by category between the periods March–September 2020 and September 2020–September 2021 in the four countries



Source: ECLAC based on Dataprovider.com database.

C. The Internet economy in Brazil, Chile, Colombia and Mexico: results of the combination of web data and administrative records

The analysis presented in this section examines the Internet economy in Brazil, Chile, Colombia, and Mexico. This analysis is based on a combination of website data and company administrative records, made possible through collaboration with esteemed institutions such as the Brazilian Institute of Geography and Statistics (IBGE), the Regional Center for Studies for the Development of the Information Society (CETIC.br) in Brazil, the National Institute of Statistics of Chile (INE), the National Administrative Department of Statistics of Colombia (DANE), the National Institute of Statistics and Geography of Mexico (INEGI), and NIC Chile.

Table 1
Number of companies by category by online activity for each country, 2020

Category	Brazil (March)	Mexico (March)	Chile (September)	Colombia (March)
Company without website (A)	4 081 207	5 294 567	1 065 336	783 490
Companies with passive online presence (B1)	305 053	44 497	23 097	36 685
Companies with active online presence (B2)	77 206	8 837	6 697	12 888
Online Stores (C)	11 990	3 485	2 386	4 812
Online Services (D)	6 757	993	672	1 995
Internet-related ICT services (E)	16 354	897	692	1 844
Total	4 498 567	5 353 276	1 098 880	841 714

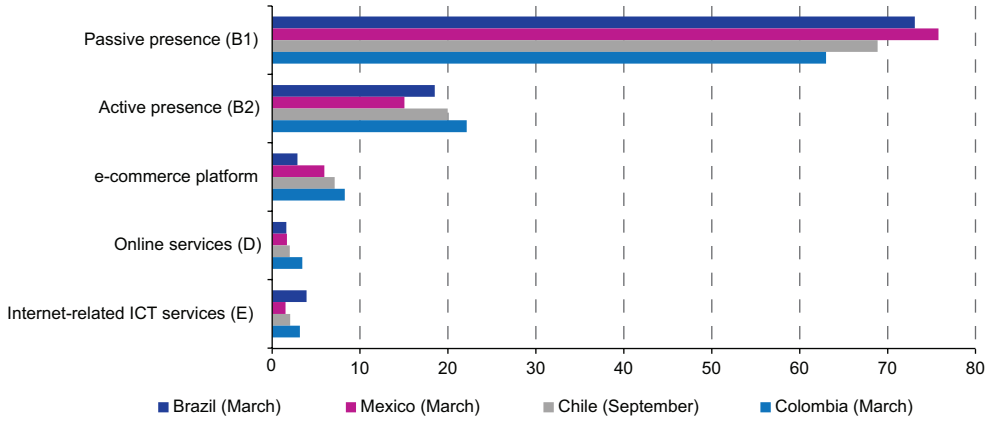
Source: ECLAC, based on data from IBGE (Instituto Brasileiro de Geografía y Estadística), CETIC.br (Centro Regional de Estudios para el Desarrollo de la Sociedad de la Información) of Brazil, INE (Instituto Nacional de Estadísticas) of Chile; DANE (National Administrative Department of Statistics) of Colombia; and INEGI (National Institute of Statistics and Geography) of Mexico and Dataprovider.com.

The highest level of linkage between company records and corresponding websites was observed in Brazil, with approximately 9% of companies in the business registry being successfully linked. Colombia followed with 7%, Chile with 3%, and Mexico with 1%. It is worth noting that the majority of companies, more than 90%, were categorized as Category A, indicating that they did not have a website.

For further details, please refer to table 1, which provides the total number of companies in each country's register and the corresponding category assignments after the linking process. Overall, this analysis sheds light on the state of the Internet economy in these countries and offers valuable insights into the digital landscape of Brazil, Chile, Colombia, and Mexico.

Figure 15 illustrates the distribution of linked websites by category and type of online presence across different countries. The distribution of categories remains consistent across the four countries. As anticipated, most companies have a passive online presence, accounting for 63% to 73% of the total. Following this category are companies with an active online presence. Core Internet companies (categories C, D, and E) make up a smaller proportion, but there are variations among countries. For instance, in Colombia, online stores constitute approximately 8% of the total, whereas in Brazil, they represent only 3%.

Figure 15
Percentage of companies successfully linked to a website by category, 2020
(Percentages)



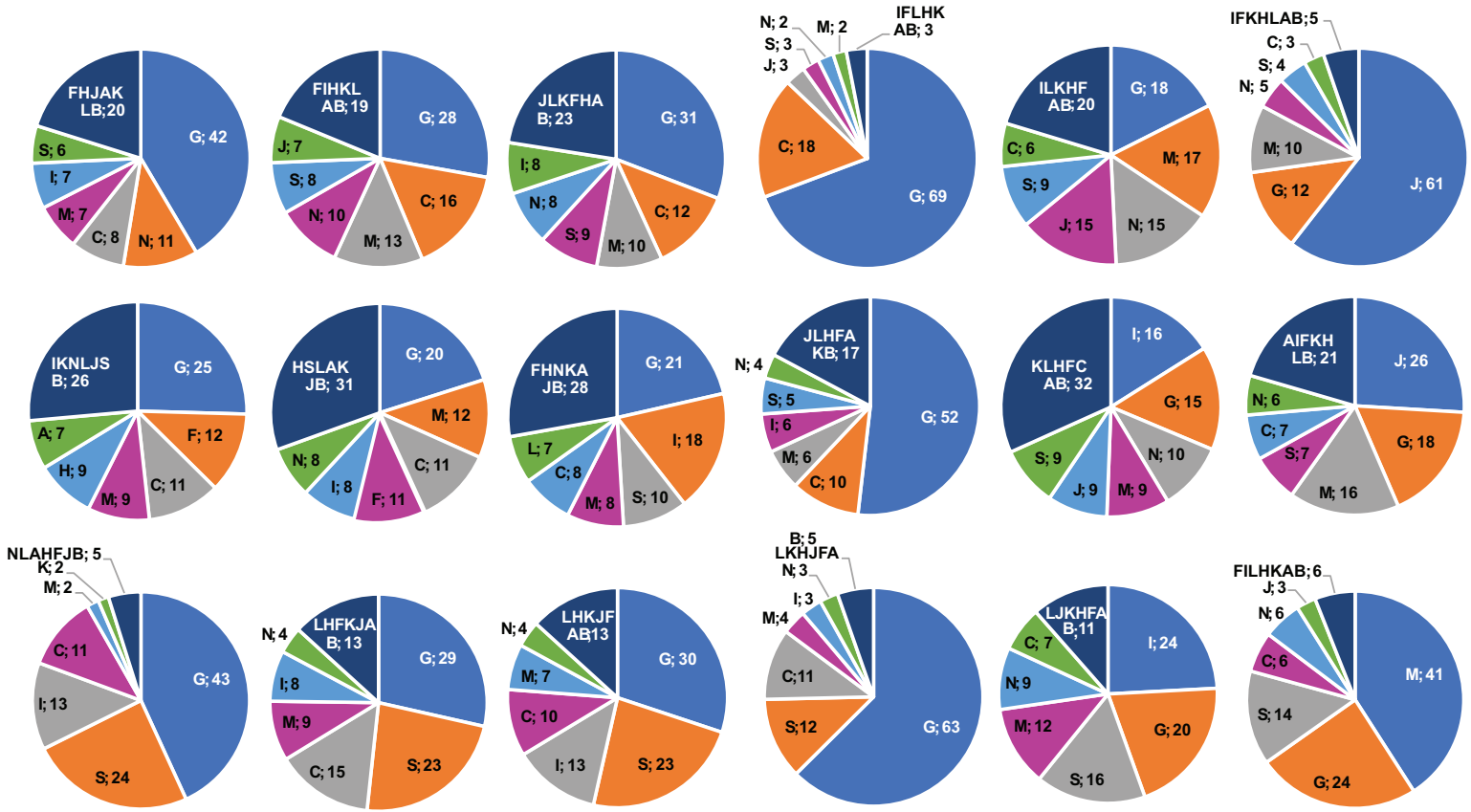
Source: ECLAC, based on IBGE (Instituto Brasileiro de Geografía y Estadística), CETIC.br (Centro Regional de Estudios para el Desarrollo de la Sociedad de la Información) of Brazil, INE (Instituto Nacional de Estadísticas) of Chile; DANE (National Administrative Department of Statistics) of Colombia; and INEGI (National Institute of Statistics and Geography) of Mexico and Dataprovider.com.

The allocation of ISIC (International Standard Industrial Classification) codes within each category for Brazil, Chile and Mexico was examined[7] (see figure 16). In almost all categories except D (online services) and E (ICT services related to the Internet), the sector with the greatest presence is the trade sector (wholesale and retail stores, sector G). Likewise, this sector contributes significantly to the category of online stores (between 52% and 69%). In category B1 (passive presence), following the trade sector, the main activities are manufacturing (C), professional, scientific and technical activities (M) and other activities (S). In category B2 (active presence), the distribution of activities is similar to category B1, but with an important presence of accommodation and food services (I), demonstrating a significant level of digitalisation of these companies. Category D (online services), following trade, consists mainly of professional, scientific and technical activities (M), administrative activities (N), other services (S), and accommodation and food services (I). In Brazil and Chile, ICT and Internet services (E) are linked to the information and communication sectors as expected from the definition of this category.

The fact that this sector is not prominent in Mexico probably reflects the low number of companies that were linked and indicates that the data are not representative.

7 Data from Colombia were not included in this representation as they were only able to obtain data from the following SIC (C, G, H, I, J, K, M, N, P, R), none of which were represented in Category C.

Figure 16
Industry share (SIC code) by internet category
(Percentages)

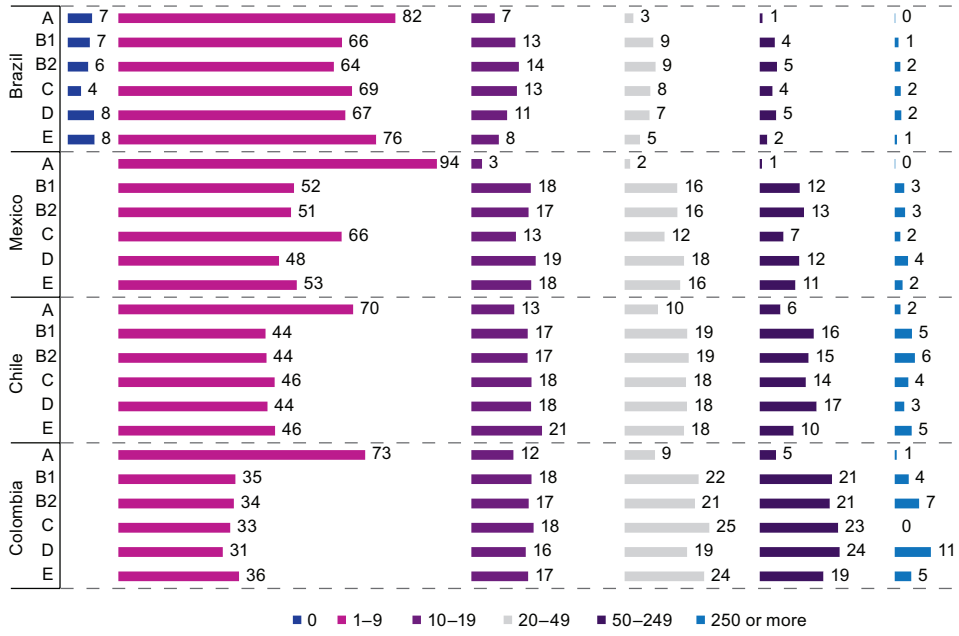


A Agriculture, forestry, livestock and fishing
B Mines and quarries exploitation
C Manufacturing industries
F Construction
G Wholesale and retail trade
H Transportation and storage
I Accommodation and food services
J Real estate activities
K Financial and insurance activities
L Real estate activities
M Professional, scientific and technical activities
N Administrative and support services
S Other service activities

Source: ECLAC, based on IBGE (Instituto Brasileiro de Geografía y Estadística), CETIC.br (Centro Regional de Estudios para el Desarrollo de la Sociedad de la Información) of Brazil, INE (Instituto Nacional de Estadísticas) of Chile; DANE (National Administrative Department of Statistics) of Colombia; and INEGI (National Institute of Statistics and Geography) of Mexico and Dataprovider.com.

In all countries, it was possible to evaluate the categories of the Internet economy according to the size of the company. Figure 17 shows the distribution of firm size by number of employees for each category. In all four countries, companies without a web presence are mostly small businesses with fewer than 9 employees and the percentage of large companies (>250 employees) without a website is low. The rest of the categories (B1-E) are evenly distributed in all countries, except in Brazil where the classification of 50 to 249 employees is relatively lower. Likewise, in Brazil, large companies are evenly distributed in categories B1 to E. Remarkably, in Colombia, a significant proportion of large companies are in the online services category (D).

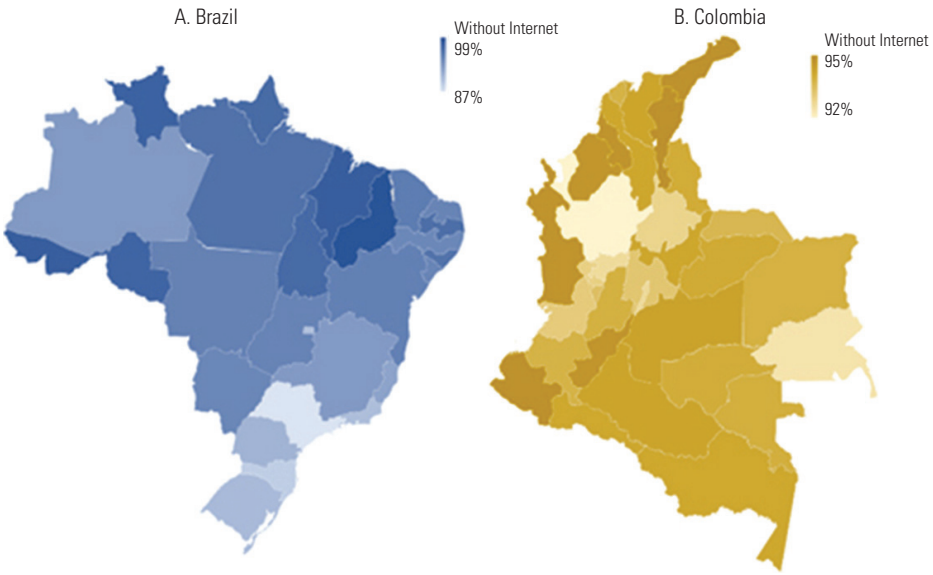
Figure 17
Participation of companies by category of Internet presence according to their size (number of employees)
(Percentages)



Source: ECLAC, based on IBGE (Instituto Brasileiro de Geografía y Estadística), CETIC.br (Centro Regional de Estudios para el Desarrollo de la Sociedad de la Información) of Brazil, INE (Instituto Nacional de Estadísticas) of Chile; DANE (National Administrative Department of Statistics) of Colombia; and INEGI (National Institute of Statistics and Geography) of Mexico and Dataprovider.com.

Through business records, it is possible to analyze the geographical distribution of companies with an online presence. In the case of Brazil, this analysis reveals a strong correlation between the low online presence of companies and the most economically disadvantaged regions, such as the northeast region. In Colombia, the results illustrate a similar correlation, but additional variables like armed conflict also seem to play a role. Notably, the regions with the lowest number of companies with an online presence in Colombia are El Cesar, La Guajira, Córdoba, Sucre, Choco, and Nariño (map 1).

Map 1
Regional differences of companies without a website for Brazil and Colombia
(Percentages)



Source: ECLAC based on IBGE (Brazilian Institute of Geography and Statistics), CETIC.br (Regional Center for Studies for the Development of the Information Society) of Brazil, DANE (National Administrative Department of Statistics) of Colombia and Dataprovider.com.

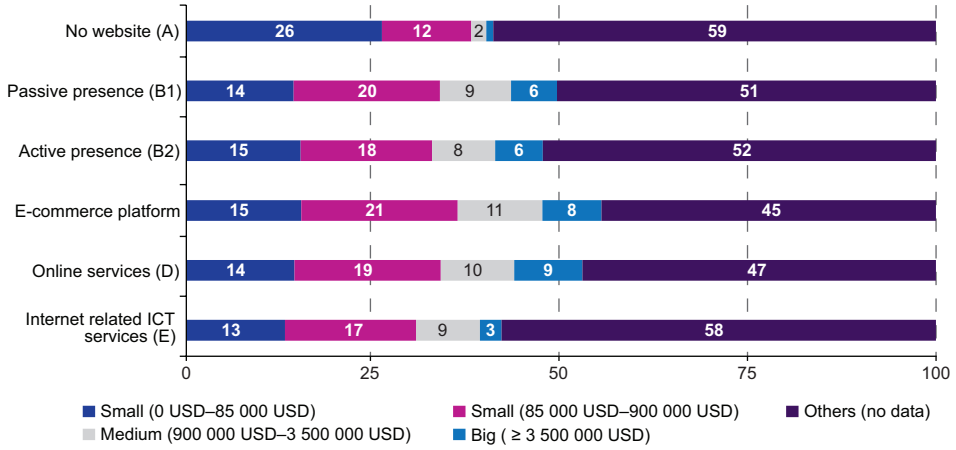
Note: The boundaries and names shown on this map do not imply official endorsement or acceptance by the United Nations.

Not all country business registers provide the same level of information. In Chile, access to more comprehensive data allows for the analysis of annual sales by category, taking into account the type of online presence of companies as well as their export/import orientation.

Figure 18 illustrates the distribution of companies based on their annual sales volume, categorized by their online presence. As of March 2020, micro-enterprises showed relatively equal participation across all categories, except for category A (no online presence), where they dominated, indicating a low level of digitization within this segment. On the other hand, small, medium, and large companies demonstrated a lower level of involvement in category A due to their higher degree of digitalization. Conversely, categories C (online stores) and D (online services) witnessed a greater presence of medium and large companies, reflecting their respective levels of sales. This highlights the existence of a digital divide among different business segments, but also underscores the potential for generating higher revenues through digitalization.

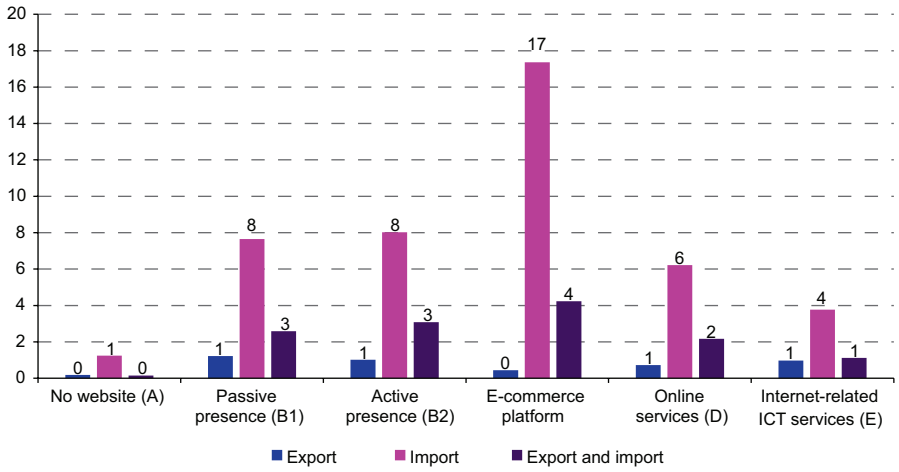
When examining international trade orientation of companies based on their online presence, it becomes evident that, as of March 2020, Chilean companies lacking an online presence predominantly focused on the domestic market. This indicates that these companies primarily operated within their local sphere (see figure 19). Category C, which represents online stores, had the highest number of companies classified as importers. This was closely followed by passive (B1) and active (B2) online presence. On the other hand, categories D and E, which fall under the Internet core, had a lower proportion of importing and exporting companies. This can be attributed to the fact that their activities are primarily focused on non-tradable services.

Figure 18
Share of companies by annual sales volume by online presence category in Chile, March 2020
(Percentages)



Source: ECLAC based on INE (National Institute of Statistics) of Chile and Dataprovider.com.

Figure 19
Participation of companies identified as Exporters, Importers or both, Chile, March 2020
(Percentages)



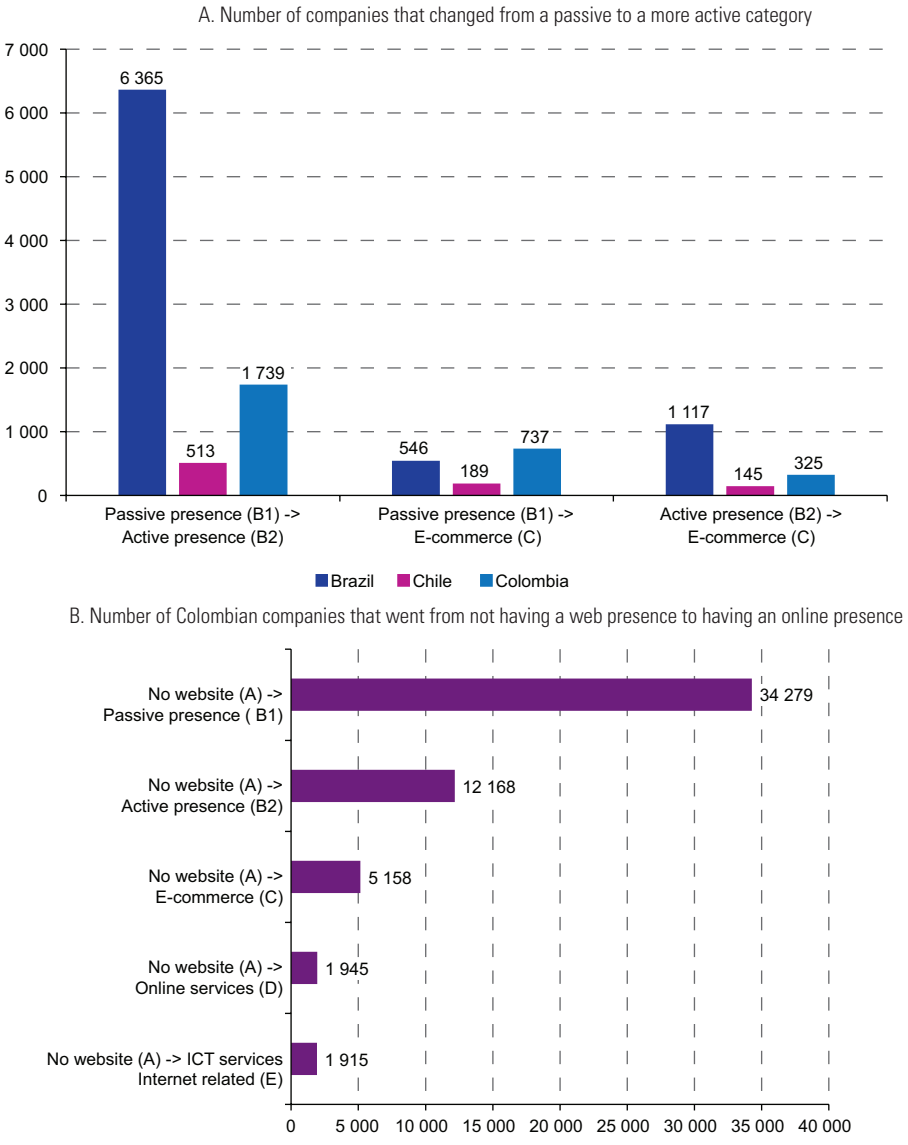
Source: ECLAC based on INE (National Institute of Statistics) of Chile and Dataprovider.com.

1. Effects of the pandemic: tracking changes between March and September 2020

The COVID-19 pandemic had major effects on the behavior of online businesses. As stated previously, there was an increase in the number of websites in general and a greater sophistication of these sites' features for customers to interact online, as well as increased e-commerce offerings. The data below analyzes companies that were able to be successfully linked with administrative records of the study countries, both in March and September 2020, and that successfully transitioned from a passive to an active online presence.

Figure 20.A shows the number of firms that switched from a passive to an active online presence or added an e-commerce component in Brazil, Chile, and Colombia (data for Mexico were not available). A change from category B1 to category B2 means that this website added an additional feature that was not initially present, for example, the possibility to make a reservation or subscribe to a newsletter. A change from B1 or B2 to C reflects the addition of e-commerce features, for example, a shopping cart system or the acceptance of payment methods. Figure 20 (b) shows the shift for Colombia from companies that were not online (Category A) in March to the different online categories in September 2020. This data demonstrates not only the impact of the pandemic on the behavior of the digitalization of companies, but also their ability to move to more advanced levels of Internet use.

Figure 20
Companies that changed their type of online presence



Source: ECLAC, based on IBGE (Instituto Brasileiro de Geografía y Estadística), CETIC.br (Centro Regional de Estudios para el Desarrollo de la Sociedad de la Información) of Brazil, INE (Instituto Nacional de Estadísticas) of Chile; DANE (National Administrative Department of Statistics) of Colombia and Dataprovider.com.

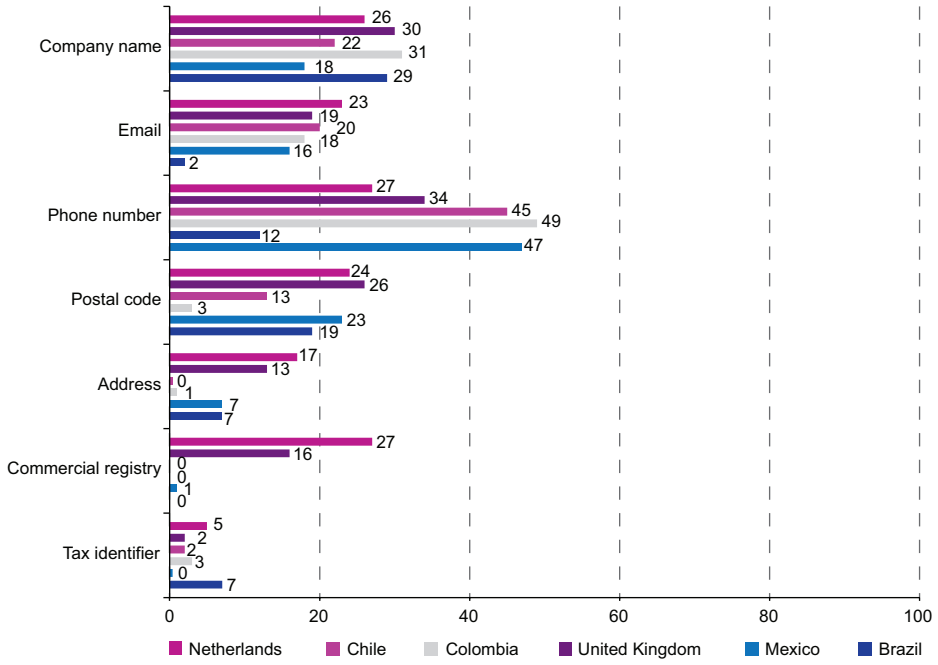
2. Methodological considerations in combining web data and business records

One of the main challenges in merging the website database with a country's national business registers was finding the right information to make the data link. For example, in the study conducted by Statistics Netherlands (CBS), the main combination point was the business registration number (CoC), a data that most companies in the Netherlands tend to publish on their website. In addition, many companies register their website in the commercial register, so two main sources of data could be accessed to create a link. Additional information such as email addresses and telephone numbers were used only in exceptional cases, where the above variables were not available.

The availability of data is a key aspect that differentiates the case of the Netherlands from the four countries participating in this study, as the number of websites providing a business registration number in these countries is very low. Therefore, other variables were included to allow a correspondence between a company's website and its registration in the national commercial register. Information used to make the link included company name, address, zip code, area code phone number, and email address.

Figure 21 shows the availability of this information in the four countries of the study and its comparison with the Netherlands. It is striking that less than 1% of websites provide a business registration number, compared to 17% in the Netherlands. This difference in the availability of information is also relevant for other variables such as the telephone number, however, the availability is comparable when considering other information such as the name of the company.

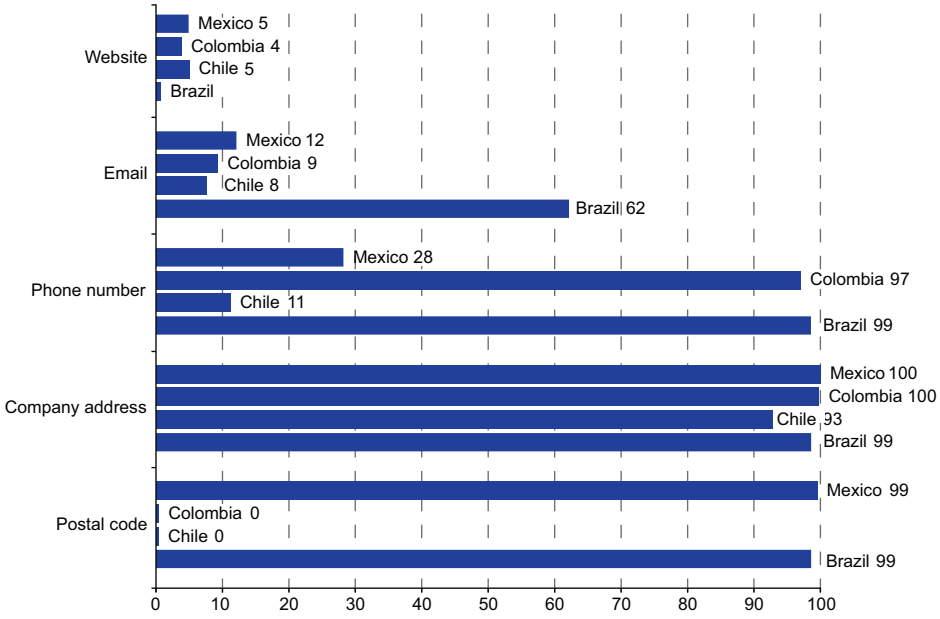
Figure 21
Information availability in the web database for six variables used for the combination (information as of November 2021)
(Percentages)



Source: ECLAC based on Dataprovider.com.com data.

Figure 22 examines the information available in business registers. In all countries the number of companies that have a URL registered in the name of their company ranges from 0.7% in Brazil to 5% in Chile. Similarly, except for Brazil, email addresses are available for less than 13% of companies, posing major challenges in combining data collected from the web and data available in company registries. Address availability is high in all four countries, but due to the lack of reliable information on postal codes in Chile and Colombia (information used to verify addresses in the web database), there is very little overall data availability in these two countries.

Figure 22
Availability of information within each country's business register for five variables used for the combination
(Percentages)



Source: ECLAC, based on IBGE (Instituto Brasileiro de Geografía y Estadística), CETIC.br (Centro Regional de Estudios para el Desarrollo de la Sociedad de la Información) of Brazil, INE (Instituto Nacional de Estadísticas) of Chile; DANE (National Administrative Department of Statistics) of Colombia; and INEGI (National Institute of Statistics and Geography) of Mexico.

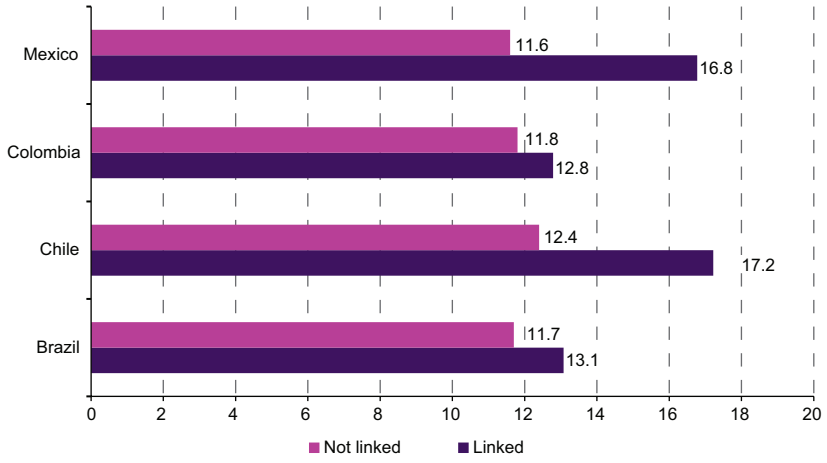
The discrepancies indicated above on the availability of information on business records and web information created challenges when trying to combine a greater amount of information. In Colombia, most businesses were matched using phone numbers (42,559) followed by the URL (14,972). Only three companies were linked based on their tax number. In Mexico, most companies were combined using the URL (45,654), while in other cases (8,072) the URL of the web database could be linked to the name of the company. Only 474 companies could be linked using the business registration number in the web database with the federal taxpayer registration number (RFC). In Brazil, the combination was carried out in two stages with the help of the NIC.br. In a first combination, the web database was linked to the NIC.br register to obtain commercial numbers (CNPJ), then, in a second stage, the numbers were combined with the commercial register (CEMPRE). A total 377,001 (85.8%) of the merged businesses were successfully identified through their website. The remaining 14.4% were compared by identifying the CNPJ through the website and email information from other sources (e.g. through the Central Bank of Brazil – BACEN).

3. Characteristics of the websites that were linked to the business register

In this section we present some preliminary data examining the differences between the websites of companies that were and were not successfully linked with national business registers. This section serves to better understand the representativeness of the data and to better understand the characteristics of the companies that were successfully connected.

Figure 23 shows the average economic footprint of websites that were and those that were not linked. Websites that successfully mapped to companies in the business register exhibit an overall higher economic footprint than websites that were not linked.

Figure 23
Average economic footprint of websites that were merged and those that were not merged in national business registries
(Index)



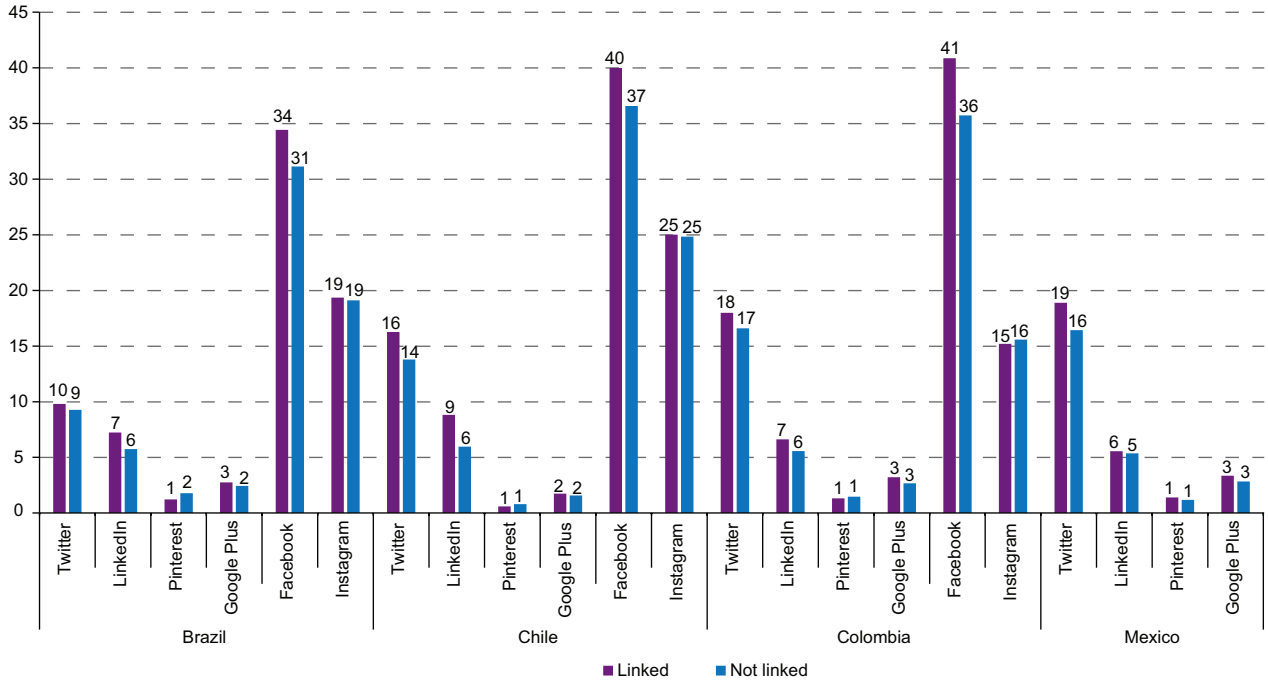
Source: ECLAC based on Dataprovider.com.

Additionally, successfully linked companies have a greater presence on some social media platforms, including Facebook and Twitter, while there is little difference with regard to their presence on other social media platforms, e.g. Instagram (see figure 24). Taken together, websites that were successfully linked with companies in the business registries tended to have higher online activity and overall greater economic impact. Presumably, these sites offer more information about the company on their websites, including important contact details.

Figure 24

Percentage of linked and unlinked websites based on social media profiles listed on the website

(Percentages)



Source: ECLAC, based on IBGE (Instituto Brasileiro de Geografía y Estadística), CETIC.br (Centro Regional de Estudios para el Desarrollo de la Sociedad de la Información) of Brazil, INE (Instituto Nacional de Estadísticas) of Chile; DANE (National Administrative Department of Statistics) of Colombia, and INEGI (National Institute of Statistics and Geography) of Mexico and Dataprovider.com.

IV FINAL THOUGHTS

This study serves as a proof of concept with the primary objective of characterizing the Internet economy in four Latin American countries. By delving into how companies operate online, it aimed to provide a better understanding of the digital landscape in these countries. In the region, this study stands as the first attempt to merge traditional data sources with unstructured data from the web. This innovative approach has the potential to generate groundbreaking indicators and shed light on crucial aspects of the digital economy that would otherwise remain elusive. However, despite the immense possibilities, there are still significant challenges that need to be overcome to fully harness the potential of these new data sources. In this regard, the section below summarizes some of these challenges and provides specific recommendations to enhance the outcomes of similar research endeavors. Additionally, it highlights the most noteworthy findings about the Internet economy.

Perhaps the biggest challenge faced by the project was successfully linking the information available on a given website with the information in national business registers due to the lack of overlapping data. In similar studies carried out in other countries (e.g. the one carried out in the Netherlands), most companies publish their commercial registration or tax number on their website. In addition, the commercial register in these countries also presents information about the companies' websites. Both practices are less common in the four countries that participated in the current research.

To address this issue, it would be beneficial for national registries to include inquiries about companies' online presence (including websites and social media profiles) in their surveys. Additionally, policies should be encouraged for companies to disclose basic business information on their websites, such as their business or tax records. This would not only facilitate better identification of companies, but also provide greater transparency for consumers. For example, in Chile, the Ministry of Economy, Development, and Tourism implemented the Electronic Commerce Regulation in 2021 (Library of National Congress, 2021). This regulation requires companies operating on e-commerce platforms to provide essential information, such as tax registration, legal address, and email. Such policies can serve as a model for other countries to enhance transparency and accountability in the business sector.

Among the main findings of the study is that most companies still maintain a passive presence online. However, it was also revealed that companies have the capacity to move towards more advanced stages in their levels of digitization and use of the Internet. The impact of the COVID-19 pandemic on this transformation is clearly visible. Additionally, the study has uncovered some of the challenges faced by smaller companies when establishing and managing a website. This study does not capture the activities of companies, particularly smaller ones, that rely on platforms and social networks to maintain their online presence. Future research should aim to incorporate and examine these activities to gain a better understanding of the dynamics of these businesses.

While data was not available for all countries regarding company revenues, the available data from Chile revealed some interesting findings. For instance, categories within the core of the Internet economy, such as online stores and online services, tend to generate higher revenues compared to companies with lower levels of digitization. Additionally, the study found a geographical disparity in the online presence of companies

in countries where data was accessible, such as Brazil and Colombia. This disparity may be correlated with income levels and other challenges, such as infrastructure deployment. Furthermore, the study confirmed the expected dominance of the services sectors in the core of the Internet but also highlighted the noteworthy participation of the manufacturing sector in these categories.

Finally, despite facing significant challenges, this project has successfully generated valuable knowledge, opening the door for future research to capitalize on the ever-increasing amount of data derived from the web and web-related activities. Notably, the project has demonstrated its effectiveness in analyzing business behavior and the Internet economy. It has also shed light on the profound impact of the COVID-19 pandemic on online business operations. It is undeniable that companies are increasingly compelled to digitize their activities, a trend that is expected to gain even more momentum in the years to come. For governments it is therefore crucial to access relevant information in order to monitor and understand these dynamics and formulate policies that facilitate and foster a productive digital transformation.

BIBLIOGRAPHY

Bailey, M. et al. (2012), "The Menlo Report", *IEEE Security & Privacy*, vol. 10, No. 2, March.

ECLAC (2022), *A Digital Path for Sustainable Development in Latin America and the Caribbean,* Santiago, Economic Commission for Latin America and the Caribbean, November.

___(2021), "Foreign Direct Investment in Latin America and the Caribbean 2021", *ECLAC* (LC/PUB.2021/8-P).

___(2020), *Analysis of the digital footprint in Latin America and the Caribbean: lessons learned from the use of big data to assess the digital economy*, Big Data Project for the Measurement of the Digital Economy, United Nations Development Account, Santiago, Economic Commission for Latin America and the Caribbean, July.

___(2018), "Data, algorithms and policies: the redefinition of the digital world", ECLAC, United Nations, Santiago, April.

ECLAC & Adenauer, F. K. (2021), *Economic recovery after the COVID-19 pandemic: empowering Latin America and the Caribbean to better take advantage of electronic and digital commerce,* Santiago, ECLAC, United Nations, October.

ECLAC, CAF, & Digital policy law (2020), *The opportunities of digitalization in Latin America in the face of COVID-19,* Santiago, Economic Commission for Latin America and the Caribbean, April.

Davahli, M. R. et al. (2020), 'Identification and Prediction of Human Behavior through Mining of Unstructured Textual Data', *Symmetry,* vol. 12, 19 November.

European Commission (2014), "Commission expert group on taxation of the digital economy", Paris, European Commission, May.

Layney, D. (2001), *3D Data Management: Controlling Data Volume, Velocity, and Variety,* META Group, January.

Markovikj, D. et al. (2013), "Mining Facebook Data for Predictive Personality Modeling", *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 7, No. 2.

Nacional, B. del C. (2021), "Library of the National Congress | Ley Chile", *www.bcn.cl/leychile*, [online] <https://www.bcn.cl/leychile> [accessed on 20 April 2023].

Nathan, M. et al. (2016), 'Measuring The UK's Digital Economy With Big Data'.

OECD (2020), "A Roadmap Toward A Common Framework For Measuring The Digital Economy", Saudi Arabia, Report for the G20 Digital Economy Task Force, January.

___(2011), *OECD Guide to Measuring the Information Society 2011*, Paris, Organisation for Economic Cooperation and Development, August.

Oostrom, L. et al. (2016), 'Measuring the internet economy in The Netherlands: a big data analysis'.

UNCTAD (2022), "Handbook for the Production of Statistics on the Digital Economy", Geneva, United Nations Conference on Trade and Development.

Ward, J. S. & Barker, A. (2013), "Undefined By Data: A Survey of Big Data Definitions". United Nations publication.

The adoption of digital technologies is a crucial tool for bridging the productivity gaps between countries in Latin America and the Caribbean and more developed nations, generating new sources of growth, and creating high-quality jobs. Evidence-based policies are needed to harness the potential of these technologies, guide technological change, seize opportunities and mitigate risks.

This publication presents an exploratory study conducted in Brazil, Chile, Colombia and Mexico, in which various sources of information are combined, including web data and official statistics, to measure the online activity of businesses. This methodology allows companies to be classified based on their Internet use, going beyond traditional industrial classifications and giving rise to a new classification. Moreover, the study explores the possibilities offered by big data techniques and tools for enhancing our understanding of the digital transformation and serves as a foundation for future research in this field.

ECLAC

Economic Commission for Latin America and the Caribbean (ECLAC)
Comisión Económica para América Latina y el Caribe (CEPAL)
www.cepal.org/en

LC/TS.2023/51