

Demographic figures at risk in the digital era: Resisting commodification, reclaiming the common good

Big Data & Society
 October–December: 1–15
 © The Author(s) 2025
 Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/20539517251403948
journals.sagepub.com/home/bds



Edith Darin¹ 

Abstract

The digital era has transformed the production and governance of demographic figures, shifting it from a collective, state-led endeavour to one increasingly shaped by private actors and extractive technologies. This paper analyses the implications of these shifts by tracing the evolving status of demographic figures through the lens of Ostrom’s typology of goods: from a club good in royal censuses, to a public good under democratic governance, and now towards a private asset whose collection has become rivalrous and its dissemination excludable. Drawing on case studies involving satellite imagery, mobile phone data, and social media platforms, the study shows how new forms of passive data collection while providing previously unseen data opportunities, disrupt also traditional relationships between states and citizens, raise ethical and epistemic concerns, and challenge the legitimacy of national statistical institutes. In response, the paper advocates for the reconstitution of demographic figures as a common good, proposing a collective governance model that includes increased transparency, the sharing of anonymised aggregates, and the creation of a Public Demographic Data Library to support democratic accountability and technical robustness in demographic knowledge production.

Keywords

Demography, digital trace, census, democracy, public data library, commons

Well-informed individuals are citizens;
 poorly informed, they become subjects.

— Alfred Sauvy

Introduction

And here I am, almost by accident, becoming an expert in census-taking in conflict zones. It started in 2020, as Burkina Faso struggled to complete its national census due to escalating insecurity in the north-eastern region, when I was approached to “fill in the gaps”—not on the ground, but from my office, thousands of kilometres away. Drawing on my expertise in population modelling with satellite imagery at a British research institute, I was able, with the support of advanced statistics, pixels, and funding from the Bill and Melinda Gates Foundation, to estimate the populations of areas that had become inaccessible to surveyors of Burkina Faso’s National Statistics Office.

The estimates were based on non-open data provided by a Canadian private company, Ecopia.AI. Mission accomplished: I had left “no one behind,” as proclaimed by the Sustainable Development Goals. Yet a lingering doubt remained. Had I not undermined two foundational relationships inherent in census-taking: that between the enumerator and the enumerated, and that between the state and its citizens?

This question marks the starting point of the present article, which explores how the rise of digital data has reshaped the ways in which populations are enumerated. Far from representing a mere technical shift, the emergence of new sources of demographic information, such as satellite imagery, social media, mobile phone data, and administrative registers, is redefining not only the practices of data

¹Leverhulme Centre for Demographic Science, Nuffield Department of Population Health, University of Oxford, Oxford, UK

Corresponding author:

Edith Darin, 42-43 Park End Street, Oxford OX1 1JD, UK.
 Email: edith.darin@demography.ox.ac.uk



collection, but also the actors involved and the very purposes of population figures. These digital traces are characterised by their relatively low cost (they are initially gathered for other purposes), their passivity and non-reactivity (they do not require direct interaction), and their near-instantaneous availability. In a context marked by an increase in humanitarian crises, shrinking public budgets, rising non-response rates in traditional surveys, and a growing demand for faster and more granular figures, such sources are emerging as a major strategic opportunity.

However, this enthusiasm should not obscure the inherent limitations of these new data sources, particularly the coverage biases they entail, which result in the absence or underrepresentation of certain subgroups. Far from being marginal, these limitations underscore the need to develop specific modelling capacities, signalling a shift from the *enumeration* to the *estimation* of populations. This technological shift is also reshaping the power dynamics associated with the production of demographic figures, as illustrated by the experience outlined above. On the one hand, it tends to replace the intersubjective relationship established through field surveys with a distant, disembodied gaze that severs the direct link between enumeration and the communities concerned. This rupture follows a data extractivist logic, in which the digital traces mobilised are originally collected for operational or commercial purposes unrelated to the collective good. On the other hand, if this shift is not actively reclaimed, it risks generating neo-colonial dynamics by granting external expertise a substitutive power over national institutions, undermining their legitimacy, autonomy, and role in data governance. Given that demographic figures constitute a cornerstone of our democracies, it is high time to critically examine the ethical implications of producing population estimates outside the framework of censuses conducted by public statistical institutes.

The analysis based on examining secondary literature unfolds in four stages, guided by a reflection on the historical transformation of demographic figures status: from a “club good” in the era of royal censuses, to a “public good” under modern census regimes, and finally to a “private good” in the contemporary context. This trajectory underpins the central concern of the article: the need to establish demographic figures as a “common good” (Ostrom, 1990). The first part highlights the importance of demographic figures—population count broken down by age, gender, and geography—for both public and private actors, and reviews the growing diversity of data sources in the digital era that underpin their production. Here, I use conceptual analysis, drawing on the economics of public goods, to show how digital technologies have altered the very nature of demographic data inputs and the figures derived from them: data, once non-rivalrous, has become rivalrous in collection, and knowledge, once non-excludable, have become excludable in access, together

shifting demographic figures from a “public” to a “private” good. The second part examines the effects of restricting access to demographic figures, using a historical-comparative approach to examine to which extent operational and commercial logics echo the evolution of census practices from taxing subjects to representing citizens. The third part turns to the increasing rivalry among data sources and the need for a reconciliation through statistical modelling with insights from the sociology of science showing how the authority of national statistical institutes is being reconfigured and how discourses of “expert colonialism” emerge when new actors claim epistemic authority. Finally, the fourth part sets out a political and technical agenda for reconstituting demographic figures as a common good: *rivalrous* in data collection but *non-excludable* in knowledge dissemination. Here, I draw on commons and governance theory to outline possible institutional arrangements, such as data pooling, shared governance mechanisms, and ultimately the proposal for a public library of demographic data. Using demographic figures as a lens, this work revisits recent proposals on public data stewardship (United Nations Economic Commission for Europe, 2024) by (1) extending their scope beyond the public sector and (2) underscoring a structural shift: what was once a dedicated, curated data stream readily converted into population figures has become multiple, more biased—though often more granular—streams that now require innovative modelling to yield population figures.

Demographic figures: Between a public and a private good

The importance of population figures

Population figures have always been a central tool of the state, used to administer territory, collect taxes, organise conscription, and plan infrastructure. In the context of increasingly quantified societies, following the institution of the post-WWII welfare state, their uses have intensified and diversified significantly (Desrosières, 2008). It is no longer sufficient to count individuals; today’s figures must be precise, disaggregated, and up-to-date in order to capture social dynamics, mobility patterns, vulnerabilities, and spatial inequalities. Such figures play a crucial role in the design, implementation, and evaluation of public policies in areas such as health, education, urban planning, and social protection. For instance, in France, 300 to 400 administrative rules are population-dependent: a pharmacy may only open in municipalities with over 2500 inhabitants; the electoral system for municipal councils differs between communes with fewer than 1000 and those with more than 1000 residents (Barlet and Lefebvre, 2024).

Population figures have also become indispensable to a variety of non-state actors, researchers, NGOs, donors,

and businesses, who use them to plan interventions, direct investment, claim rights, or document exclusion. Inadequate population figures can produce a distorted picture of reality and lead to inappropriate responses; for example, in eastern Chad, mortality rates were underestimated fourfold between 2006 and 2010 due to an overestimation of the population (Bowden et al., 2012). The expanding need for robust demographic figures carries significant ethical implications. Populations that are absent or poorly represented are more vulnerable to political and social invisibility. Hence, the production of complete and inclusive population figures is not only a technical imperative but also a democratic one: it underpins a system's capacity to recognise its citizens, allocate resources equitably, and ensure accountability.

In response to the growing demand for population figures at finer temporal (Billari, 2022) and spatial (Tatem, 2017) scales, data collection is being undertaken by an increasingly diverse set of actors. In contexts where government institutions are stable and demographic change is gradual, data collection typically takes the form of a census conducted by national statistical offices. These large-scale operations, counting individuals one by one, require considerable human and financial resources and are difficult to undertake more than once per decade (United Nations, 2017). In situations marked by rapid demographic change, resulting from large-scale forced displacement caused by natural disasters or conflict, alternative data collection methods are employed such as registration data for displaced populations. Where displaced populations are less formally organised, sampling methods are used to estimate average occupancy. These techniques are often managed by the United Nations High Commissioner for Refugees (UNHCR), which becomes the de facto provider of crisis demographic figures. Broader estimations of population movements often rely on data collected from key informants, as exemplified by the Displacement Tracking Matrix developed by the International Organization for Migration (Abdelmagid and Checchi, 2018).

These various methods of population counting all share one feature: they involve a direct encounter between the enumerator and the individual being counted, even when this encounter is virtual, as in the case of online census questionnaires. However, the necessity of planning this interaction introduces an inherent time lag between the demographic consequences of an event and their statistical registration, and, by extension, between the event and the possibility of an organised response. This lag is largely overcome when data collection takes place digitally.

The digital revolution

Originally, the purpose of digital technologies, at least the computer as an automated processing tool, was precisely to manage the vast volumes of data generated by the U.S. census. It was in this context that IBM was founded, as a

technological extension of the Census Bureau (Ruggles and Magnuson, 2020). The advent of computing also facilitated the establishment of centralised population registers at the national level. Although population registers, defined as “a continuous recording mechanism of information pertaining to every member of the population, designed in such a way that at any given time one can accurately determine the size and characteristics of that population” (United Nations, 2014), were introduced as early as 1665 in Sweden, they were kept locally by individual parishes. It was with the rise of digital technologies in the 1960s that the development of national population registers truly accelerated. Today, eleven countries, primarily in Europe, are capable of producing monthly population counts exclusively from such registers¹. Fifteen other countries have established database integrating various administrative records that, while not exhaustive yet, can be combined with coverage surveys to generate population figures at shorter and more regular intervals².

Yet the impact of digital technologies extends far beyond improved technical conditions for recording population data. As Kashyap (2021) has highlighted, the digital revolution has introduced entirely new channels of demographic data, particularly through collection techniques that do not involve direct interaction with the individuals concerned. Two primary sources are relevant here: satellite imagery and digital traces. The former entails the analysis of remote sensing images to detect and quantify human presence, for example, through the observation of built environments, settlement patterns, or night-time light emissions. The latter is generated through individuals' digital activities, such as mobile phone use or interactions on social media platforms, and can provide insights into population mobility and density. While these data do not originate from traditional demographic survey mechanisms, they serve as proxies for demographic variables, enabling the inference of key features such as population size, spatial distribution, or age-structured migration behaviours. In this sense, they function as demographic data, not because they were collected with that purpose in mind, but because they can be repurposed to describe and model population dynamics. Moreover, they significantly reduce traditional data collection times due to their “ready-made” nature, that is, their continuous and passive availability (Salganik, 2019). They thus epitomise the “velocity” component of so-called ‘big data’, alongside volume and variety (Laney, 2001), and are particularly suited to rapidly evolving demographic contexts. For instance, mobile phone data were used to map population shifts following the 2010 earthquake in Haiti (Lu et al., 2012), while Facebook user data have provided insights into internal displacement following the Russian invasion of Ukraine (Leasure et al., 2023).

These sources also offer the potential for extensive population and geographic coverage without the heavy infrastructure typically required for censuses, centralised registers

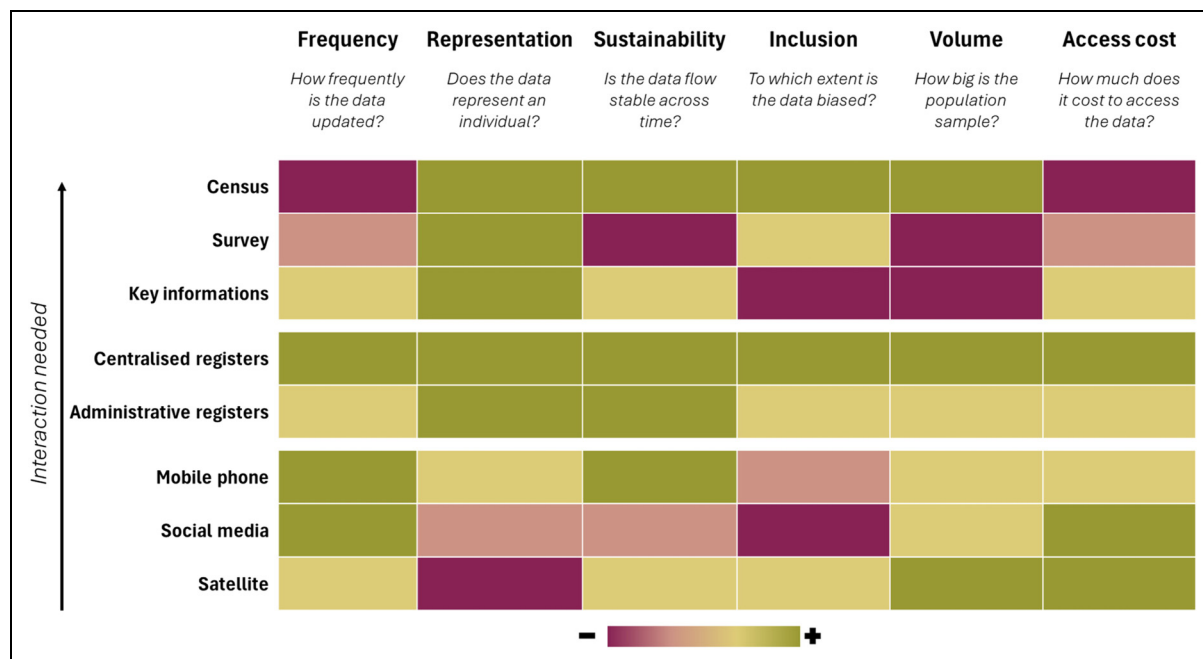


Figure 1. Technical advantages and limitations of different population data sources in relation to the level of intervention required for their collection. The colour scale indicates the performance of each source, ranging from least favourable (pink) to most favourable (green).

or administrative records. As such, they are especially attractive in lower-resource contexts. For example, Burkina Faso's National Statistics Office used satellite imagery to supplement census enumeration in areas inaccessible to their teams (Darin et al., 2022), a method also adopted by Colombia's statistical office (Sanchez-Cespedes et al., 2023).

However, the limitations of digital demographic data must not be overlooked. Since these data are not collected for statistical purposes, they do not constitute a direct representation of individuals, requiring transformation, often involving statistical modelling, into usable demographic units. Additionally, they may systematically exclude certain subgroups, particularly the most vulnerable, who might not own a mobile phone or access digital platforms. Figure 1 provides a synthetic overview of the technical advantages and limitations associated with each population data source, with a view to producing reliable national figures at high spatial and temporal resolution. It compares the main sources of demographic data across six criteria: update frequency, individual representation, data stream continuity, inclusiveness, volume, and access cost. The sources are classified according to the degree of intervention they require, distinguishing traditional approaches (such as censuses and surveys) from digital sources (including mobile phone data, social media, and satellite imagery). The census, for instance, offers an accurate snapshot of the population but is infrequent and costly. By contrast, digital data are more abundant and regularly updated, yet more susceptible to coverage bias. Taken together, the figure highlights the methodological trade-offs specific to each type of data.

What breaks with the census?

“It is ironic that the information society may ultimately lead to the demise of the traditional census” (Whitby, 2020). Whitby's closing remark on the history of population enumeration signals a deeper transformation in population data collection practices. These changes raise not only technical concerns about data quality, but also reshape the political order of population figures production, and by extension, the power relations embedded in the sovereign authority to know the size of the population. This article challenges three common assumptions about data: first, that data are passively “given” rather than constructed (inertia); second, that it is impossible to organise collectively around individual data (atomisation); and third, that data must be captured immediately and reflected on only afterwards (immediacy) (Ruppert et al., 2017). It examines, therefore, how technical shifts in data collection are driving deeper structural transformations.

Figure 2 illustrates the implications of the digital shift on population figures, from data collection to demographic figures, including the emergence of new power relations.

First, a transformation in the mode of data collection can be observed. As noted earlier, the process shifts from active collection, where the transfer of demographic information depends on the conscious participation of citizens, to passive collection, which can, to varying degrees, become entirely invisible to the individuals whose data are being extracted. For example, with population registers and

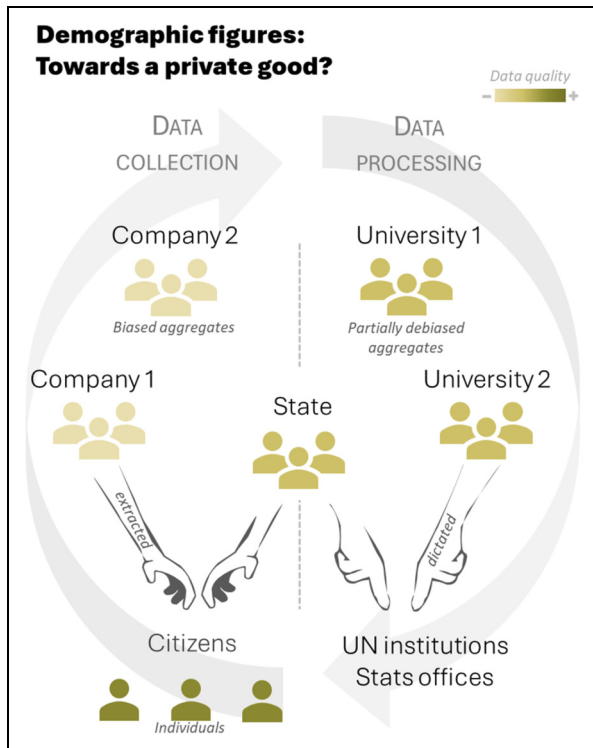


Figure 2. Schematic representation of the lifecycle of population data in the digital era.

administrative records, although individuals are aware of providing information—for instance, when updating their address—they are not always clearly informed about secondary uses of this data, nor do they have effective means of objection, which has led some Western countries to forgo such systems (Poulain et al., 2013). But the most striking example is satellite imagery, where the capturing device operates hundreds of kilometres away from the subjects. Second, there is a diversification of collectors. At the national level, there may be several mobile phone operators, multiple social media platforms, and various public administrations, each holding a partial view of the population. This break in the monopoly over demographic data collection is coupled with a shift in the purpose of collection, now often operational and/or commercial, no longer aligned with the rigour of statistical standards. Third, the monopoly over demographic figures production is being disrupted. The diversification of data sources leads to a proliferation of population figures from various producers. For example, five different teams have produced displacement estimates for Ukraine: Leasure et al. (2023) using Facebook data, Rowe et al. (2025) using mobile geolocation data from a data broker, Shibuya et al. using data from another broker (2024), Rufener et al. (2024) using satellite imagery, and Liu et al. (2024) using data from X. This multiplication of estimates breaks with the governance model historically centred on the national statistical institute, shifting towards

academic or even private control over population estimation, requiring specialised technical expertise.

Demographic figures: A public good turned private in the digital era

Economic theory categorises goods along two axes: rivalry, where one individual's use precludes another's, and excludability, where access can be restricted (Ostrom, 1990). This framework helps us understand how digital technologies have reshaped demographic figures, contributing to its transformation into a private good. As Hess and Ostrom (2006) observed with formerly collective resources such as outer space or the deep seabed, technological advances can render these resources exploitable—making them rival, even if non-excludable—and subject to dynamics of appropriation and scarcity. Managing such common resources thus demands sophisticated governance to counter enclosure and harmful commodification.

To apply Ostrom's framework to demographic figures, it is useful to distinguish two key phases: the collection of data and the dissemination of figures. I argue that rivalry primarily affects the collection phase, while excludability shapes dissemination. Before the digital era, the high cost of data collection meant it was non-rival and conducted by governments. Dissemination, in turn, became non-excludable as democratic norms promoted broad access. This balance, however, has been profoundly disrupted by digital technologies.

Today, data collection is increasingly rivalrous. While, in theory, individuals can share their data with many actors, in practice, capture by one limits access for others. This is due to (1) competition among firms, such as telecoms, where individuals appear in only one dataset; and (2) rising digital fatigue, which reduces individuals' willingness to share data, especially through more invasive means like surveys. Simultaneously, knowledge dissemination has become excludable: population figures now represent strategic, commodified assets that confer competitive advantage (Sadowski, 2019). For instance, telecommunications firms that once shared user data during humanitarian hackathons now commercialise it, converting demographic figures into private goods.

This fragmented ecosystem of data collection and control, akin to the enclosure movement, justifies recognising demographic data as a knowledge commons (Hess, 2008), highlighting its structural vulnerability. However, this vulnerability arises not from overuse, as in the tragedy of the commons, but from underuse, as in the tragedy of the anti-commons (Heller, 1998): the proliferation of exclusive control rights restricts access and prevents the optimal use of a resource that is, by all accounts, valuable. Today, no single actor holds data that are both granular and comprehensive, and only collective data governance can ensure

both the quality of knowledge produced and preserve its essential non-excludability. In this sense, escaping the current prisoner's dilemma where each actor retains exclusive and scattered fragments of data, is critical: the persistence of siloed, incomplete raw datasets harms all stakeholders, from public service providers to private-sector stakeholders.

Moreover, the current enclosure regime, driven by digital actors, undermines citizens' fundamental right to access and control data derived from their own person. This right is largely framed at the individual level, as in the EU's General Data Protection Regulation (GDPR), which ultimately shifts responsibility for a collective resource onto each user without equipping them to understand, let alone mitigate, systemic risks (Gordon-Tapiero et al., 2025). In light of this individualisation of data governance, it is urgent to reimagine their integration into a knowledge commons, a resource that ultimately belongs to all citizens. What is now required is the reinvention of a collective governance framework for demographic figures, one that is commensurate with the social, political, and ethical stakes they entail, a proposal developed in the final section of this article.

Digital era and excludability: A return to extractive practices in population figures

The transformation of demographic figures into a private good, catalysed by the digital era, manifests through two interrelated dynamics: on the one hand, data collection has become rivalrous; on the other, dissemination has become lucrative and, consequently, exclusive. In this section, I argue that while the rivalry in data collection is a specific feature of the digital era (discussed in Section "Digital data and rivalry: a necessary modelling of demographic data"), restricting the access to demographic knowledge has deeper historical roots, originating from a time when producing population figures was not intended for the collective good.

The first censuses were ordered by sovereigns to quantify and consolidate their power, rather than to serve the needs of the populations being enumerated. This asymmetry between data collectors and the subjects of the data corresponds to what Sadowski (2019) describes as an "extractive data practice." This critical perspective helps examining in the first subsection, the history of census-taking and the gradual emergence of citizen empowerment since 1703, whereby the duty to "be counted" is progressively coupled with the right to "be taken into account." In doing so, demographic figures evolved from being a club good (non-rivalrous collection and exclusive dissemination) to a public good (non-rivalrous collection and non-exclusive dissemination). The second subsection examines the emergence of population figures derived from

administrative records as successors to traditional census forms, highlighting the ethical tensions they continue to raise. Finally, the last subsection examines how digital data collection disrupts the empowerment mechanisms embedded in the census as a public good.

From the "mirror of the prince" to the "mirror of the nation": The evolution of census ends

The origins of census-taking lie in the sovereign's need to survey their domain, a symbolic extension of their own body (Desrosières, 2010). In this "royal census" model, population enumeration served as the "mirror of the prince" to affirm state power by identifying taxable and conscriptable subjects. Emerging in early civilisations such as Mesopotamia (~3000 BCE), pre-censuses became more systematic under the Roman Empire (from 508 BCE) and Han Dynasty (from 2 CE), where they served military and fiscal purposes. In medieval Europe, instruments like England's *Domesday Book*, France's *états des feux*, and Tuscany's *catasto* remained largely extractive. This logic extended into colonial settings—Peru (1548), the French Antilles (1660), North America (1666), and Ireland (1679)—where censuses helped consolidate metropolitan control and allocate wartime resources (Cahen, 2022; Hacking, 1990; Thorvaldsen, 2017). Notably, English liberal thinkers like Thornton vehemently opposed the introduction of a census in Britain, fearing it would deal "a fatal blow to the last vestiges of English liberty" (Whitby, 2020). Such extractive approaches continued into the twentieth century; under Nazi Germany, enumerators accompanied military campaigns, aligning demographic data with authoritarian aims (Aly and Roth, 2017). This model sometimes entailed secrecy, with data becoming state property rather than collective knowledge (Wastergaard, 1932).

Such practices exemplify an extractive circulation of data in service of *biopower*, the management of populations through detailed knowledge (Foucault, 2004). Within interstate competition, demographic knowledge became strategic, and thus excludable. Yet its high fixed costs and state monopoly kept its collection non-rival, classifying the royal census as a *club good* accessible only to ruling elites. As a result, scholars of the time were compelled to develop alternative methods of political arithmetic to produce population figures (Cahen, 2022).

A turning point came in 1703, when Denmark dispatched officials to Iceland to count all inhabitants during a severe glacial period (Tomasson, 1977). Sweden followed in 1748, motivated by smallpox concerns (Whitby, 2020). These initiatives reoriented the census toward public welfare. It became a "mirror of the nation" (Desrosières, 2010), a shift reinforced by the rise of democratic regimes. The U.S. Constitution (1790) tied enumeration to

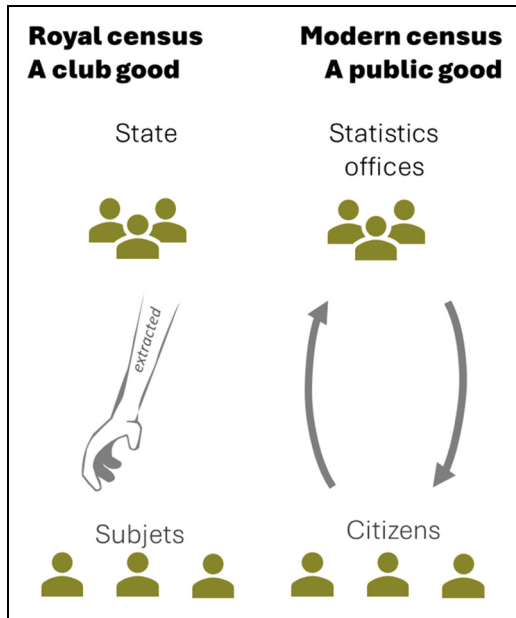


Figure 3. Diagram comparing population data flows in the age of the prince and the age of the nation.

representation, and by 2020, 54 countries referenced the census constitutionally, with 32 linking it to electoral systems (Elkins et al., 2014). This nexus of census and democracy has also inspired protest and reform. In 1911, British suffragettes refused enumeration, declaring: “if I am not a citizen, I shall not perform the duties of one” (Liddington and Crawford, 2014). Conversely, the 1967 repeal of Section 127 of the Australian Constitution, which excluded Aboriginal people from the census, affirmed citizenship recognition (Attwood, 2007).

The pursuit of inclusive representation also spurred technical innovation, not only to improve accuracy, but also to enhance transparency and co-production. For example, U.S. census forms were publicly displayed for validation until 1840 (Gatewood, 2001), while Prussian citizens filled in their own forms (von Oertzen, 2021). In this transformation of both purpose and dissemination, population figures evolved from a *club good* to a *public good*, produced by the state not only due to cost, but through democratic mandate (Figure 3).

It is important to note that both versions of the census—royal or modern—are not mutually exclusive and if their birth is historically consecutive, the tension between those two ends is inherent to the aim of a comprehensive enumeration of a population requiring a cartographic gaze thus a dominating gaze (Anderson, 1991). To counterbalance this top-down power, explicit systems of citizen empowering have been developed. A similar tension can be found in another form of population data collection: the recording of every individual in administrative records.

The ambiguous case of administrative records

Population figures derived from administrative records, whether centralized or not (see Section “The digital revolution”), provide a continuous system of demographic data collection for administrative purposes, reframing tensions between extractive motives and the common good, and between restricted and open dissemination. Unlike the episodic census, they rely on the ongoing registration of individuals—for instance, when applying for benefits or updating residency—making data collection more “passive,” as a by-product of administrative procedures, and less “reactive,” since opting out usually entails losing access to essential services.

This passivity also makes knowledge dissemination more exclusive, as it tends to serve operational objectives and reduces the transparency of collection, contributing to an inherently more extractive model vis-à-vis citizens. One of the most striking examples is the use of population registers by the Nazi regime and its collaborators to track Jewish individuals during the Second World War. In the Netherlands, 73% of the Jewish population was exterminated between 1941 and 1945, compared to 40% in Belgium and 24% in France, a difference that Blom (1989) attributes in part to the quality and comprehensiveness of the Dutch registration system.

Nevertheless, the systematisation of administrative records also reflects a commitment to the common good. These systems ensure that individuals cannot “disappear without a trace,” and thereby operationalise the “right to have rights,” as articulated by Hannah Arendt. This principle is embedded in international law: civil registration is recognised as a fundamental right under the Convention on the Rights of the Child. In the United States, for instance, individuals must demonstrate their descent from someone listed in the Dawes Rolls, a register established in 1898 to enumerate Native Americans, in order to be officially recognised as Indigenous and claim corresponding rights (Miller, 2015).

Hence, administrative records straddle a delicate boundary. While they can entrench extractive practices, they may also protect individuals’ legal recognition and facilitate access to rights, particularly when designed with transparent and equitable governance in mind.

Digital data: An impossible empowerment?

Digital traces amplify the passive nature of demographic data collection. They are the residual footprints left behind through interactions with digital services, such as social media or mobile networks. In the most extreme case, satellite imagery, data can be collected, without any conscious action on the part of the individuals observed. This form of data collection is also non-reactive: since it takes place in the background, it does not influence

behaviour. Yet it echoes earlier extractive practices where data were gathered without the participation of those concerned. A telling example is the use of satellite imagery in the 1991 South African census to bypass the inability of apartheid-era enumerators to enter the township of Soweto (Khalfani et al., 2008).

This episode highlights what is at stake when field-based methods of data collection are replaced: the social consent required for a door-to-door census. The disappearance of the interaction between enumerator and respondent, the “data at the doorstep” (Schlicht et al., 2021), also erodes the possibility of a direct encounter between the state and its citizens. In such encounters, citizens are not only counted, but also seen and heard, embodying the “pastoral power” of the state (Foucault, 2004). The adoption of silent collection methods also disrupts the census as a ceremonial moment of national belonging: every individual is subject to the same operation, at the same time, across the entire country (Coleman, 2013). It enables a dual process of identification: individuals recognise themselves as members of the population (subjectivation), while the state identifies and aggregates individuals into a population (objectivation) (Ruppert, 2009). However, when data collection serves operational or even profit-driven objectives, their dissemination becomes excludable thus eroding their credibility, undermining trust in the governing institutions, and severing any aspiration toward reciprocity, belonging, or the possibility for citizens to actively engage in the process. In such cases, the entire system of demographic figures production shifts toward an extractive logic, breaking away from the democratic ideal of a shared body of knowledge, one that is both sustained by and in service to the citizens.

Digital era and rivalry: A necessary modelling of demographic data

While the previous section addressed the increasing excludability of demographic figures dissemination, this section examines how demographic data collection has become a rivalrous process and explores the implications for the mandate and role of national statistical institutes. Digital technologies have enabled a proliferation of actors capable of conducting partial or approximate population enumeration. Instead of seeking a single, complete, high-resolution source of data, I advocate for a modelling framework that integrates this abundance of available data. This shift necessitates rethinking the role of statistical institutes, traditionally focused on the deterministic production of data, rather than the stochastic estimation of latent processes. The section concludes by reflecting on the implications of outsourcing the production of demographic figures beyond the statistical sphere, and its impact on the stewardship of democratic figures.

From diversity to equivalence: Modelling as a key for leveraging digital data

With the explosion of data sources in the digital era, demographic figures are no longer produced solely through standardised statistical procedures but arise from a multiplicity of heterogeneous data sources: mobile phones, social networks, satellite imagery, and administrative records. These data flows are opportunistic, shaped by the technical infrastructures in place rather than deliberate observation strategies.

The first challenge is completeness. No data source covers the entire population. Mobile phone data, for instance, pertain only to a specific operator’s subscribers, excluding clients of other networks and those without a phone. Data from advertising platforms like Meta (Facebook) are limited to active users and shaped by opaque algorithmic criteria. Conversely, certain groups can be over-represented: heavy users, duplicate accounts, corporate accounts, bots, or individuals not removed after international migration. These representativeness biases, specific to each data source are neither random nor stable—they vary by age, gender, education, geography, and season. The second challenge is indirectness. Digital data tend to not measure individual presence directly; rather, they capture proxies, connections, posts, buildings, that must be translated into demographic units. For example, mobile phone data rely on home detection algorithms that determine night-time location via the most frequently used cell tower. Such inference is based on behavioural assumptions that may be invalid for night workers, mobile populations, or shared devices. A third challenge lies in the structural heterogeneity of data sources. Different sources, telephony, social platforms, satellite imagery, do not share comparable units, common identifiers, or synchronised spatial and temporal resolutions. Even administrative records contain individuals that would not be counted in a census population with the difference between their *de jure* and *de facto* residence.

Thus, producing robust, unbiased, high-resolution demographic figures requires reconciling disparate, segmented data streams. This calls for moving beyond a deterministic view of enumeration, where each individual must be counted, to an integrated modelling approach that can (1) convert signals into people, (2) corrects for coverage bias and (3) combine heterogeneous streams together. A promising direction is probabilistic modelling, which treats the population as an unobservable latent parameter to be estimated from partial and biased inputs (Darin et al., 2025; Leasure et al., 2020).

Official statistics and modelling: Methodological caution and institutional resistance

Unlike humanitarian statistics, where short timeframes limit the scope for sophisticated modelling, official statistics

have been working to integrate digital population data in two areas. First, numerous countries are incorporating administrative data into their population figures either to supplement or complement census data—as seen in New Zealand (Stats NZ, 2023), Uruguay (Instituto Nacional de Estadística, 2024) and Canada (Statistics Canada, 2022)—or to produce experimental population figures as in Ecuador’s REBPE and Colombia’s REBP (Departamento Administrativo Nacional de Estadística, 2021; Instituto Nacional de Estadística y Censos, 2024). These efforts aim at consolidating administrative records through extensive data cleansing and harmonisation. However, coverage biases have so far been addressed only through post-enumeration surveys (Pfeffermann et al., 2019). The need for more sophisticated probabilistic modelling has clearly been demonstrated (Bryant and Graham, 2015; Office for National Statistics, 2023). The second data source attracting institutional interest is mobile phone data celebrated for capturing population mobility, particularly during the COVID-19 pandemic (Wang et al., 2020). Statisticians have advocated for stochastic modelling approaches to: (1) integrate multiple data sources; (2) account for uncertainty in transforming signals into statistics; (3) ensure resilience to technological shifts; and (4) minimise required input data (Salgado et al., 2021). Despite this, these data continue to be processed by statistical institutions using deterministic algorithms (Instituto Nacional de Estadística, 2021).

Institutional reluctance towards statistical modelling stems from the origins of official statistics. While early demographic modelling emerged already in the seventeenth century with political arithmetic—such as Graunt’s use of mortality schedules—modern governmentality led to a focus on empirical data. The Royal Statistical Society’s original motto, *Aliis exterendum* (“for others to thresh out”), made this ethos explicit, stating that its aim was solely “to gather facts” (Hilts, 1978). Even today, exhaustive enumeration remains the benchmark of statistical legitimacy, and modelling remains fraught with suspicion, as shown by the controversy over sampling in the 2000 U.S. census (Anderson and Fienberg, 1996). Where statistical models are employed, they are usually limited to correcting omissions or undercounts via post-enumeration surveys or demographic analysis. In these cases, inference remains tied to a priori sampling schemes, rather than probabilistic models (van den Brakel and Bethlehem, 2008). Consequently, statistical errors addressed are those of sampling, not measurement or selection bias. Stochastic modelling is thus largely confined to intercensal periods through forecasting models, introduced reluctantly in the 1950s and not adopted universally despite decades of methodological development (Gans, 2002). The resistance is partly technical—modelling introduces non-determinism, non-replicability, sensitivity to unverifiable assumptions, and temporal instability—partly political: official statistics are designed to be data providers and not estimates providers.

Asymmetries of knowledge and power in the digital era: Towards a return of expert colonialism?

A particular case of using digital traces for the official production of population figures deserves attention: the use of satellite imagery to estimate populations in areas inaccessible to enumerators, not through simple multipliers, as in apartheid-era South Africa (see Section “Digital data: an impossible empowerment?”), but through statistical modelling. First trialled in Afghanistan in 2017, this approach has proven effective in several contexts: Colombia’s 2018 census (Sanchez-Céspedes et al., 2023), Burkina Faso’s 2020 census (Darin et al., 2022), and Mali’s 2022 census (Institut National de la Statistique du Mali, 2023). These modelling exercises estimate population figures based on building footprints, supplemented by auxiliary geographic variables. However, they are typically led not by national statistical officers but by academic partners due to the specific technical skills required and the novelty of the methods. A similar exercise in Papua New Guinea sparked widespread criticism when the estimates nearly doubled those of the 2011 census projections (Laveil, 2023). Behind the tension between estimation and enumeration lies a deeper issue: the contest between national statistical offices and external technical experts. Competition over data collection thus escalates into competition over the production of population figures, culminating in struggles over epistemic authority.

Yet the census is the cornerstone of national statistical authority. As Desrosières (2005) noted, it is through the census that public statisticians established “a new and original professional identity,” distinct from both classical administration and academic research. Indeed, fifty-four countries enshrine the census in their constitutions as the principal mandate of their national statistical institutes (Elkins et al., 2014). Moreover, the data revolution launched by the United Nations with the Millennium Development Goals positioned statistical offices at the heart of global information systems (Sustainable Development Solutions Network, 2015). The clearest expression of this sovereign control over data is perhaps Tanzania’s 2015–2019 legislation banning public discussion or challenge of official statistics (Nyeko, 2019). Thus digital-era data may represent the latest chapter in a broader trend of marginalising and partially privatising public statistics, especially in Africa (Bédécarrats et al., 2016).

Circumventing national statistical systems, while potentially efficient in the short term, raises serious questions of democratic legitimacy and opens the door to neo-colonial forms of expertise and power, reminiscent of colonial censuses. Global North actors are often better equipped to harness digital-era data due to superior: (1) infrastructure; (2) technical skills; and (3) economic incentives, especially given that salaries in sub-Saharan African statistical offices are closely tied to field deployment. Population estimates

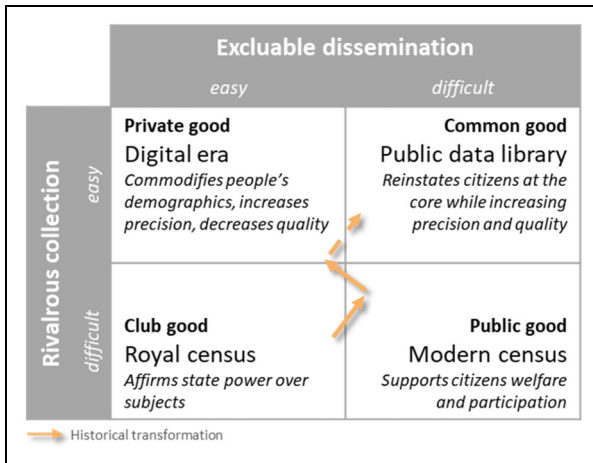


Figure 4. Historical evolution of the nature of demographic figures following Ostrom's typology of goods (1990).

today must transcend the “caveat emptor” (buyer beware) logic that characterises information derived from digital traces (Johns, 2023). Instead, they must be validated and endorsed through processes overseen by legitimate authorities accountable to citizens, in line with a vision of democratic sovereignty over population data governance. Colombia's 2018 census is a good example where academics were approached to teach the novel methodology to the Colombian statistics office (Sanchez-Cespedes et al., 2023).

Towards a reconstitution of demographic data as a common good: Proposal for a collective governance

In line with Roca and Letouzé (2016), I contend that the digital revolution needs to be reframed as a gateway to renewed dialogue and a revised social contract, with citizen participation and accountability as both tools and objectives. Specifically, in the field of population figures, I traced the evolution of their status over time: from a club good, costly to collect and restricted in use, to a public good shaped by democratic demands. Today, however, as illustrated in Figure 4, the multiplication of data sources due to the digital revolution brought an increasingly exclusive and commodified dissemination which has transformed population figures into a private good, extractive in its collection and commercial in its purpose. I argue that it is both possible and necessary to claim demographic figures as a common good: one that recognises the diversification of data collectors while promoting the open sharing of raw aggregates and estimated figures. This would enable both technically robust and politically democratic population figures. To counter increasingly rivalrous data collection, all data collectors should bear a duty to share anonymised aggregates to enable joint modelling. And to counter increasingly restrictive and thus extractive dissemination practices,

all institutions producing demographic figures should assume a duty of transparency and dissemination. Ultimately, I call for a citizen-centred governance model in which demographic estimates are built on an open public demographic data library.

Addressing exclusive dissemination: The transparent dissemination of demographic figures

To restore the democratic tradition of citizen empowerment and counter extractive demographic estimation, I argue for a dissemination duty incumbent upon any institution engaged in population modelling, whether public, semi-public, or private. Inspired by the long-standing practices of national statistical offices laid out in the UN Fundamental Principles of Official Statistics, the OECD's Good Statistical Practice, and the European Statistics Code of Practice, this duty must now extend across the broader ecosystem of demographic figures producers.

The barriers to demographic knowledge hampers both the provision of appropriate public and private services and citizens' ability to understand and shape their collective identity, participating in the social construction of population (Ruppert, 2009). Effective dissemination must comply with the FAIR principles: Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2016), and the CARE principles: Collective benefit, Authority to control, Responsibility, and Ethics (Global Indigenous Data Alliance, 2019). In practice, this requirement of dissemination implies that demographic figures should be accompanied by rigorous, clear, and accessible documentation, drawing on the standards promoted by the open science community. Where feasible, documentation should be supplemented by the release of source code with sufficient explanation for reproducibility, as demonstrated for instance by Statistics Canada's open-source practices (Statistics Canada, 2025).

However, privacy must be protected. I do not advocate publishing figures that could harm individuals. Privacy-preserving measures such as anonymisation, pseudonymisation, tiered access, aggregation, differential privacy, or synthetic data can reconcile openness with personal data protection. The choice of a privacy protection model must be context-specific, but it is essential to provide a well-reasoned justification that assesses the balance between societal value, risks, and potential harms, in order to legitimise the chosen approach to ensuring such protection.

Finally, operational feasibility must be acknowledged. For many producers of demographic figures, especially in low-resource settings, capacity constraints pose real challenges to comprehensive dissemination. In such contexts, international collaboration and resource pooling, through regional statistical initiatives or humanitarian data platforms

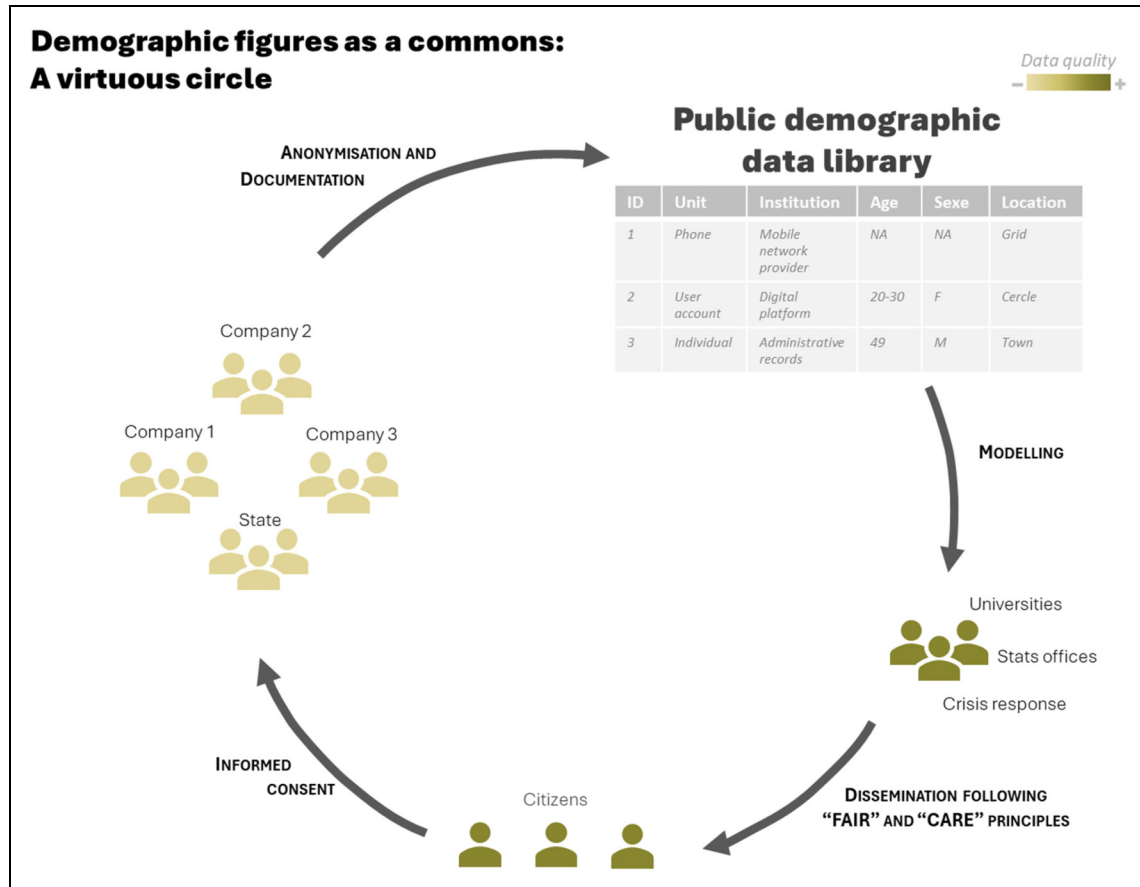


Figure 5. Schematic representation of virtuous demographic data and figures flows.

like the Humanitarian Data Exchange (OCHA Services, 2025) offer practical and scalable solutions.

Addressing rival collection: The obligation to share aggregate data

Building on open government practices which requires public administrations to open administrative data by default, I propose extending this principle to demographic data, including those collected by private actors. The point is not to advocate for additional or more exhaustive collection aiming at having a digital twin of the population, but to recognise the many ongoing collection practices and ensure their better use through systematic data sharing. This rests on a key distinction: collective ownership, reflecting the societal value of demographic data, versus delegated management by private entities. Private operators should not retain full ownership, as such data are the basis of a knowledge commons and must be governed accordingly. Thus, in contrast to initiatives based on the voluntary donation of data by individuals—that is, the re-sharing of personal data for public interest purposes—which align with the individualising logic of the GDPR

but have so far engaged only a minority of users (van Driel et al., 2022) I argue that it is more effective to intervene, through legal means, directly at the level of data-collecting actors, particularly when their coverage exceeds a certain population threshold. One possible articulation of this intervention is through a licensing framework, operating at the entity and/or individual levels. At the entity level, organizations collecting demographic data would be required to sign data-sharing licenses mandating reciprocal contribution, drawing on models such as the UK Open Government Licence. At the individual level, individual consent would be conditioned on the anonymised reversion of demographic data into a common repository, raising therefore also citizen awareness of data collection efforts. This could be structured in ways comparable to PSD2's "open banking," where financial institutions must grant third-party access with user consent (Dimachki, 2019); and international initiatives in genomics, such as the Global Alliance for Genomics and Health, where individual data contributions are conditional on inclusion in collective repositories (Knoppers, 2014).

To avoid mistrust and misuse, particularly in the context of surveillance capitalism (Zuboff, 2015), I do not suggest sharing individual-level data but rather aggregated data

across different dimensions such as: temporal (day instead of minute), geographical (grid instead of GPS coordinates), demographical (age group instead of birthdate). Additionally, to protect business interests, the identity of the data-collecting company could be anonymised under a standardised typology (e.g., “social network,” “telecom provider”), with modelling used to impute any gaps.

Governing the demographic commons: Towards a public demographic data library

To operationalise data sharing, I suggest drawing inspiration from the EU Data Governance Act, which promotes the creation of data intermediaries, technical and legal infrastructures facilitating secure, transparent, and traceable data sharing. These non-profit entities do not extract direct economic value from the data but enable smoother exchange. Similar intermediaries already exist in agriculture (AgDataHub) and business data (Dawex). Compared to initiatives like OpenSafely (for medical data (Andrews et al., 2022)), I advocate for a more operational and thematically focused database encompassing diverse data collectors, not solely academic researchers. Unlike the Humanitarian Data Exchange (OCHA Services, 2025), our model promotes rawer, more standardised, and thematically targeted data sharing. Unlike Open Humans (Greshake Tzovaras et al., 2019) or Posmo (Yatsenko, 2024), data would be contributed by institutions rather than individuals.

A key feature of this intermediary, unique to demographic figures, is its role in maintaining public trust and ensuring democratic governance. Here, a parallel can be drawn from the ongoing UK discussion on a “National Data Library” (Wellcome Trust and Economic and Social Research Council, 2024), which explores technical data-sharing models combining interoperability, confidentiality, and sovereignty. In our proposal, a “Public Demographic Data Library” would serve this function. While governance could be led by national statistical institutes as they have inherent and unique expertise to lead data stewardship (United Nations Economic Commission for Europe, 2024), alternative trusted third parties could also play this role depending on the context (for example in midst of a humanitarian crisis, to ensure data sharing practices beyond borders, or to mitigate political change risk). The Human Technology Foundation (2020) offers useful recommendations on trust (“transparency by design,” independent governance, conflict resolution), data quality (standard protocols, communication interfaces, anonymisation policies, contributor communities), and data reuse (application-based access). Contributors would be required to document and regularly update raw data (“update by design”), and users would be required to share and document their estimates. Figure 5 summarises the principles of a virtuous demographic data circulation.


Conclusion

Transforming the “raw metal” of diverse demographic data generated in the digital era into the “(almost) pure gold of general knowledge” (Desrosières, 2005) requires not only the technical capacity to aggregate, standardise, and contextualise data, but also the political and institutional will to organise their governance transparently, inclusively, and sustainably. The concept of a “public demographic data library” may seem utopian. Yet precedents such as meteorological or satellite imaging data, historically open, interoperable, and collectively valued, demonstrate that large-scale public data can be managed as shared resources for the common good. This transition depends less on technological innovation than on our collective ability to institute new, trustworthy intermediaries capable of orchestrating the shared governance of demographic figures. It is imperative to move beyond the current function of merely “guiding attention” performed by an abundance of dashboards (Johns, 2023), and to re-anchor digital traces in empirically grounded models that represent populations accurately and inclusively. Such a paradigm shift is essential to overcome the fragmentation of the social body induced by the multiplication of digital interfaces, and to reaffirm a cohesive and democratic vision of demographic knowledge, a foundational condition for instituting demographic figures as a genuine common good.

Acknowledgements

The author would like to express their gratitude for the constructive feedback on the manuscript provided by Hugo Bailly and Vincent Straub. This piece would not have been possible without the unwavering support, guidance, and intellectual generosity of Prof. Ridhi Kashyap and Dr Douglas Leasure, whose influence shaped the work at every stage. Discussions held during workshops of the Collaborative Group for Nowcasting Population, bringing together researchers from the Universities of Liverpool, Manchester, Oxford, and Southampton, also helped shaping the normative arguments of the article. Feedback from participants of the United Nations Population Division Seminar helped consolidate the overview of the characteristics of new digital data sources.

ORCID iD

Edith Darin  <https://orcid.org/0000-0002-8176-092X>

Funding

This work was supported by the Leverhulme Trust under Grant RC-2018-003 for the Leverhulme Centre for Demographic Science. The author gratefully acknowledges the resources provided by the International Max Planck Research School for Population, Health and Data Science (IMPRS-PHDS).

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Notes

- 1 In Europe: Austria, Denmark, Finland, Monaco, Norway, Slovenia, Svalbard, Sweden and the Vatican. In Asia: Bahrain and Singapore. In North America: Greenland (United Nations Statistics Division, 2025).
- 2 In Europe: Germany, Belgium, Spain, Estonia, Iceland, Italy, Latvia, the Netherlands, Poland, Czech Republic, Switzerland. In Asia: Israel, Turkey. (United Nations Statistics Division, 2025)

References

- Abdelmagid N and Checchi F (2018) Estimation of Population Denominators for the Humanitarian Health Sector. Guidance for Humanitarian Coordination Mechanisms. Health Cluster, World Health Organisation. Available at: <https://healthcluster.who.int/publications/m/item/estimation-of-population-denominators-for-the-humanitarian-health-sector>.
- Aly G and Roth KH (2017) *The Nazi Census: Identification and Control in the Third Reich*. Philadelphia, PA: Temple University Press. ISBN-13:978-1592132591.
- Anderson B (1991) Census, map, museum. In: *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London, UK: Verso, 163–185. ISBN-13: 978-1844670864.
- Anderson M and Fienberg SE (1996) An adjusted census in 1990: The supreme court decides. *CHANCE* 9(3): 4–9. Taylor & Francis.
- Andrews C, Schultze A, Curtis H, et al. (2022) OpenSAFELY: Representativeness of electronic health record platform OpenSAFELY-TPP data compared to the population of England. *Wellcome Open Research* 7: 91.
- Attwood B (2007) *The 1967 Referendum: Race, Power and the Australian Constitution*. Canberra, Australia: Aboriginal Studies Press. ISBN: 9780855755553.
- Barlet M and Lefebvre O (2024) Le recensement annuel fête ses 20 ans!. *Le blog de l'INSEE*: 1–1. <https://blog.insee.fr/le-recensement-annuel-fete-ses-20-ans/>.
- Bédécarrats F, Cling J-P and Roubaud F (2016) The data revolution and statistical challenges in Africa: Introduction to the special report. *Afrique Contemporaine* 258(2): 9–23. ISBN: 9782807390072
- Billari FC (2022) Demography: Fast and Slow. *Population and Development Review* 48: 9–30.
- Blom JCH (1989) The persecution of the Jews in the Netherlands: A comparative Western European perspective. *European History Quarterly* 19(3): 333–351. Sage Publications Sage CA: Thousand Oaks, CA.
- Bowden S, Braker K, Checchi F, et al. (2012) Implementation and utilisation of community-based mortality surveillance: A case study from Chad. *Conflict and Health* 6(1): 11.
- Bryant JR and Graham P (2015) A Bayesian approach to population estimation with administrative data. *Journal of Official Statistics* 31(3): 475–487.
- Cahen F (2022) *Le Nombre Des Hommes. La Mesure de La Population et Ses Enjeux (Xvie-Xxie Siècle)*. Paris, France: Classiques Garnier. DOI: 10.48611/isbn.978-2-406-13146-5.
- Coleman D (2013) The twilight of the census. *Population and Development Review* 38(s1): 334–351.
- Darin E, Kashyap R and Leasure DR (2025) How accurate are high resolution settlement maps at predicting population counts in data scarce settings? *Population, Space and Place* 31(6): e70083.
- Darin E, Kuépié M, Bassinga H, et al. (2022) The population seen from space: When satellite images Come to the rescue of the census. *Population* 77(3): 437–464.
- Departamento Administrativo Nacional de Estadística (2021) *Registro Estadístico Base Población - REBP 2018*. Informes de Estadística Sociodemográfica Aplicada 4. Bogotá, Colombia: Departamento Administrativo Nacional de Estadística. Available at: <https://www.dane.gov.co/files/investigaciones/poblacion/informes-estadisticas-sociodemograficas/2021-10-01-Registro-Estadistico-Base-Poblacion-REBP-2018.pdf>
- Desrosières A (2005) Décrire l'État ou explorer la société : Les deux sources de la statistique publique. *Genèses* 58(1): 4–27. Belin.
- Desrosières A (2008) *Gouverner par les nombres : L'argument statistique II. Sciences sociales*. Paris: Presses des Mines. DOI: 10.4000/books.pressesmines.341
- Desrosières A (2010) *La politique des grands nombres. Histoire de la raison statistique*. Paris, France: La Découverte. DOI: 10.3917/dec.desro.2010.01.
- Dimachki ME (2019) *More Data and More Data Sharing: Navigating an Open Banking World*. London, UK: Henry Stewart Publications. Available at: <https://www.ingentaconnect.com/content/hsp/jdb001/2019/00000003/00000003/art00003>.
- Elkins Z, Ginsburg T, Melton J, et al. (2014) Constitute: The world's constitutions to read, search, and compare. *Journal of Web Semantics* 27: 10–18. Elsevier.
- Foucault M (2004) *Sécurité, territoire, population: cours au Collège de France, 1977-1978*. Paris, France: Éditions Gallimard. ISBN: 978-2020307994.
- Gans Hd (2002) Law or speculation? *Population* 57(1): 83–108.
- Gatewood G (2001) *A Monograph on Confidentiality and Privacy in the U.S. Census*. Suitland, MD: US Census Bureau. Available at: <https://www2.census.gov/library/publications/2001/communications/monograph-on-confidentiality-and-privacy.pdf>.
- Global Indigenous Data Alliance (2019) CARE Principles for Indigenous Data Governance. Available at: <https://www.gida-global.org/care>
- Gordon-Tapiero A, Ligett K and Nissim K (2025) On the rival nature of data: Tech and policy implications. In: Proceedings of the 2025 Symposium on Computer Science and Law,

- New York, NY, USA, 25 March 2025, pp. 17–25. CSLAW '25. Association for Computing Machinery. DOI: 10.1145/3709025.3712211
- Greshake Tzovaras B, Angrist M, Arvai K, et al. (2019) Open humans: A platform for participant-centered research and personal data exploration. *GigaScience* 8(6): giz076.
- Hacking I (1990) *The Taming of Chance*. Cambridge, UK: Cambridge University Press. DOI: 10.1017/CBO9780511819766.
- Heller MA (1998) The tragedy of the anticommons: Property in the transition from Marx to markets. *Harvard Law Review* 111(3): 621–688.
- Hess C (2008) Mapping the New Commons. 1356835, SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. Available at: <https://papers.ssrn.com/abstract=1356835>
- Hess C and Ostrom E (2006) *Understanding Knowledge as a Commons: From Theory to Practice*. Cambridge, USA: MIT Press. DOI: 10.7551/mitpress/6980.001.0001
- Hilts VL (1978) Aliis extendum, or, the Origins of the Statistical Society of London. *Isis* 69(1): 21–43. The University of Chicago Press.
- Human Technology Foundation and Exploratoire Sopra Steria Next (2020) *Le data altruisme: les données au service de l'intérêt général*. Courbevoie, France: Sopra Steria Next.
- Institut National de la Statistique du Mali (2023) *Rapport préliminaire. Résultats globaux du RGPH5*. Available at: https://www.instat-mali.org/laravel-filemanager/files/shares/rgph/rapport-resultats-globaux-rgph5_rgph.pdf
- Instituto Nacional de Estadística (2021) Mobility study based on mobile telephony 2020–2021. Sub-Directorate General of Socio-Demographic Statistics. Available at: https://ine.es/en/experimental/movilidad/exp_em_proyecto_en.pdf
- Instituto Nacional de Estadística (2024) *Metodología del Censo Combinado 2023. 2023 Censo Uruguay*. Montevideo, Uruguay: Instituto Nacional de Estadística. Available at: <https://www5.ine.gub.uy/documents/CENSO%202023/Metodolog%C3%ADa%20del%20Censo%20Combinado%202023.pdf>
- Instituto Nacional de Estadística y Censos (2024) *Registro Estadístico Base de Población del Ecuador (REBPE)*. Quito, Ecuador: Instituto Nacional de Estadística y Censos.
- Johns F (2023) Populations: From statistics to data science. In: Johns F (ed) *#Help: Digital Humanitarianism and the Remaking of International Order*. Oxford, UK: Oxford University Press. DOI: 10.1093/oso/9780197648872.003.0003.
- Kashyap R (2021) Has demography witnessed a data revolution? Promises and pitfalls of a changing data ecosystem. *Population Studies* 75(Suppl. 1): 47–75. Routledge.
- Khalfani AK, Zuberi T, Bah S, et al. (2008) Race and population statistics in South Africa. In: *White Logic, White Methods: Racism and Methodology*. Lanham, MD: Rowman & Littlefield Publishers, 63–90. ISBN: 9798881853419.
- Knoppers BM (2014) Framework for responsible sharing of genomic and health-related data. *The HUGO Journal* 8(3). DOI: 10.1186/s11568-014-0003-1.
- Laney D (2001) 3D Data management: Controlling data volume, velocity and variety. *META group Research Note* 6(70): 1. Stanford.
- Laveil M (2023) PNG needs a census, not more population estimates. *The interpreter | Lowy Institute* 1: 1–1. <https://www.lowyinstitute.org/the-interpreter/png-needs-census-not-more-population-estimates>.
- Leasure DR, Jochem WC, Weber EM, et al. (2020) National population mapping from sparse survey data: A hierarchical Bayesian modeling framework to account for uncertainty. *Proceedings of the National Academy of Sciences* 117(39): 24173–24179. National Academy of Sciences.
- Leasure DR, Kashyap R, Rampazzo F, et al. (2023) Nowcasting daily population displacement in Ukraine through social media advertising data. *Population and Development Review* 49(2): 231–254.
- Liddington J and Crawford E (2014) *Vanishing for the Vote: Suffrage, Citizenship and the Battle for the Census*. Manchester, UK: Manchester University Press. Manchester University Press. ISBN: 978-0719087493
- Liu Z, Li J, Chen H, et al. (2024) Prediction of changes in war-induced population and CO₂ emissions in Ukraine using social media. *Humanities and Social Sciences Communications* 11(1): 1–11. Palgrave.
- Lu X, Bengtsson L and Holme P (2012) Predictability of population displacement after the 2010 Haiti earthquake. *Proceedings of the National Academy of Sciences of the United States of America* 109(29): 11576–11581. National Academy of Sciences.
- Miller M (2015) Dawes cards and Indian census data. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 48(4): 214–229. Routledge.
- Nyeko O (2019) Tanzania Drops threat of prison over publishing independent statistics | Human Rights Watch. In: *Human Rights Watch*. Dar es Salaam, Tanzania, 1–1. Available at: <https://www.hrw.org/news/2019/07/03/tanzania-drops-threat-prison-over-publishing-independent-statistics>.
- OCHA Services (2025) The humanitarian data exchange. Available at: <https://data.humdata.org>
- Office for National Statistics (2023) Dynamic Population Model, Improvements to Data Sources and Methodology for Local Authorities, England and Wales: 2011 to 2022. Methodology. Office for National Statistics. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/methodologies/dynamicpopulationmodelimprovementstodatasourcesandmethodologyforlocalauthoritiesenglandandwales2011to2022>
- Ostrom E (1990) *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge, UK: Cambridge University Press. DOI: 10.1017/CBO9781316423936.
- Pfeffermann D, Ben-Hur D and Blum O (2019) Planning the next census for Israel. *Statistics in Transition New Series* 20(1): 7–19.
- Poulain M, Herm A and Depledge R (2013) Central population registers as a source of demographic statistics in Europe. *Population* 68(2): 183–212.

- Roca T and Letouzé E (2016) La révolution des données est-elle en marche Implications pour la statistique publique et la démocratie. *Afrique Contemporaine* 258(2): 95–111. Louvain-la-Neuve: De Boeck Supérieur.
- Rowe F, Iradukunda R and Pietrostefani E (2025) Producing population-level estimates of internal displacement in Ukraine using GPS mobile phone data. In: 30th International Population Conference, Brisbane, Australia, July 2025. Available at: <https://ipc2025.popconf.org/abstracts/250655>
- Rufener M-C, Ofli F, Fatehkia M, et al. (2024) Estimation of internal displacement in Ukraine from satellite-based car detections. *Scientific Reports* 14(1): 31638. Nature Publishing Group.
- Ruggles S and Magnuson DL (2020) Census technology, politics, and institutional change, 1790–2020. *Journal of American History* 107(1): 19–51.
- Ruppert E (2009) Becoming peoples. *Journal of Cultural Economy* 2(1–2): 11–31. Routledge
- Ruppert E, Isin E and Bigo D (2017) Data politics. *Big Data & Society* 4(2): 2053951717717749. SAGE Publications Ltd.
- Sadowski J (2019) When data is capital: Datafication, accumulation, and extraction. *Big Data & Society* 6(1): 1–12.
- Salgado D, Sanguiao L, Oancea B, et al. (2021) An end-to-end statistical process with mobile network data for official statistics. *EPJ Data Science* 10(1): 20.
- Salganik MJ (2019) *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press. ISBN: 978-0691196107.
- Sanchez-Cespedes LM, Leasure DR, Tejedor-Garavito N, et al. (2023) Social cartography and satellite-derived building coverage for post-census population estimates in difficult-to-access regions of Colombia. *Population Studies* 78(1): 3–20.
- Schlicht L, Ledebur S and Echterhölder A (2021) Data at the doorstep. *Science in Context* 34(4): 411–421.
- Shibuya Y, Jones N and Sekimoto Y (2024) Assessing internal displacement patterns in Ukraine during the beginning of the Russian invasion in 2022. *Scientific Reports* 14(1): 11123. Nature Publishing Group.
- Statistics Canada (2022) Use of administrative data to impute non-responding households in areas with low response rates for the 2021 Census. Available at: <https://www12.statcan.gc.ca/census-recensement/2021/ref/imputation-eng.cfm>
- Statistics Canada (2025) Github repositories. Available at: <https://github.com/StatCan>
- Stats NZ (2023) Methodology for Using Admin Data to Count People in the 2023 Census. Wellington, New Zealand: Stats NZ. Available at: <https://www.stats.govt.nz/methods/methodology-for-using-admin-data-to-count-people-in-the-2023-census/>
- Sustainable Development Solutions Network (2015) Data for Development - A Needs Assessment for SDG Monitoring and Statistical Capacity Development. Available at: <https://sdgs.un.org/publications/data-development-needs-assessment-sdg-monitoring-and-statistical-capacity-development> (accessed 22 May 2025).
- Tatem AJ (2017) Worldpop, open data for spatial demography. *Scientific Data* 4(1): 1. Nature Publishing Group: 170004.
- Thorvaldsen G (2017) *Censuses and Census Takers: A Global History*. London: Routledge.
- Tomasson RF (1977) A millennium of misery: The demography of the Icelanders. *Population Studies* 31(3): 405–427. Routledge.
- United Nations (2014) Millennium Development Goals Report 2014. United Nations Publications.
- United Nations (2017) Principles and Recommendations for Population and Housing Censuses, Revision 3. Statistical Papers (Ser. M). New York: UN. Available at: https://www.un-ilibrary.org/population-and-demography/principles-and-recommendations-for-population-and-housing-censuses-revision-3_bb3ea73e-en
- United Nations Economic Commission for Europe (2024) Data Stewardship and the Role of National Statistical Offices in the New Data Ecosystem. Conference of European Statisticians Statistical Standards and Studies. United Nations. Available at: <https://www.un-ilibrary.org/content/books/9789213585542>
- United Nations Statistics Division (2025) Census Dates. Available at: <https://unstats.un.org/unsd/demographic-social/census/censusdates/>
- van den Brakel J and Bethlehem J (2008) Model-Based Estimation for Official Statistics. 08002, Discussion paper, 4 March. Centraal Bureau voor de Statistiek: Statistics Netherlands. Available at: <https://www.cbs.nl/nl-nl/achtergrond/2008/10/model-based-estimation-for-official-statistics>
- van Driel II, Giachanou A, Pouwels JL, et al. (2022) Promises and pitfalls of social media data donations. *Communication Methods and Measures* 16(4): 266–282. Routledge.
- von Oertzen C (2021) True to form: Media and data technologies of self-inscription. *Science in Context* 34(4): 439–458.
- Wang Y, Li J, Zhao X, et al. (2020) Using Mobile phone data for emergency management: A systematic literature review. *Information Systems Frontiers* 22(6): 1539–1559.
- Wastergaard H (1932) Contributions to the History of Statistics. P.S. King and Son Ltd Westminster London. Available at: <http://archive.org/details/dli.ernet.5741>
- Wellcome Trust and Economic and Social Research Council (2024) UK National Data Library: Technical White Paper Challenge. Available at: <https://wellcome.org/what-we-do/our-work/uk-data-library>
- Whitby A (2020) The Sum of the People: How the Census Has Shaped Nations, from the Ancient World to the Modern Age. Hachette UK. ISBN : 978-1541619333
- Wilkinson MD, Dumontier M, Aalbersberg I, et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(1): 160018. Nature Publishing Group.
- Yatsenko O (2024) Purpose of mobility data usage and risk assessment: POSMO ethical data market case. In: *IEEE 3rd Conference on Information Technology and Data Science (CITDS), 2024*, Debrecen, Hungary, pp.1–6. IEEE.
- Zuboff S (2015) Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30(1): 75–89. SAGE Publications Ltd.