Original Research Article

Official statistics and big data in Latin America: Data enclosures and counter-movements

Oscar Arruda d'Alva[1] (iD) and Edemilson Paraná[1,2,3] (iD)

Abstract
With the rise of the data-driven economy, the state-owned sector of official statistics has been pressurised to 'modernise' and engage with big data. However, most of these data sources are controlled by tech corporations. We address the conflicts this brings about from a Latin American perspective through three case studies involving national statistical offices, international organisations, and the private sector. We investigate strategies for enabling data markets for official statistics and analyse how the statistical field has acted in this context. In doing so, we contribute to understanding the political economy of big data in Latin America and debates on how digitalisation encompasses the reshaping of state-business relations. Supported by secondary data and semi-structured interviews, we appraise Bourdieu's theory of fields, Marxian readings on the enclosure of commons, and Polanyi's double movement to analyse how data commodification challenges data as public goods – a fundamental principle for official statistics. The findings demonstrate that data enclosures have prevented the state's access to compiling official statistics and show that the initiatives for introducing big data in Latin American national statistical offices have involved testing data markets through public–private partnerships supported by international organisations and big tech. As a result, the statistical field has reacted to the data market with a 'double movement': mobilising symbolic capital such as 'trust' for partnering with businesses to access data and technologies and, conversely, defending the public value of data in counter-movements protective of the relative autonomy of the national statistical offices and the state's control over informational capital.

Keywords
Official statistics, big data, big tech, data enclosures, Latin America, counter-movements

This article is a part of special theme on Critical Data Studies in Latin America. To see a full list of all articles in this special theme, please click here: https://journals.sagepub.com/page/bds/collections/critical_data_studies__in_latin_america

Introduction

Official statistics are central instruments for politics, economy and society, constituting a fundamental element for public information in democratic systems. Between the 19th and 20th centuries, nation-states mobilised economic and cognitive investments for the consolidation of national statistical offices (NSOs). This ensured quality and reliability in the production and use of common measures in the most diverse aspects: demographic, economic, social, environmental and political (Desrosières, 1998).

Our core argument is that due to the increasing digitalisation of the economy and society in the 21st century, the capacity of nation-states to accumulate informational capital (Bourdieu, 2014) and produce relevant and timely

[1]Graduate Program in Sociology, Federal University of Ceará (UFC), Fortaleza, CE, Brazil
[2]Department of Social Sciences, LUT University, Lappeenranta, Finland
[3]Brazilian National Council for Scientific and Technological Development (CNPq), Brasília, Brazil

Corresponding author:
Edemilson Paraná, Department of Social Sciences, LUT University, Room 4616, Yliopistonkatu 34, Lappeenranta 53850, Finland.
Email: edemilson.parana@lut.fi

statistics is compromised by data enclosures carried out by private corporations (Verdegem, 2022). In this context, privately held sources of big data constitute assets in emerging data markets and official statistics, traditionally a state-own sector, becomes a market niche disputed by tech companies. The logic of capitalisation within new markets aims to convert data into digital commodities and, in so doing, to enable the extraction of knowledge-rents (Rotta and Paraná, 2022). This logic confronts the conception of data as public goods, a fundamental principle for official statistics, which triggers disputes and different strategies within the statistical field.

The article focuses on how this process has been developing recently in Latin America and pursues three main goals: (i) to understand the political economy of the introduction of big data for official statistics in the region; (ii) to investigate strategies of private and public agents to enable data markets for official statistics; and (iii) to analyse how the statistical field has acted in this context.

The main findings confirm that data enclosures have prevented the states' access to relevant sources of big data for compiling official statistics and further show that: (i) the initiatives for introducing big data in Latin American NSOs have involved the testing of business models and the promotion of data markets through public–private partnerships (PPPs) and platformisation with the support of international organisations (IOs) and big techs; and (ii) the agents within the statistical field react in different ways regarding data markets, featuring a double movement (Polanyi, 2001). We identify cooperation with pro-market initiatives as counter-movements in defence of the public value of data and NSOs' autonomy. By doing so, agents seek to preserve their positions in the field, which involves the mobilisation of symbolic capitals and control of informational capital (Bourdieu, 1998).

More broadly, the paper contributes to debates on how the new data-driven economy reshapes relations between states and businesses. The issue is relevant as the unprecedented advance of the market over official statistics may pose risks to national sovereignty over data, insofar as it can compromise the relative autonomy of the NSOs. The Latin American countries addressed in the article have historically built strong statistical systems and recently started to use big data. Our investigation offers a privileged point of view on how this process may escalate through the region and the Global South.

The article is structured into five sections: (i) Initially we present the research's methodological and theoretical approaches and then (ii) address the canonical and recent literature, highlighting the innovation and relevance of our framework. Next, (iii) a case studies section provides empirical evidence on the introduction of big data for official statistics in Latin America. This is followed by (iv) a discussion section in which we analyse the empirical data in light of the theoretical approach mobilised. We conclude by (v) summarising the main findings and indications for further research.

Theoretical approach and methodology

The article presents three case studies involving Latin American NSOs, IOs and the private sector. Two cases address Colombia and Brazil, and a regional case also involves Chile and Mexico. The cases are supported by secondary data that is collated with primary data collected through 21 virtual interviews conducted with representatives of NSOs and IOs between May and October 2022.[1] We have used a snowball sampling technique in which key informants were indicated to interviewers from their referential network. The data analysis considered two dimensions: (i) actions guided by a logic of commodification (pro-market movements); and (ii) actions oriented by a logic of public value (protective counter-movements). A cross-sectional analysis was carried out regarding the types of capitals involved in each case (see Table 1).

Regarding the theoretical approach, Bourdieu's theory of fields and total capital (Bourdieu, 1986, 1998) in light of recent research (Ruppert and Scheel, 2021) supports our understanding of disputes over different forms of capital within the statistical field. We follow Bourdieu and Wacquant's (1992: 97) definition of the field as a network of objective relations between positions (occupied by agents and institutions), which are defined by the structure of the distribution of species of capitals 'whose possession commands access to the specific profits that are at stake in the field'. Informational and symbolic capitals are especially approached through the cases. The first refers to a broader definition of cultural capital that encompasses all forms of realisation of theoretical knowledge, whether in an objectified state as goods, an institutionalised state as official titles, or an embodied state in the dispositions of minds and bodies (Bourdieu, 1986). The latter comprises any form of capital 'when it is perceived by social agents endowed with categories of perception which cause them to know it and to recognise it, to give it value', for example, honour, prestige, trust etc. (Bourdieu, 1998: 47).

We also instrumentalise the Marxian elaboration on the enclosure of the commons (Marx, 1976), in line with recent literature (Dean, 2016; Verdegem, 2022). With this, we seek to characterise the movement of private appropriation of data from a political economy perspective. Finally, we mobilise Polanyi's (2001) concept of double movement. The author identified how the state acted to meet the social resistance against market efforts to convert land, labour and money into fictitious commodities at the birth of industrial capitalism so as to attend to the interests of the self-regulated market economy. From this perspective, we analyse double movements concerning data commodification within the statistical field.

Addressing the literature

Statistics, as a tool for validating scientific knowledge and as an instrument for political action, became an object of

Table 1. Case studies framework.

Case studies	Pro-market movements	Protective counter-movements	Disputed capitals
Case 1 – The Latin American Hub of the UN Global Platform (Brazil)	Public–private partnership and plataformisation. Development of technologies (PET) and services for data sharing. Private third-parties commercial interests in providing services and technologies.	Defence of a new statistical law to enhance NSO's access to new forms of data. NSO's development of autonomous in-house knowledge and technology for compiling official statistics from publicly available big data.	Symbolic: trust (trusted partners), public good, social good and data for development (SDGs). Informational: access to data, technologies and knowledge. Economic: business case for platformisation and PPPs, new data sharing technologies, products and services for NSOs.
Case 2 – Data-Pop Alliance and the OPAL project at DANE (Colombia)	Public–private partnership. Defense of deregulating the national data environment. Development of technologies (PET), business cases and products for data sharing.	Public officers' alleged resistance to 'disruptive' big data and data sharing. Limitation of national PPPs in big data for official statistics.	Symbolic: Social good (not-for-profit social innovation). Informational: access to data, technologies and knowledge. Economic: business model and commercial product for data sharing.
Case 3 – ECLAC-Dataprovider (Brazil, Chile, Colombia, Mexico)	Public–private partnership. Purchase of data from private third-party. Promotion of a European business model (Google-sponsored PPP) in LAC.	NSOs' autonomous adaptation of web scraping without a private third party. Recommendations to NSO to explore legal alternatives for accessing data.	Symbolic: trust (seal of quality) Informational: access to data, technologies and knowledge. Social: access to NSO network. Economic: data sale, commercial contracts with NSOs and IOs.

investigation in the social sciences from the 1970s onwards (Beaud, 2015). Pioneering works on the social history of statistics address two main dimensions. On the one hand, studies on the systems of statistics production, and on the other hand, research on statistical reasoning. Early contributions to a sociological and philosophical approach to statistics include works by Bourdieu and Foucault.

Bourdieu (2014) reflected on statistics as instruments for accumulating informational capital that provided the territorial states the means for concentrating economic and military capacities for its centralization. According to Bourdieu (2014), a qualitative transformation of a series of private capitals into public capitals, characterises the transition to the modern nation-state. We take this framework to investigate official statistics as public goods controlled by a field of practices that differentiated itself within the state's meta-field.[2] For Bourdieu (2014: 141), official statistics, in their public form and as state acts, 'impose a legitimate view of the social world'. This is done by a work of classification and universalization, which contributes to the inculcation of common cognitive structures and produces a 'belief effect', that provides the nation-states with an extraordinary symbolic power.

The original contribution of Foucault (2007) to the social studies of statistics was to understand it as a tool that provides the instrumental articulation of the two technological sets that constitute the security mechanism of liberal governmentality: the police and diplomatic-military apparatuses. In his account, statistics operate as a biopolitical and rationalisation tool for assessing and controlling the internal forces of the state (population and territory), on the one hand, and on the other hand, it constitutes an instrument for calculating the dynamics of forces between states into a global competition space. We take this 'double governance' as a source of recurring tension in the statistical field. This derives from the contradictions between its internal role, subject to national sovereignty, and its external role linked to the transnational network of statistical internationalism.

In the 1990s, the sociology of quantification by Desrosières (1998) opened up a field of research dedicated to the study of statistics as a tool of proof and government. Since then, new branches of social studies of quantification have proliferated (Espeland and Menniken, 2019). According to Desrosières (1998), statistical practices rely on a 'space of equivalence conventions', which is demarcated mainly by the state and sustained by a set of institutions, norms, practices and tools. The struggles and controversies within these cognitive and institutional networks demonstrate that such a space is never definitively fixed. By understanding the historical development of statistical tools vis-à-vis the relationship between the state and the market, Desrosières (2008) identifies a crisis in national statistical systems from the advent of the neoliberal regime, encompassed by the emergence of new private and transnational centres of quantification. We follow his findings to understand how this process unfolds in the contemporary data economy.

As digital technologies become central to the global economy, new areas of social studies have emerged.

Recent literature has addressed aspects such as data extraction (Pasquinelli and Joler, 2021), data enclosures (Verdegem, 2022) and data colonialism (Couldry and Mejias, 2019). Studies have also tackled the political economy of new digital technologies. Exemplary are the notions of platform capitalism (Srnicek, 2016), intellectual monopoly capitalism (Rikap, 2021) and AI capitalism (Dyer-Witheford et al., 2019). We dialogue with these works in their investigation of asymmetric relations of technological and geopolitical power characterised by new dispossessions and monopolies that emerge in the data-driven economy. Our original contribution is to provide evidence from the social construction of data markets for official statistics and its implications for the global-south.

Few works have sought to articulate contemporary studies of datafication with social studies of statistics. An exception can be found in Ruppert and Scheel (2021), whose empirical research analyses how the use of big data by European NSOs has been transforming statistical practices and shaping new forms of subjectivation through data. Our approach, however, differs by emphasising the political economy aspects of this process.

A field of study on the social history of Latin American statistics has developed since the 2000s (Lanata-Briones et al., 2022; Otero, 2018). This corpus of research has contributed to demonstrating that Latin American countries are not only passive consumers of foreign methods but have innovated, actively adapted and built their statistical systems with relative autonomy. However, most of the research focuses on the 19th and 20th century periods and still does not address the contemporary process of big data.

Overall, the scholarly literature on the use of big data for official statistics is European-oriented and tends to focus on technical and instrumental aspects, such as applications and constraints of new sources and methodological issues (Allin, 2021; Kitchin, 2015; MacFeely, 2019; Struijs et al., 2014). The political economy of big data for official statistics is an area still little explored by research. With this work, we contribute to the literature with a critical approach on the subject from a Latin American perspective.

Latin American case studies

Latin America is a diverse region. Territorial, political, economic and social differences reflect considerable heterogeneity in its national statistical systems. However, the central role played by the NSOs and the existence of specific statistical laws regulating the systems, stand out as a strength and common feature of the Latin American statistical field (ECLAC, 2010).

According to the Statistical Performance Indicators – SPIs (World Bank, 2023) there is high statistical capacity in a set of Latin American countries when compared to other global regions. This stems from historical factors, such as the role played by statistics in the post-colonial

making of nation-states in the 19th century (Otero, 2018), the importance of NSOs for the developmental experiences in the mid-20th century, and more recent opportunities for the statistical agenda in the democratisation processes (Dargent et al., 2018).

Grouping Latin American and Caribbean (LAC) countries according to their SPI (Figure 1),[3] countries like Brazil, Chile and Mexico (cluster 1) show a statistical performance compatible with the European average, while countries like Colombia (cluster 2) stand out above the Global South average. Accordingly, these countries have assumed a more prominent position in the use of big data in the region and are included in the following case studies.

Case study 1: The Latin American hub of the UN global platform in Brazil

Within the scope of the 2030 development agenda of the United Nations, the case for a 'data revolution for sustainable development' (UN, 2014) gave room within the UN Statistics Division (UNSD) for a strategy directed at intensifying the use of big data for official statistics. According to the 'data revolution' discourse, the traditional statistics produced by NSOs would be insufficient for monitoring the sustainable development goals (SDGs). The use of big data by NSOs through PPPs with business was then encouraged as an essential action to measure and achieve the objectives of the agenda.

Following the UN framework, a set of so-called 'data for development' and 'data for social good' initiatives emerged, focusing mainly on the Global South countries. Since NSOs act as the countries' focal points at the UN Statistical Commission (UN-StatCom),[4] the 'UN Big Data' initiative has significant repercussions in these institutions. In this regard, two developments within the UN stand out: (i) the Global Partnership for Sustainable Development Data (GPSDD); and (ii) the UN Global Platform.

Based at the UN Foundation,[5] GPSDD bills itself as a network dedicated to spreading the 'data revolution' as a 'force for good' (GPSDD, 2020: 2). The initiative's corporate partners include Facebook, Google, Microsoft and IBM (GPSDD, 2023b). The Bill & Melinda Gates Foundation and Google.org stand among its donors (GPSDD, 2023a). GPSDD operates in 30 countries in the Global South and its plans for 2023 included increasing the number of governments using big data from private sources, brokering at least 10 PPPs, and developing technical advice for at least 20 NSOs (GPSDD, 2020). GPSDD is also responsible for managing the UN Global Platform (UN-StatCom, 2021: 8).

The UN Global Platform is an offshoot of the 3rd International Conference on Big Data for Official Statistics of the UN. The conference was configured as a space for the networking of senior managers of NSOs and IOs with representatives of big techs (Microsoft, Google,
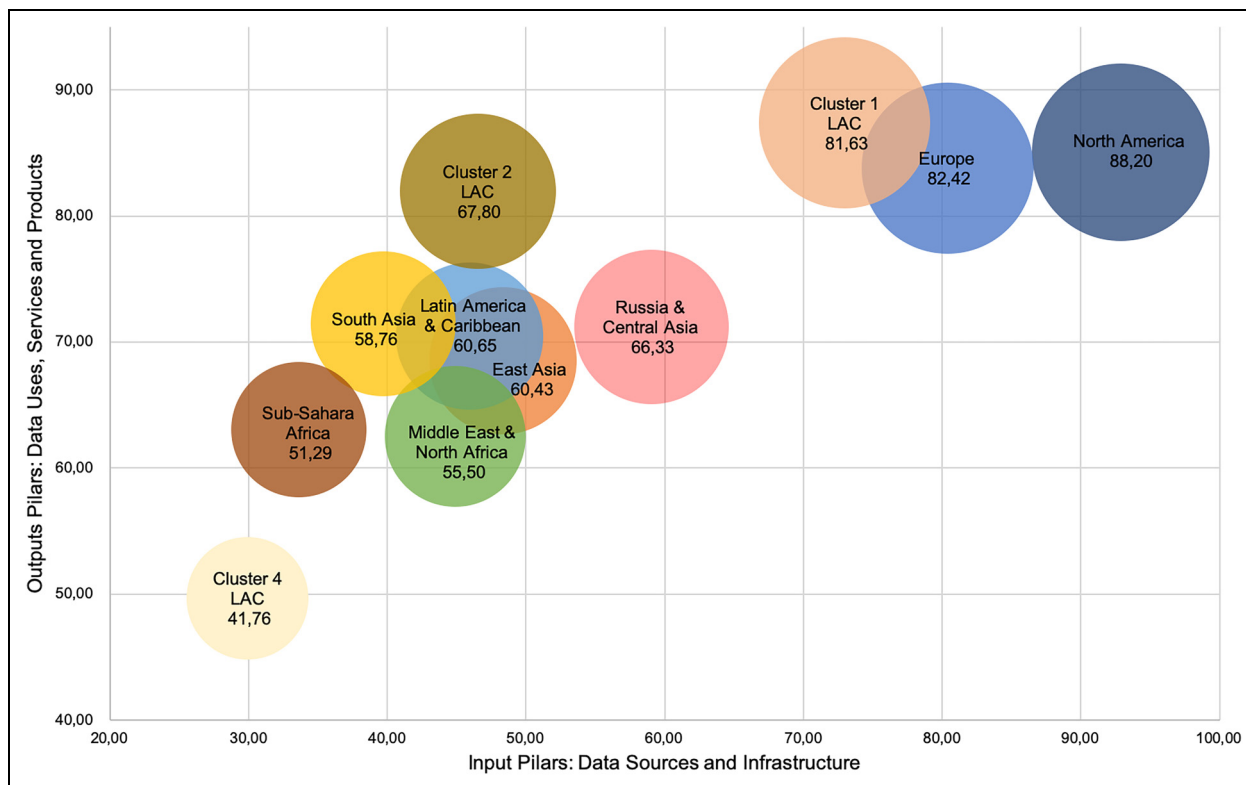
Figure 1. Statistical performance indicators-global regions and Latin American and Caribbean (LAC) clusters.
Source: World Bank (2023). Authors' own computation.

Amazon and IBM) and other tech companies. According to a press release of the event:

> This play between public and private sector in the field of official statistics raises the question of what National Statistical Offices have to do to stay relevant […] Partnerships with the private sector in the use of big data for official statistics seem to be the only way forward. If you can't beat them, join them. (UN, 2016: 1)

Following the conference, the UN Global Platform was proposed as an initiative to support the international statistical community in developing 'a public–private partnership that should extend current and future initiatives to make better use of innovative data sources at the national and regional levels' (UN-StatCom, 2017: 1). Notably, trust is quoted as an important asset of the Platform:

> The UN Global Platform is a policy, technical and business infrastructure which supports an international ecosystem of statisticians, data scientists and other partners […] It delivers a trusted environment for collaborative data analysis based around four pillars: Trusted Partners, Trusted Data, Trusted Methods (algorithms), and Trusted Learning. (UN-StatCom, 2020a: 10)

The Platform's business plan states its contribution to 'the delivery of the SDGs and the modernization of the statistical system whilst also supporting sustainable profits for commercial partners' (UN-StatCom, 2019: 6). Besides the NSOs, the Platform's 'trusted partners' include companies such as Microsoft, Samsung, Positium, Flowminder and other business and research representatives (UN-StatCom, 2020b: 9). According to the plan:

> Collaboration on the platform will give multinational companies opportunities to test their products and services in a global community and gain access to potential government and other customers […] Commoditization of products on the platform will gain from quality and trust of the platform and from access to platform partners in this global marketplace. Trusted partner status will have reputational value for commercial organizations that will translate into positive messages for company and product promotion. (UN-StatCom, 2019: 34)

The Platform comprises the implementation of four regional hubs in the Global South at NSOs in Asia, the Middle East, Africa and Latin America. The hubs aim to build capacities, expand partnerships and support the Platform's implementation (UN-StatCom, 2020b). The

Latin American hub was implemented at the Brazilian Institute of Geography and Statistics (IBGE) in 2021.

The institutional strategy of the IBGE from 2017 to 2027 foresaw new data acquisition strategies, a data intelligence committee, and plans for a new statistical law. The proposition of a new regulation aimed to strengthen the coordination role of the NSO in the Brazilian statistical system and to assure systematic access to administrative registers and new forms of data (IBGE, 2017). As of 2019, the discontinuity of this strategy relates to changes in the NSO's management during the extreme right-wing Bolsonaro's government.

Under Bolsonaro's (2019–2022), the IBGE was directed by two presidents with no former experience in national statistics coordination and was exposed to intense controversies. In particular, the reduction of the 2020 census budget and scope, allegedly justified by alternative use of administrative records, was heavily criticised by former presidents of the IBGE and led to the resignation of the first nominated president.

It was amidst this controversial context that the IBGE settled a memorandum of understanding (MoU) with the UN for the implementation of the 'Regional Hub for Big Data in Brazil in support of the United Nations Global Platform' (IBGE, 2021). The objectives include the facilitation of big data projects in Latin America and the 'additional development and maintenance' of the Platform (IBGE, 2021). The regional hub has been carrying out research and workshops with NSOs in the region and partners of the Platform. In addition, IBGE officers are participating in the UN Big Data task teams.[6]

There are at least three experiences underway at the IBGE with new data sources: web scraping for e-commerce statistics, web scraping for price statistics, and the experimental use of mobile phone data.[7] The most successful initiative, which effectively evolved from the experimental phase to the production of official statistics, is web scraping for an airline ticket price index. The project was developed autonomously by IBGE technicians and new applications are under development (e.g. accommodation prices). Initiatives using web scraping at the IBGE tend to advance since there is greater ease of access to web data. Conversely, experiments involving the use of privately held data have been limited to tests.

As part of the UN task team on mobile telephone data, the IBGE developed a project to measure internet access using data from a local mobile phone operator in partnership with the company Positium (ITU, 2021). According to an IBGE officer, despite demonstrating potential, the experience did not go beyond tests. This stems from the difficulty of accessing data from private companies:

> We can't get the data. The president of the IBGE is very interested, but what I told him is that there's no point in us taking the test again […] we already know […] some areas where we can use the data. Now we have to get this

data in a consistent way, sit down with these companies, sit down with the government, I think it's a very political thing. (Interview, 25 May 2022)

We have identified different views among IBGE officers on data access strategies. For a coordinator involved in the big data regional hub, there is a need for new regulation: 'all countries need to have legislation that says that all data owners of any nature have the obligation to provide them [to the NSOs]' (Interview, 16 May 2022). According to a former president of the IBGE, the proposal for a new statistical law which, among other things, aimed to ease access to new sources of data, was under consultation within the NSO, the Ministry of Planning and a Congress Committee, but from 2019 'everything changed. The new IBGE president came in and completely abandoned that initiative. And why? In her mind she had to modernise the IBGE and bring big data' (Interview, 9 June 2022).

A different approach to data access comes from an IBGE statistician, who sees emerging methods of data sharing using privacy enhancing technologies (PETs) as a possible solution, 'much closer than this discussion of access to data that is being fought with great difficulty' (Interview, 15 May 2022). Still, according to this statistician, Microsoft Research has been actively involved at the UN Big Data in the development of new services for data sharing in the official statistics sector: 'I think that's what Microsoft is thinking there, to be the provider of this service. It will perhaps be the first [tech] giant doing this'. Notably, Microsoft, Google and IBM's products and services in encryption, secure multi-party computation and differential privacy are listed among the proposed methodologies and approaches in the UN PET guide (UN, 2023).

The participation of private companies specialised in mobile phone data within the UN Big Data initiative was also pointed out by an IBGE officer: 'they discovered this market niche […] and why are they there at the UN? Because they end up teaching others to use it, what is their interest? They will sell consultancy' (Interview, 25 May 2022). Conversely, the same officer points to the benefits for NSOs partnering with private third parties:

> These companies have a deep knowledge of data, how to process this data, they have experience, we don't. We want to use this data and they can teach us [..] they have the software ready […] Of course, I think there will come a time when we will be self-sufficient in this, but we are far from it. (Interview, 25 May 2022)

According to a senior manager of the UNSD, the experts from the private sector at the UN Global Platform work 'without any financial incentives', apart the knowledge that is taken 'back home to their work'. Allegedly, this doesn't include 'products that they would be able to

commercially exploit'. He nevertheless recognises the commercial interests of private corporations and raises concerns with certain businesses' involvement in the initiative:

> when we work in this group of experts on these technologies [PETs] there is in-house knowledge, there is also academic knowledge, but there is still a technology company that puts together some of the applications that are needed to work with differential privacy or with multiparty computation […] so as a service provider, there might be, commercial interest involved of this private sector company […] we are critical in the sense if we think that the experts from the private sector have certain kinds of commercial benefits in mind then we'll double check that, it is possible because in the area of the privacy enhancing technologies, we are a little bit careful with Facebook or Google working with us, but we do work with smaller kind of private sector companies, we have like Samsung research working with us. So, all of them work with us and if they work with us, they do it without any commercial benefit. I mean if you count that we develop some kind of knowledge in general, so they take that back home to their work. So that's their benefit, but we don't develop any products that they would be able to commercially exploit. (Interview, 12 Sep. 2022)

The risk of competition with the private sector was quoted in some of the interviews with IBGE members. 'Trust' and 'public good' came out in some speeches as important differentials of official statistics amidst a trend of data commodification. The following quote is illustrative:

> The statistical offices must be careful, it's a competition. If Google decides to compete with the IBGE, we have no chance […] we have to take care of what we do best, which is trust and sustainability, because Google may be interested in that today and tomorrow it won't be […] We are good at making statistics, we guarantee the sustainability of this, we have immense historical series, we are doing this because statistics are a public good, we are not interested in making money, nor in the market. But we have to recognise that more and more information is a commodity […] and people are going to be interested in it for the simple market value and not […] for the good it can generate. (Interview, 25 May 2022)

To conclude, this case illustrates how transnational initiatives involving IOs and Big Techs have been promoting the use of big data in Latin American NSOs. In Brazil, this has led to struggles that comprise a double movement. On the one hand, the UN Global Platform approach induces cooperation with the private sector within a platformisation framework for official statistics in the region, and on the other, within IBGE, in-house independent development of

statistics with new sources of data and demands for a new legislation indicates protective counter-movements aiming to secure free data access and NSO's autonomy. Both movements aim to control informational capital, which they seek through the mobilisation of symbolic capitals – 'trust' and 'public good'. In the following case, we gather further evidence on the meaning of Global North-induced experiments in Latin America and the kind of resistance they have been encountering.

Case study 2: Data-Pop Alliance and the OPAL Project at DANE (Colombia)

The US not-for-profit Data-Pop Alliance (DPA) is a GPSDD partner that promotes the use of big data for official statistics in Latin America and has developed a partnership with the National Administrative Department of Statistics of Colombia (DANE).

DPA's strategy for Latin America expressly advocates that NSOs in the region should 'engage with the private sector, evaluate current models for corporate data sharing and set up agreements for Public–Private Partnerships' (DPA, 2016: 39). For DPA, governments in Latin America have the potential to play a leading role within the national big data ecosystems, 'becoming, on the one hand, a facilitator and, on the other, a consumer of […] big data' (DPA, 2021: 3). Accordingly, the DPA's recommendations to Colombia for enabling data markets in the country backed the removal of administrative and legal barriers: 'In terms of public investment, best practices to encourage the big data value chain focus on deregulating the ecosystem' (DPA, 2019: 22).

The Open Algorithms Project (OPAL), implemented by the DPA, consisted of a PPP with DANE and Telefónica Colombia to identify use cases of Call Detail Record (CDR) data. For DPA, OPAL is an example of action 'essential to boost the data market in the country' (DPA, 2019: 22). In contrast, according to GPSDD, OPAL is a 'not-for-profit sociotechnological innovation' and 'a new paradigm for using private data for social good' (GPSDD, 2018: 1). In technical terms, OPAL consists in a PET and data-sharing method in which algorithms run on pseudonymised data that remain on the servers of the partner company and only provide aggregated statistics for selected users.

For the implementation of the pilot project, OPAL received funding from GPSDD and the World Bank (GPSDD, 2018). Subsequently, OPAL raised 1.5 million euros from the French Development Agency (AFD, 2018). In 2020, the project escalated to a market test in partnership with Flowminder. Now, OPAL focuses 'on low and middle-income countries, starting with Haiti in collaboration with Digicel' (OPAL, 2022: 1).

In the OPAL model (Hardjono and Pentland, 2019), the NSO is conceived as a user with access to certain

aggregated statistics performed by algorithms developed by an intermediary in a company's database. As a consumer, the NSO's degree of autonomy in statistical production is reduced. Since it does not have access to the raw database but only to aggregate data, there are limitations to the statistical process.

According to a DPA research report, the main obstacles encountered for 'adopting disruptive technologies such as big data' in Latin America are the resistance of the public officers and the lack of public–private alliances at the national government scale (DPA, 2021: 38). The DPA has found that most of the projects in the region are experimental and depend on IOs incentives. Therefore, it advocates the need for a 'cultural change', alleging that:

> The still ingrained belief that they [analogue data] are more reliable is erected as a barrier to boost the use of digitised information as a valuable source in the production of official statistics […] there is a distrust of government data sharing that makes it difficult to understand the value that big data can bring with it […] Misconceptions about the meaning and the potential of big data are, in this sense, a great challenge for the consolidation of a potential big data ecosystem. (DPA, 2021: 38–39)

In summary, this case has shown how Global North-induced experiments of big data for official statistics have sought to build new business models for data markets in Latin America. In this case by developing a PET for data sharing, brokering a PPP, influencing a national government and mobilising symbolic capitals such as 'social good' which led to the development of a commercial product for an emerging data market for official statistics. It also revealed alleged resistances to big data among public officials and limitations to PPPs at national levels. In the following case, we investigate the role of big techs and IOs in sponsoring and disseminating such experiments and evidence of emerging counter-movements.

Case study 3: ECLAC – DATAPROVIDER (Brazil, Mexico, Chile and Colombia)

Between 2016 and 2021, the Economic Commission for Latin America and the Caribbean (ECLAC) implemented the project 'Big Data for measuring and fostering the digital economy' (ECLAC, 2021). The initiative involved NSOs from Brazil (IBGE), Mexico (INEGI), Chile (INE) and Colombia (DANE). Among other actions, the project tried to replicate the e-commerce measurement methodology adopted by Statistics Netherlands (CBS) through a Google-sponsored PPP.

According to CBS: 'Google has approached Statistics Netherlands to carry out a study to deepen the understanding of the internet economy using an innovative approach'

(CBS, 2016: 3). In the Netherlands, the project consisted of combining data from the register of companies at CBS with data on the presence of businesses on the internet, collected by the company Dataprovider (CBS, 2016). The pilot project then escalated into an ongoing e-commerce measurement with CBS acquiring data from Dataprovider on a regular basis (CBS, 2020). In 2019, a collaboration agreement was signed, and Dataprovider received a 'seal of quality' from CBS. According to one of the CBS directors, 'market actors can expand and enrich the range of their products and services by working with us. Our network contributes to this' (CBS, 2019: 1). Indeed, CBS's network has enabled Dataprovider to expand its operations. Besides ECLAC, the company also signed an agreement with the South Korean NSO (CBS, 2019: 1). According to a Dataprovider director:

> The seal of quality granted by CBS means a lot to us […] we've been invited by the United Nations to deliver a keynote presentation at an assembly together with CBS, in front of more than 100 statistical directors from all over the world […] This fits well within the future plans of our company: to continue growing and collaborating with new actors in order to expand our range of data. (CBS, 2019: 1)

Conversely, the PPP strategy of CBS for accessing big data sources was questioned by a recent report commissioned by the Dutch Ministry of the Interior, which highlighted the regulatory route as an alternative to be considered for data access. According to the report, although CBS 'can buy data […] the path of regulation is open and it can lead to third-party obligations regarding data sharing' (Geonovum, 2021: 16).

Moving on to Latin America, the study coordinated by ECLAC followed the same methodology adopted by CBS. According to a statistician involved in the project: 'ECLAC paid for the database from the web scraping company […] We delivered some keywords that we classified by category. It's as if we had almost translated what the Netherlands did' (Interview, 30 May 2022).

We discovered that the Latin American NSOs encountered difficulties in pairing the databases. In the case of Brazil, it was possible to identify only 9.2% of the companies in the Dataprovider database (IBGE, 2020). According to interviews, in the other countries, the percentages were even lower. Constraints were also identified in the Netherlands. A CBS report warned of 'substantial instability and problems [of consistency] arising from the combination of variability in the big data and the method used to delineate the internet economy and its categories' (CBS, 2020: 5).

Nonetheless, the local learning generated by the project resulted in a new initiative developed autonomously by the Brazilian NSO. According to one of the statisticians: 'due to

this development, there is now a project at the IBGE to deal with this issue, but we are collecting the information on our own, we are not using Dataprovider or anything like that' (Interview, 17 May 2022).

To sum up, this regional case brought further evidence of how big techs and IOs have supported and disseminated pro-market business models in big data for official statistics. It also presented another example of how 'trust' has been mobilised by NSOs for partnering with businesses. Finally, the case highlights counter-movements that were found both in Europe (recommendations on legal alternatives for data access) and Latin American NSOs (replication of the initiative without business partners). In the following section, we discuss the main findings of the case studies considering theory and historical context.

Evidence and discussion

In its etymology, statistics means 'science of the state' and its historical origins relate back to the processes of the constitution of the states themselves. The birth of the territorial states is inseparable from an immense accumulation of informational capital, provided mainly by statistics. From the monarch's administration secrets, the official statistics of the nation-states became a public good controlled by a statistical field. This process followed the qualitative transformation of a series of private capital into public capitals that characterised the relative autonomisation of modern bureaucracy. The nation-state has ever since been associated with a 'total' knowledge of the social world, which is provided precisely by statistics (Bourdieu, 2014). This status and its associated symbolic power have been challenged in the present by new forms of knowledge and action controlled by transnational tech corporations.

From the 19th century on, the implementation of decennial censuses and the regular production of official statistics by the national bureaux of statistics corresponded to an effort to forge a social unity space and a national identity within the nation-states. This encompassed the construction of 'national spaces of equivalence conventions' through a series of economic and cognitive investments (Desrosières, 1998). Although Europe was the intellectual matrix of this process, statistics were also crucial for the new post-colonial Latin American states to justify their existence as independent nations (Otero, 2018).

The construction of national statistical systems and institutions by sovereign nation-states was followed by efforts of international coordination.[8] The active participation of Latin American countries in such initiatives since the 19th century shows that networking in a transnational field of statistics and adapting international methods are features that structured the historical construction of national statistical systems in the region (Otero, 2018). This remains to the present, as we have seen in the case of the UN Big Data.

The production of statistics on a national scale, from very early on, imposed the challenge of dealing with large datasets on the NSOs. Not by chance, the first mechanic computer used for data processing that marks the birth of IBM was developed for the US Census of 1880 (Anderson, 2015) and was later introduced in Latin America for the Brazilian Census of 1920 (Segura, 2017). Innovations have historically been necessary for NSOs to deal with large volumes of data in a context of increasing demands for information and were often accompanied by controversy.

The introduction of probabilistic samples and electronic computers in national statistics in the middle of the 20th century were innovations that reshaped the statistical field. Until then, NSOs were made up of public administrators guided by the ideal of exhaustiveness of the census with limited enumeration and tabulation tools. Introducing inferential statistics and complex data processing represented not only methodological and technological innovations but a complete change in the professional profile of the field that now required the mastery of advanced mathematical statistics and computer science (Desrosières, 2008).

The statistical practice was developed ever since as a field of tensions characterised by the conflicting interface between a bureaucratic and a scientific habitus (Bourdieu, 1990),[9] on the one hand, and on the other, by the pressures that emerge from the instrumental articulation of a governmentality that is both national and transnational (Foucault, 2007). If data science and big data come to represent another turn in official statistics, this is still to be determined. According to Grommé et al. (2021: 239), introducing big data to European NSOs has triggered struggles and competition between national statisticians and an emerging faction of data scientists 'over the relative valuation of cultural capital and habitus required to work with big data'. We understand the contemporary turn within the political economy of data, from public goods to commodities, as a structural factor of such disputes, which are also emerging in the Latin American context, as demonstrated in the case studies.

A relevant aspect of understanding NSOs is the social construction of trust in official statistics. Reliance on unemployment rates, GDP, inflation index etc. is a condition for public debate and decision-making in democratic societies. Conversely, the willingness of subjects to provide their data to statistical agencies also depends on the confidence in the confidentiality and the public purpose of the information provided. Therefore, trust is a recurring subject in the official statistics literature (Lehtonen, 2019). For Radermacher (2020: 140), 'trust is the main and overarching goal of statistical governance' and for Kitchin (2015: 477), it is the NSO's 'number one priority'.

We understand trust as a symbolic capital vital to ensuring the state's control of informational capital and the effectiveness of statistics as a technology of government.

According to Bourdieu (2014), to discern the public from the private, 'the state must theatricalize the official and the universal, it must put on the spectacle of public respect for [...] the official truths in which the totality of society is supposed to recognise itself' (Bourdieu 2014: 28). In the statistical field, this theatricalisation is accentuated to the extent that trust in statistics relies on the perception of its public value and accuracy, which is reinforced by a conviction in the objectivity of numbers (Porter, 1995).

Constant care with the preservation of trust in the NSOs is reflected, among others, in the search for autonomy, in the methodological rigour, in the quality and stability of the data, in the publicity of the results and in the permanent risks assessment (Radermacher, 2020). Reconciling innovation and trust implies an inherent tension within the statistical field. This is a relevant element to understanding the disputes at stake with the advent of new methods and data sources.

In the 21st century, data began to be generated through digital interactions in unprecedented volumes, varieties and velocities, constituting what are conventionally called big data (Kitchin, 2014). A distinctive feature of the emerging data economy is the effort to extract and enclose data from users to convert it into digital commodities (Rotta and Paraná, 2022) to obtain knowledge rents (Rotta and Teixeira, 2018).

Analysts have pointed to asymmetries in the data economy, especially in the AI industry (Dyer-Witheford et al., 2019). According to Rikap (2021), companies such as Google, Amazon, Facebook, Apple and Microsoft act as intellectual monopolies using their disproportionate power to extract knowledge rents. For Couldry and Mejias (2019), these asymmetries have led to new forms of geopolitical power, in which data is being used by Global North tech companies and governments to maintain dominance and influence over developing countries, configuring contemporary forms of colonialism. One of the aspects of this process is the advance of big techs in several traditionally state-owned sectors, which results in even more data extraction from users, enhancing their power at a global level. In the statistical field, this process takes place through the big data agenda.

With the rapid digitalisation of social processes, NSOs have been pressured to modernise. Discourses on this new set of tensions emphasise budget constraints and demands for more timely information (Radermacher, 2020: 49); the burden of surveys on informants and new phenomena difficult to capture with traditional sources (MacFeely, 2019: 32–38). Allegedly, these pressures also involve competition with big techs and other players (Struijs et al., 2014: 2). As raised by a UN chief statistician: 'whether NSOs like it or not, they're competing with lots of other people who are producing estimates [...] they're lagging behind the private sector [...] Google are producing stuff, Facebook are producing stuff, why can't the statistical offices do this?' (Interview, 19 May 2022).

For Ruppert and Scheel (2021: 2), 'these pressures and calls have prompted numerous experiments with sources of big data'. As we have found in the case of the UN Global Platform, the punch line 'if you can't beat them join them' somewhat sums up how the alleged competition with the private sector was mobilised to present big data as a potential solution for 'modernising' official statistics by partnering with business. Similar arguments could also be found in a Eurostat report prepared for the Conference of European Statisticians in 2013:

> The private sector may take advantage of the big data era and produce more and more statistics that attempt to beat official statistics [...] It is unlikely that NSOs will lose the 'official statistics' trademark, but they could slowly lose their reputation and relevance unless they get on board. (UNECE, 2013: 2)

Likewise, opportunities to compile official statistics from big data were also raised by technical literature. Kitchin (2015: 472) highlighted that the new sources could 'complement, replace, improve, and add to existing datasets and refine existing statistical composition.' Struijs et al. (2014: 2), advocated 'a huge potential for new statistics', such as the use of mobile phone data for population and tourism statistics, social media for consumer indicators and web prices for inflation rates.

Despite the alleged potential, according to MacFeely (2019), the Big Data Project Inventory of the UN Statistics Division indicates that the use of Big Data in official statistics is still very limited around the world. Although the inventory comprises 109 initiatives from 34 NSOs, 'several projects are speculative or aspirational, where the big data source has not yet been identified or where access to the data (particularly mobile phone) has not yet been secure' (MacFeely, 2019: 31). Similar to our findings, MacFeely (2019) highlights that price statistics using web data are among the most frequent projects, as these approaches 'typically have fewer data access problems'.

Concerning Latin America and the Caribbean (LAC), a recent survey (HubBrazil, 2022) with sixteen NSOs revealed that there are seven countries already using big data in official statistics in the region, four countries in a testing or experimental phase only and three countries planning tests (see Figure 2).[10] Regarding official statistics, the main sources of big data under use are satellite imagery (33%), web scraping (27%) and energy meters (20%). Big data sources from public or open sources (satellite images, web data) or from public services concessionaires (energy meters and health records) are among the most used (87%), while sources held by private companies (credit card and scanner data) represent only 13%. Other privately held data such as mobile phones and social media are under tests or in an experimental phase only and have not yet been used in official statistics (see Figure 3).
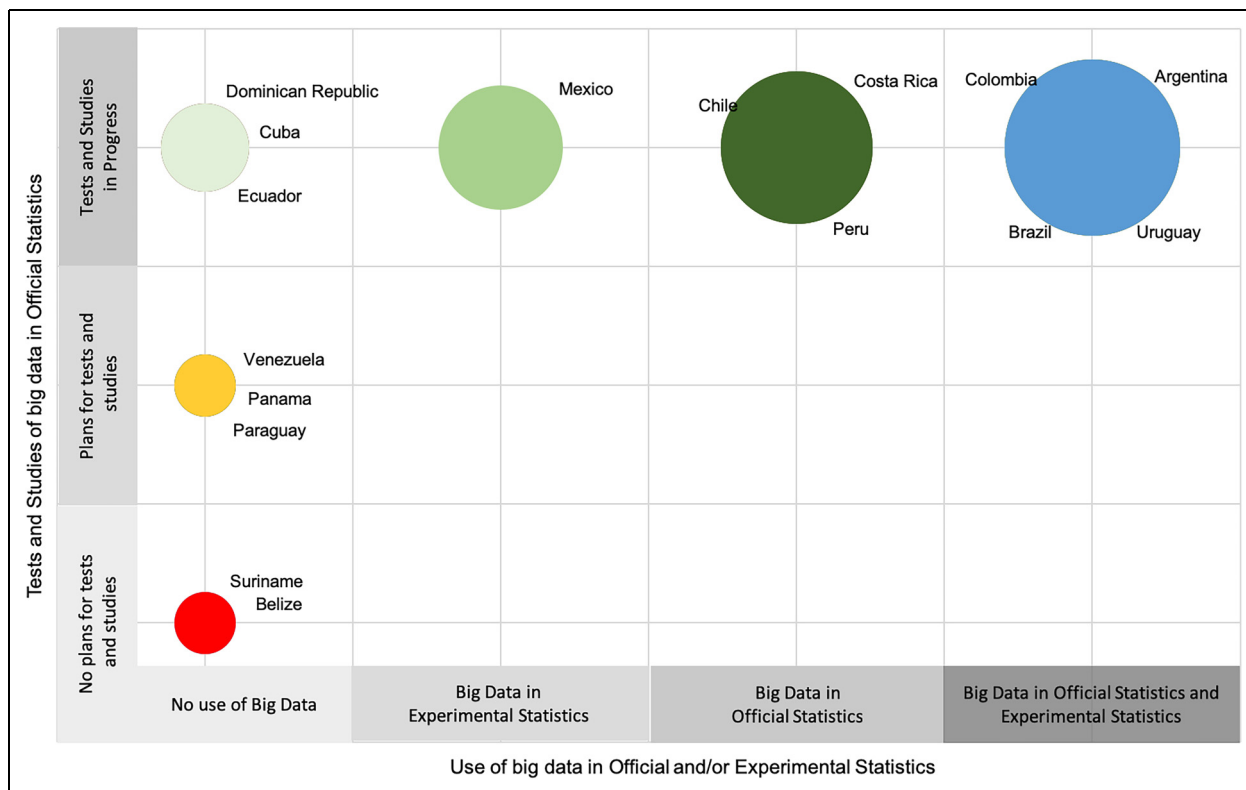
Figure 2. Latin American and Caribbean (LAC) countries clusters according to tests and use of big data in official statistics.
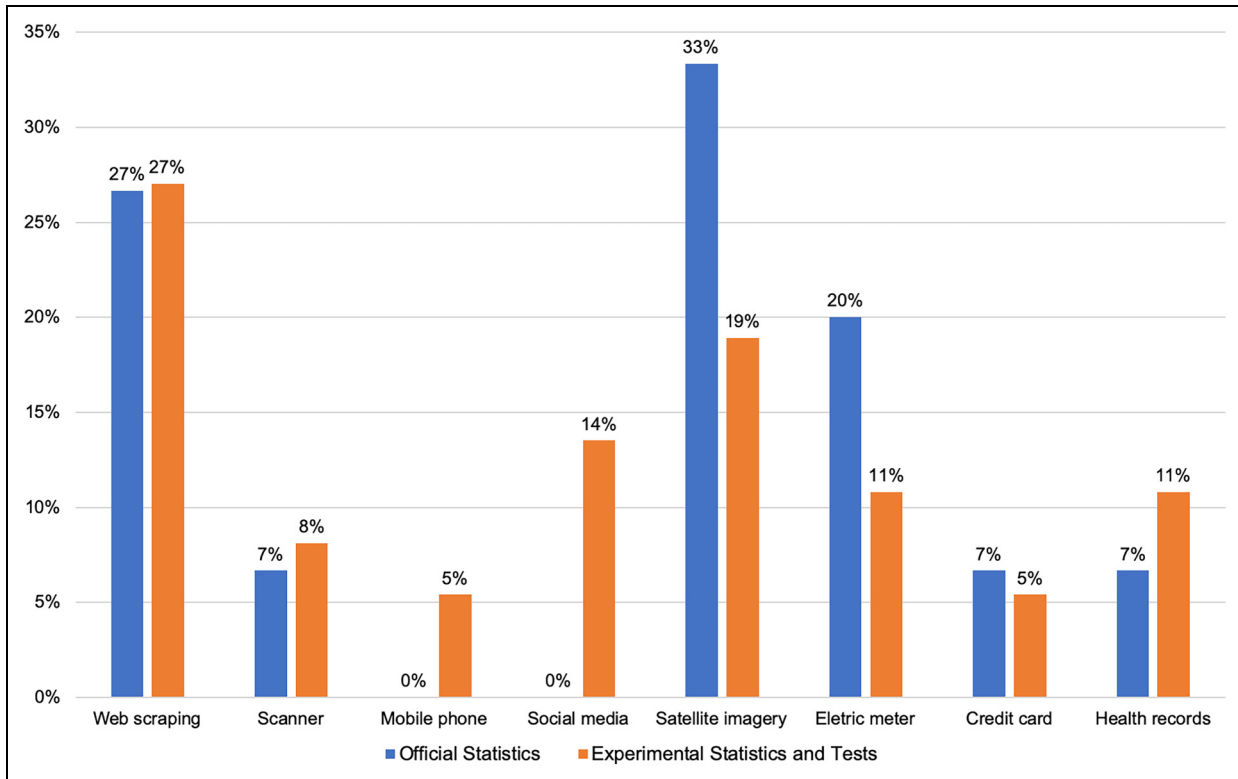*Source:* HubBrazil (2022). Authors' own computation.

The secondary data on the use of big data for official statistics corroborates our findings in the Brazilian case and shows that the lack of access to privately held data has prevented the use of these sources for official statistics. These findings are supported by other research. For MacFeely (2019), 'one of the biggest barriers to using big data is lack of access. Many big data are proprietary i.e., data that are commercially or privately-owned'. According to a recent report from Eurostat (2022: 13), the NSOs' lack of access to privately held data indicates the failure of the voluntary PPP approach, as the results of compiling official statistics from big data 'so far have been seriously limited in terms of (i) statistical domains and statistical output covered, and (ii) integration of new data sources in regular statistics' (Eurostat, 2022: 16).

The difficulty for NSOs to access privately held data is evidence of the contemporary process of data enclosure[11] which stands as an element of a broader dispute between nation-states and private corporations for the control of informational capital. In line with Verdegem (2022) and Dean (2016), we understand data enclosure as the process by which free access and control over the user activity information generated by digital interactions are kept away from users themselves and the public for the benefit of the service providers (platforms and tech companies) where the data is generated. By this, open or shared arrangements of access and control over data are made proprietary and exclusive. Enclosed, big data constitutes informational capital under dispute and a digital commodity in potential.

Therefore, while those in the statistical field seek access to data to maintain their positions as relevant producers of statistical information, for the private sector the data enclosures are necessary for enabling new data markets. The case studies have shown the quest for new business models on big data for official statistics that are still under development, namely: (i) PPPs (cases 1, 2 and 3); (ii) platformisation frameworks (case 1); (ii) development of new technologies (PETs), commercial products and services for data sharing (cases 1 and 2); (iii) sale of data to NSOs and IOs (case 3); (v) sale of services to NSOs (cases 1 and 2).

The case studies have shown how the trust in official statistics has been mobilised by the statistical field for partnering with businesses, through titles such as 'trusted partner' (case 1) and 'seal of quality' (case 3), in which the symbolic capital of trust is granted to private partners. In this institutional form, trust can then be converted into informational capital for NSOs (access to data) and economic capital for the private sector (profits in the data market). Other forms of symbolic capital were also mobilised to justify partnerships with companies through alleged not-for-profit initiatives, namely 'data for social good' and 'data for development' (cases 1 and 2).

Figure 3. Big data for official statistics in Latin America-sources
Source: HubBrazil (2022). Authors' own computation.

The three case studies have Big Tech's (Microsoft and Google) direct or indirect sponsorship and involvement. As we have found in the first case study, it was alleged that direct partnerships with certain companies were avoided, indicating concerns of public distrust. According to Kitchin (2015: 11), the loss of public trust is one of the main risks in the use of big data for official statistics, as 'partnering with a commercial third party and using their data […] exposes the reputation of an NSO to that of the partner'. This risk was also borne out by a UN chief statistician: 'If the public finds out that the NSO is using these new types of data and they're not comfortable with it, then that could end up causing them a lot of trouble' (Interview, 19 May 2022).

To illustrate, surveys have shown that people tend to trust national public authorities more than private corporations with regard to the use and protection of personal data (EC, 2015; Latinobarómetro, 2020). Additionally, scandals involving the abuse of data, such as the Facebook-Cambridge Analytica case, have contributed to boosting public suspicions. In 2022, 36% of users affirmed they do not trust the internet, the worst historical index measured by IPSOS (2022), and 79% expressed concerns about privacy and protection of personal data. More recently, the open campaigning of big techs against platform regulation in countries like Brazil tends to deepen distrust in these companies in Latin America (Boadle, 2023).

In sum, trust in official statistics is a symbolic capital in dispute amidst the social construction of new data markets. For the private sector, official statistics represent a data market niche, which, in addition to profits, can add reputational value and credibility. For the statistical field, mobilising trust for partnering with the private sector comprises a risky strategy for accessing big data, as the association with certain businesses may justify public distrust. The PPPs in official statistics may also jeopardise NSOs' autonomy and national sovereignty over public data, especially in the global-south, in the face of asymmetric relations of power with global tech corporations and the absence of proper regulation. So far, although partnerships with businesses have allowed for NSO's initial testing of new data sources, they have not proven to be an effective path for assuring the sound access to data necessary for compiling official statistics at the national level.

The case studies have also evidenced protective counter-movements following a logic-oriented by the public value of data for official statistics and NSO's relative autonomy. Examples are the advocacy for a new statistical law in Brazil (case 1), the independent development of NSOs' in-house knowledge to use new sources of data and technologies without private parties' involvement (cases 1 and 3), and the alleged resistance to 'disruptive' big data by public officers in LAC (case 2). Protective counter-

movements could also be found in Europe, such as the official recommendations for CBS to look for law enforcement alternatives for accessing data (case 3).

We conclude that the opposing logics of commodification and public interest concerning the use of big data for official statistics leads to a double movement (Polanyi, 2001) within the statistical field (Table 1). On the one hand NSOs and IOs cooperate with businesses to get access to new sources of data and technologies, capitals at stake in the field, and on the other hand, they work to defend the public value of data and the autonomy of the statistical institutions accordingly to a distinctive bureaucratic logic of 'interest in the general interest'[12] (Bourdieu, 2014) that assure symbolic profits and control over informational capital.

The Polanyian concept of double movement that we instrumentalise in the analysis states that modern society is governed by two opposing principles. The movement to expand the extension of the self-regulating market – or to disembed the economy from society, is often met by a resisting social protective counter-movement 'aiming at the conservation of man and nature as well as productive organization' (Polanyi, 2001: 138). The state – also a market participant and architect – takes sides at both poles of this double movement. It is this 'contradictory position' that sustains the relative autonomy of the state in the face of immediate economic interest.[13]

Back to the actual context, the COVID-19 pandemic marked a political shift as it has demonstrated the potential of new data sources for relevant public statistics – such as mobile phone data for social isolation statistics – and exposed the governments' lack of access to data enclosed by private companies (Biancotti et al., 2021). Therefore, in Europe, amid consultations regarding the new EU Data Act, the European Statistical System expressly defended a new law for compulsory access to privately held data (ESS, 2022). In contrast, private corporations (e.g. IBM, Telefonica, Vodafone and Fujitsu) and associations of business (e.g., U.S. Chamber of Commerce) sent objections to the EU, against any approach to data access based in mandatory requirements and in defence of a voluntary partnership model (EC, 2021). Finally, the Data Act included a legal mechanism granting governments compulsory access to privately held data in public emergencies and exceptional circumstances (EC, 2022).

New forms of data regulation may characterise a broader double movement coming from states: on the one hand, meeting society's demands for protecting the public value of data and, on the other, building a framework for enabling regulated data markets. In the case of the EU Data Act, a developing mechanism for payments to compensate companies for transaction costs in data sharing operations seems to point in this direction. Based on the Brazilian case, where segments of the NSO are already advocating for new legislation, and considering the influence in the

region of other European regulatory models such as the GDPR, it is likely that the debate over the access to privately held data for official statistics will soon affect Latin America. In the UN environment as well, new initiatives for protecting public data are also emerging. According to one of the UN's chief statisticians:

> We're looking at this idea of a Global Data Compact and one of the issues that we're grappling with is not all data should be public goods, but some data clearly should be. So how to protect the public good part of the data to make sure that it's not privatised or sucked in and become private. (Interview with a UN chief statistician, 19 May 2022)

We conclude that there is a set of ongoing transformations in the statistical field with the emergence of new data sources, practices and agents in the context of a new data-driven economy. As demonstrated, Latin America has been a privileged stage for Global North-induced experiments in big data for official statistics, which reveals a quest for new business models to enable data markets in the region. The top-down geopolitical framework of these initiatives and the asymmetries between the agents may characterise renewed forms of data colonialism (Couldry and Mejias, 2019). Nevertheless, the Latin American NSOs investigated have strong capacities and have actively and creatively reacted to such incidences, featuring a double movement. The conflicting logics of data as public goods and as commodities constitute a fundamental element of the new tensions that arise within the statistical field, which relates to a broader disputation between nation-states and tech corporations. The developments of such disputes will have important consequences for sovereignty over data, especially in the Global South where such experiments have been carried out.

Final remarks

In this article, we have investigated the introduction of big data for official statistics in Latin America through three case studies: (i) the regional hub of the UN Global Platform in Brazil; (ii) a data-sharing project in Colombia; and (ii) a regional project of ECLAC involving NSOs of four countries (Brazil, Chile, Colombia and Mexico).

We have found that data enclosures have blocked the access of NSOs to privately held sources of big data, preventing its use for official statistics. In this contested terrain, segments of the transnational field of statistics have encouraged NSOs' partnerships with businesses to 'modernise' official statistics. These initiatives, assessed in the case studies, have been testing and promoting new business models through PPPs in big data for official statistics aiming to enable new data markets in Latin America.

We conclude that big data represents an element of a broader dispute between nation-states and private

corporations for informational capital accumulation and control. In the case of official statistics, the disputes involve the conflicting logics of data as public goods and data as commodities. The originality of our approach is to shed light on the political economy aspects of this process from a Latin American perspective. This made it possible to understand that while the region has been a stage for IOs and Global North tech corporation-induced experiments, the Latin American statistical field has actively reacted to such incidences featuring a double movement; on the one hand, cooperating with market agents and on the other defending the public value of data. Similar counter-movements were also found in Europe.

The case studies have shown how symbolic capitals such as 'trust' and 'social good' were mobilised for partnering with businesses. This has allowed private companies to test business models in the exchange of access to data, knowledge, and new technologies by the NSOs. Conversely, the public value of data for official statistics was defended in counter-movements aiming to maintain the relative autonomy of the NSOs and the control over informational capital. Finally, new legal mechanisms for governmental access to privately held data that are currently under development in the EU (Data Act) mark a new set of controversies that have recently emerged since the pandemic. The quest for a new statistical law, as appears to be the case in Brazil, may lead to new developments in Latin America that require further investigation.

Declaration of conflicting interests

Funding

ORCID iD

Edemilson Paraná https://orcid.org/0000-0003-3487-9677
Oscar Arruda d'Alva https://orcid.org/0000-0002-6979-6823

Notes

1. Direct quotes from interviewees are anonymized and referenced in parentheses with the indications 'interview' and date of completion.
2. According to Bourdieu (2014), the state 'has gradually established itself as a kind of meta-field of a field where a capital is produced, preserved and reproduced that gives power over the other species of capital' (Bourdieu, 2014: 198). In this sense, the stake in the state is the determination of the position that the different fields should legitimately occupy in relation to one another.
3. We performed a cluster analysis using the SPI score for 27 Latin American countries: Cluster 1 (four countries, SPI 81.63), Cluster 2 (eight countries, SPI 67.80), Cluster 3 (10 countries, SPI 55.98) and Cluster 4 (five countries, SPI 41.76). The graph shows a disaggregation of the SPI into two factors: statistical outputs (average scores of the pillars of uses, services and data products), and statistical inputs (average scores of the pillars of sources and data infrastructure). See attached database.
4. The UN Statistical Commission is the highest body for the coordination of international statistical activities.
5. The UN Foundation is a US not-for-profit that supports the UN and its activities (UN-Foundation, 2023).
6. The UN Big Data task teams gather NSOs, research and private companies addressing: (i) earth observation, (ii) mobile telephone data, (iii) scanner data, (iv) data from vessel tracking, (v) Big Data for the SDGs, and (vi) privacy preservation techniques.
7. Not to mention the use of geospatial data which were not listed here among 'new sources', considering IBGE's tradition on the research and use of this kind of data.
8. The quest for transnational statistical coordination began with the international statistical congresses in the 19th century, followed by an international statistical commission within the League of Nations (1920–1946) and consolidated at the United Nations Statistical Commission (1947–).
9. For Bourdieu (1990: 60), the 'habitus' could be regarded as a 'subjective but non-individual system of internalized structures, common schemes of perception, conception and action'. The habitus tends to generate behaviours which are likely to be positively sanctioned as they are adjusted to the logic characteristic of a particular field.
10. The label 'experimental statistics' has been adopted by NSOs to differentiate statistics that are in a testing phase from official statistics, compiled through consolidated data sources and methods.
11. Private property arises from appropriation that demands exclusion from access. Marx (1976: 879–885) argued that modern capitalism emerges as a result of massive processes of expropriation and the 'enclosure of the commons' that enabled the industrial revolution. This 'original form of accumulation' is actualised today in continuous dispossession that negates collective and public non-commodity forms of wealth – be it of land and natural resources, knowledge, data etc.
12. We follow here Bourdieu's (2014) conception of the public as a space where 'disinterestedness' is officially recognised and where to some extent agents have an interest in the disinterest.
13. See also Offe (1974) and Poulantzas (1978).

References

AFD (2018) Proyecto OPAL. Available at: www.afd.fr/es/actualites/proyecto-opal-el-big-data-al-servicio-del-desarrollo (accessed 13 February 2023).

Allin P (2021) Opportunities and challenges for official statistics in a digital society. *Contemporary Social Science* 16(2): 156–169.

Anderson M (2015) *The American Census: A Social History*. New Haven: Yale University Press.

Beaud JP (2015) Sociopolitique de la statistique. In: Prud'homme J, et al. (ed.) *Sciences, Technologies et sociétés de A à Z.* Québec: Les Presses de l'Université de Montréal, pp.216–219.

Biancotti C, et al. (2021) COVID-19 and official statistics: a wakeup call? Questioni di Economia e Finanza. Banca D'Italia Eurosistema.

Boadle A (2023) Brazil pushes back on big tech firms' campaign against 'fake news law'. Reuters, 2 May.

Bourdieu P (1986) The forms of capital. In: Richardson J (ed.) *Handbook of Theory and Research for the Sociology of Education.* Westport: Greenwood, pp.241–258.

Bourdieu P (1990) *The Logic of Practices.* Stanford: Stanford University Press.

Bourdieu P (1998) *Practical Reason. On the Theory of Action.* Stanford: Stanford University Press.

Bourdieu P (2014) *On the State.* Cambridge: Polity Press.

Bourdieu P and Wacquant L (1992) *An Invitation to Reflexive Sociology.* Cambridge: Polity Press.

CBS (2016) Measuring the internet economy in the Netherlands. CBS Discussion Paper, 14. Statistics Netherlands.

CBS (2019) CBS and dataprovider successfully map internet economy. Available at: www.cbs.nl/en-gb/corporate/2019/20/cbs-and-dataprovider-successfully-map-internet-economy (accessed 03 February 2023).

CBS (2020) Measuring the internet economy in the Netherlands 2016–2018: A big data analysis. CBS Discussion Paper. Statistics Netherlands.

Couldry N and Mejias U (2019) *The Costs of Connection.* Stanford: Stanford Univ. Press.

Dargent E, et al. (2018) *Who wants to know? The Political Economy of Statistical Capacity in Latin America.* New York: Interamerican Development Bank.

Dean J (2016) Big data: accumulation and enclosure. *Theory & Event* 19(3). https://www.muse.jhu.edu/article/623988.

Desrosières A (1998) *The Politics of Large Number: a History of Statistical Reasoning.* Cambridge: Harvard University Press.

Desrosières A (2008) *L'Argument Statistique II: Gouverner par les nombres.* Paris: Presses de l'Ecole des mines.

DPA (2016) Opportunities and requirements for leveraging big data for official statistics and the SDGs in Latin America. White Paper Series. May 2016. Data-Pop Alliance.

DPA (2019) Propuesta de plan de implementación de la estrategia de Big Data para el Estado. Report no.5. July 2019. Data-Pop Alliance.

DPA (2021) Estudio exploratorio regional en big data para el desarrollo sostenible en Latinoamérica y el Caribe. Report to AECID. Madrid: AECID.

Dyer-Witheford N, Kjosen A and Steinhoff J (2019) *Inhuman Power: Artificial Intelligence and the Future of Capitalism.* London: Pluto Press.

EC (2015) *Eurobarometer 83.1: Data Protection and the Internet, February–March 2015.* Ann Arbor, MI: GESIS.

EC (2021) Public consultation on the data act. Available at: https://digital-strategy.ec.europa.eu/en/consultations/public-consultation-data-act (accessed 13 February 2023).

EC (2022) Data act. Available at: ec.europa.eu/commission/presscorner/detail/en/ip_22_1113 (accessed 13 July 2023).

ECLAC (2010) Development of official statistics in the region. Report of the Economic Commission for Latin America and the Caribbean. Santiago: ECLAC.

ECLAC (2021) *Final assessment Report Project 16/17 Y.* Santiago: ECLAC.

Espeland W and Menniken A (2019) What's new with numbers? Sociological approaches to the study of quantification. *Annual Review of Sociology* 45: 223–245.

ESS (2022) European statistical system position paper on the future data act proposal. Available at: https://ec.europa.eu/eurostat/documents/13019146/13405116/main + ESS + position + paper + on + future + Data + Act + proposal.pdf (accessed 03 February 2023).

Eurostat (2022) *Empowering Society by Reusing Privately Held Data for Official Statistics.* Luxembourg: Publications Office EU.

Foucault M (2007) *Security, Territory, Population: Lectures at the College de France 1977–1978.* New York: Picador.

Geonovum (2021) Rapport verkenning publiek gebruik data van Derden. Report to the Ministry of Interior. 27 May. Amersfoort: Geonovum.

GPSDD (2018) OPAL case study. Available at: *static1.squarespace.com/static/599ef170197aeac586fed53f/t/5c182ccd8a922daa605547b1* (accessed 13 February 2023).

GPSDD (2020) GPSDD: five-years strategy. Available at: www.data4sdgs.org/news/our-five-year-strategy-2019-2023 (accessed 03 February 2023).

GPSDD (2023a) Funders. Available at: www.data4sdgs.org/about/funders (accessed 13 February 2023).

GPSDD (2023b) Our network. Available at: www.data4sdgs.org/our-network (accessed 13 February 2023).

Grommé F, et al. (2021) Statistician subjects: differentiating and defending. In: Ruppert E and Scheel S (eds) *Data Practices: Making up a European People.* London: Goldsmiths Press, pp.237–268.

Hardjono T and Pentland A (2019) *Trusted Data.* Cambridge: MIT Press.

HubBrazil (2022) Big data for statistics in Latin America and the Caribbean. Available at: //storymaps.arcgis.com/stories/3e42f28257004e4b98b0606518185656 (accessed 04 February 2023).

IBGE (2017) *Plano Estratégico 2017-2027.* Rio de Janeiro: IBGE.

IBGE (2020) Big Data para la medición de la economía de Internet. Available at: www.cepal.org/sites/default/files/presentations/bigdata-medicion-economia-internet-ibge-brasil-2020.pdf (accessed 03 February 2023).

IBGE (2021) Memorando de entendimento IBGE-ONU. Signed Document ID: 4CEC3A1E-652C-456D-9019-4AB051604611. 8 April. Rio de Janeiro: IBGE.

IPSOS (2022) Internet users' trust in the Internet has dropped significantly since 2019. Available at: www.ipsos.com/en/trust-in-the-internet-2022 (accessed 03 February 2023).

ITU (2021) Using mobile phone data for measuring SDG. Available at: www.itu.int/en/ITU-D/Statistics/Documents/bigdata/ITU_SDG_case_study.pdf (accessed 16 February 2023).

Kitchin R (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures & their Consequences.* Los Angeles: Sage Publications Ltd.

Kitchin R (2015) The opportunities, challenges and risks of big data for official statistics. *Statistical Journal of the IAOS* 31: 471–481.

Lanata-Briones C, et al. (2022) *Socio-political Histories of Latin American Statistics.* Cham: Palgrave Macmillan.

Latinobarómetro (2020) Latinobarómetro informe 2020. Available at: www.latinobarometro.org/latContents.jsp (accessed 03 February 2023).

Lehtonen M (2019) The multiple faces of trust in statistics and indicators: A case for healthy mistrust and distrust. *Statistical Journal of the IAOS.* 35: 539–548.

MacFeely S (2019) Big data and official statistics. In: Strydom S and Strydom M (eds) *Big Data Governance and Perspectives in Knowledge Management.* Hershey, PA: IGI Global, pp.25–54.

Marx K (1976) *Capital: A Critique of Political Economy, Volume One.* London: Penguin Books.

Offe C (1974) Structural problems of the capitalist state. Class rule and the political system. On the selectiveness of political institutions. *German Political Studies* 1: 31–57.

OPAL (2022) Opal history. Available at: https://www.opalproject.org/history (accessed 03 February 2023).

Otero H (2018) Socio-history of statistics on Latin America: a review. *Histoire & Mesure* XXXIII(2): 13–32.

Pasquinelli M and Joler V (2021) The Nooscope manifested: AI as instrument of knowledge extractivism. *AI & Society* 36: 1263–1280.

Polanyi K (2001) *The Great Transformation: The Political and Economic Origins of Our Time.* Boston: Beacon Press.

Porter T (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life.* Princeton, New Jersey: Princeton University Press.

Poulantzas N (1978) *State, Power and Socialism.* London: Verso.

Radermacher W (2020) *Official Statistics 4.0: Verified Facts for People in the 21st Century.* Cham: Springer Nature.

Rikap C (2021) *Capitalism, Power and Innovation: Intellectual Monopoly Capitalism Uncovered.* London: Routledge.

Rotta T and Paraná E (2022) Bitcoin as a digital commodity. *New Political Economy* 27(6): 1046–1061.

Rotta T and Teixeira R (2018) The commodification of knowledge and information. Greenwich Papers in Political Economy, Working Paper, No.60.

Ruppert E and Scheel S (2021) Introduction. In: Ruppert E and Scheel S (eds) *Data Practices: Making up a European People.* London: Goldsmiths Press, pp.1–28.

Segura M (2017) IBM no Brasil: Como tudo começou. Available at: https://www.maurosegura.com.br/ibm-no-brasil-como-tudo-comecou/ (accessed 16 June 2023).

Srnicek N (2016) *Platform Capitalism.* Cambridge: Polity Press.

Struijs P, Braaksma B and Daas P (2014) Official statistics and Big Data. *Big Data & Society* 1(1): 1–6.

UN-Foundation (2023) Our board. Available at: unfoundation.org/who-we-are/our-board/ (accessed 16 February 2023).

UN-StatCom (2017) Global platform for data, services and applications. Report to UN Statistical Commission. 48th session. Item 3(d)7-10 March. New York: UN.

UN-StatCom (2019) UN global platform business model options. Report to UN Statistical Commission. 50th session. Item 4(e) 5-8 March. New York: UN.

UN-StatCom (2021) Report of the global working group on big data for official statistics. Report to the UN Statistical Commission. 52nd session. Item 3(j)1-3 and 5 March. New York: UN.

UN-StatCom (2020b) Report of the GWG on big data for official statistics. Report to the UN Statistical Commission. 51st session. Item 3 (t)3-6 March. New York: UN.

UN-StatCom (2020a) Review of implementation of data revolution. Report to the UN Statistical Commission. 51st session. Item 3(a)3-6 March. New York: UN.

UN (2014) *A World that Counts: Mobilising the Data Revolution for Sustainable Development.* New York: United Nations Publications.

UN (2016) Public–private partnerships in use of Big Data for official statistics. Available at: https://www.un.org/en/desa/public-private-partnerships-use-big-data-official-statistics (accessed 03 February 2023).

UN (2023) The pet guide: The United Nations Guide on privacy-enhancing technologies for official statistics. UN Big Data. Statistics Division, UN Department of Economic and Social Affairs.

UNECE (2013) What does 'big data' mean for official statistics? Report for the UNECE. Conference of European Statisticians, 10 March. Geneva: UN.

Verdegem P (2022) Dismantling AI capitalism: the commons as an alternative to the power concentration of big tech. *AI & Society.* DOI: 10.1007/s00146-022-01437-8.

World Bank (2023) Statistical performance indicators. Available at: www.worldbank.org/en/programs/statistical-performance-indicators (accessed 17 February 2023).