

# Rethinking official statistics: A sociological perspective

Arman Bidarbakhtnia\*

*United Nations Economic and Social Commission for Asia and the Pacific, Bangkok, Thailand*

**Abstract.** Does the current dialogue on the development of statistical systems provide adequate scope for transforming official statistics to deliver their social role? Can statistical systems, as currently defined, provide opportunities to people and non-state institutions to influence “what” statistics and “how” should be produced and used? This paper provides a sociological framework to investigate these questions within a broader understanding of the social functions of official statistics as part of public statistics required for a democratic society.

Keywords: Sociology of statistics, information society, citizen data, data revolution, data governance

## 1. Introduction

Despite many progressive and people-centered data initiatives (e.g. data values,<sup>1</sup> inclusive data charter,<sup>2</sup> data for now,<sup>3</sup> and citizen data collaborative<sup>4</sup>), the key global action plans for the development of official statistics (i.e. Marrakech Action Plan [1], Busan Action Plan [2], Cape Town Global Action Plan [3]) do not advocate for actions aiming to increase citizens’ participation in the production of statistics. The prevailing international discourse on statistical development revolves around “evidence-based decision-making”. This narrative often confines “evidence” solely to data (or even official statistics) and “decisions” solely to choices made by officials for public action. This narrow focus restricts the broader societal impact of all types of evidence (statistical and non-statistical). Such language does not provide adequate scope to discuss and implement initiatives that broaden the mutual benefits of information and society.

The World Development Report 2021 sets an aspirational goal for national data systems to fully harness data development values. The report states that a pre-

requisite to such data systems is a new social contract around data, in which “people produce, process, and manage high-quality data” [4]. This paper highlights that establishing such data systems takes more than a whole-of-government approach toward data governance. It requires a whole-of-society approach, wherein both state and non-state actors have equal opportunities to contribute across the entire data lifecycle, from design to use. The evidence produced and used in this process goes beyond statistics and provides the informational basis required for a democratic society. The main objective of the paper is to explore whether the current understanding of the National Statistical Systems (NSSs) and official statistics, and the approach towards development and assessment of the NSSs provide sufficient room to pursue this ideal. Do we have adequate language and a relevant conceptual framework to support the global discourse on developing statistical systems that integrate people and non-state institutions into data governance structures?

There exists a well-established body of literature on the historical evolution of the mutual influence among statistics, state, and society [5,6,7,8]. Nevertheless, the sociology of quantification is an emerging field [9,10,11,12,13,14,15] and the language used to describe the social roles of data and evidence-informed governance in sociology and political science is unfamiliar to the statistical community and differs from the one used by

---

\*Corresponding author: E-mail: bakhtnia@un.org.

<sup>1</sup><https://www.data4sdgs.org/datavaluesproject>.

<sup>2</sup><https://www.data4sdgs.org/initiatives/inclusive-data-charter>.

<sup>3</sup><https://www.sdsntrends.org/datafornow>.

<sup>4</sup><https://unstats.un.org/UNSDWebsite/citizen-data/>.

statisticians when describing statistical systems [8]. If the NSS is meant to unleash the power of data<sup>5</sup> for better societies, beyond merely informing official decisions, the development of statistics becomes an interdisciplinary field requiring a common language rooted in human science (where statistics originated).

This paper is an attempt to establish a sociological framework<sup>6</sup> for discussing the social functions of statistics and the roles of statistical systems (Section 2). The framework provides for a fresh interpretation of statistics, and facilitates rethinking two fundamental questions (Section 3): What are official statistics for? and do national statistical systems, as currently defined, provide an adequate institutional framework for official statistics to deliver their social function?

## 2. Social functions of public statistics

### 2.1. A sociological framework

#### 2.1.1. Introductory

Durkheim's rule for sociological research<sup>7</sup> provides a good start for the autopsy of statistics. The most central concept in statistics (i.e. the parameter) can be described as an unknown "*thing*" that presents itself to an observer with a degree of uncertainty.<sup>8</sup> We can imagine many of the "things" (reflected in statistical parameters) experienced by ordinary members of society in the subjectively meaningful conduct of their everyday life [16] either as a result of their actions and thoughts or as circumstances imposed on them (e.g., job satisfaction, being a female, religious beliefs, illness, housing conditions, and air pollution). Or more complex "things" that society as a whole faces at the macro level, which could not be commonly understood without quantification (e.g. unemployment, poverty, growth). Statistical parameters aim to formalize such unpredictable subjective experiences or complex social realities into objects that can be predicted/understood with a known

degree of uncertainty. As such, the scrutiny of (public) statistics, statistical systems, and their social function is not possible without an understanding of the mechanism through which individual (subjective) and social (objective) realities are constructed.

A girl forced to quit school, work on a farm, lacking nutrition and leisure time with other kids can describe her deprivation in her own words (perhaps very different from how other girls and their mothers would describe their girlhood deprivation). Over generations and across social groups, a language is constructed and continues to evolve in describing the "deprived girl" reality in this society. In sociological terms, this is referred to as a constructed reality unique to a particular society [16]. The intertwined nature of individuals' everyday lives and the social realities continually shape (co-construct) one another through a reflexive relationship facilitated by a shared language [20].

#### 2.1.2. Overview of the framework

Figure 1 presents a model for reality construction I employ in explaining the fundament of statistical inquiry and the social function of statistics. The model consists of four types of social realities interacting with each other; subjective realities, constructed realities, social institutions, and complex realities.

Subjective realities refer to our experience of the lifeworld (how we interpret, understand, and engage with the world around us) which is perpetually shaped by a complex interplay of other realities that are socially constructed. The constructed realities are social objects (norms, values, identities, beliefs, culture, and history) constructed by the collective knowledge derived from dialogue and the exchange of subjective knowledge of people's everyday lives.

Institutions are the third type of social reality. The institutions (such as government, family, economy, legal system, and civil societies) are formal and informal mechanisms established based on the constructed realities and govern the relationship between social entities. The members of society and institutions are acting upon each other directly (e.g. by voting (bottom-up) or law-making (top-down)), or indirectly through constructed realities (e.g. by changing social norms (bottom-up) or making and implementing policies (top-down)).

The fourth type of reality is complex reality which is extremely important for understanding the role of statistics and defined by institutions (indirectly influenced by subjective and constructed realities). In contrast to other types of social realities, which are a construct of the natural inter-subjective exchange of knowl-

<sup>5</sup><https://www.data4sdgs.org/initiatives/power-of-data-unlocking-data-dividend-sdgs>.

<sup>6</sup>Inspired by theories of sociology of knowledge [16], communicative action [17], and Sen's capability approach [18].

<sup>7</sup>Durkheim [19] proposes to "consider social facts as things?".

<sup>8</sup>The uncertainty could be due to the observer's lack of knowledge about the truth (subjective view) or the inherent characteristic of the phenomenon (objective view). I have consciously avoided the philosophical controversy of objectivism versus subjectivism here as its relevance to our argument is trivial. Desrosières [5] thoroughly reviews different schools of thought and their technical implications.

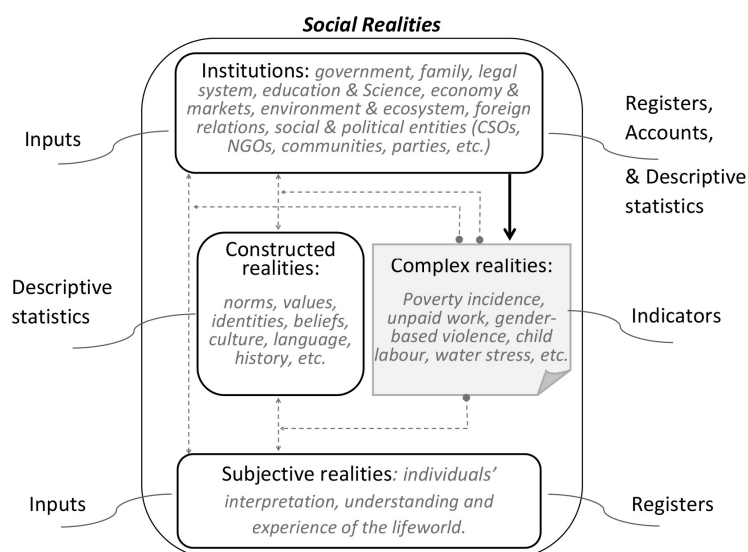


Fig. 1. A model for social reality construction and the role of public statistics. Note: There is a reflexive relation structure between subjective and constructed realities and institutions ( $\leftrightarrow$ ). Complex realities are defined ( $\rightarrow$ ) by institutions and must, in principle, reflect socially constructed realities. Complex realities influence the relation structure of realities through quantification ( $\star\rightarrow$ ).

edge and communicative socialization of entities [17], complex realities are delineated by institutions. These are products of human intellectual endeavor in modern life and aim to formalize the classification of social entities claimed (or believed) to be related to a specific phenomenon. Poverty is a socially constructed reality (without a need for direct intervention by institutions<sup>9</sup>). However, understanding the extent, pattern, and change in the prevalence of poverty requires a standard definition of poverty and a threshold (such as the poverty line) to determine who falls within the category of poor and who does not. The same applies to gender-based violence, air pollution, suicide rate, and adult literacy. All of these are concepts related to existing socially constructed realities but defined by institutions based on formalized classification systems.

### 2.1.3. Incorporating statistics

Statistical parameters, although not a precise reflection of social realities and are not expected to capture their full complexity, encapsulate features and dimensions of the realities that are socially co-constructed. Parameters provide us with standard language facilitat-

ing the quantification of specific attributes of the social phenomenon. Hence, parameters and their quantification (statistical production) are integral components of the model for constructing social reality.

Before incorporating statistics into this model, it is crucial to explain two concepts:

1. *Public vs applied statistics*: Statistical inquiry is not restricted to social realities. The application of statistical techniques in science typically deals with a different type of reality that is not socially constructed but is an inherent feature of the subject or a relationship between subjects or realities. I refer to this kind of reality as 'inherent reality'. Application of statistical techniques for analyzing clinical trials in developing coronavirus vaccines, aims to quantify the inherent realities associated with chemical and organic entities. However, statistics on deaths caused by coronavirus infection disaggregated by vaccination status contribute to the ongoing social debate and (re)construction of several social realities and institutions, such as death, public health, trust in science, and human-nature harmony. To distinguish between the two in this paper, the first one (quantification of inherent realities of subject/nature and relations) is referred to as '*applied statistics*' and the latter (quantification of the social realities in the public sphere) as '*public statistics*'. With

<sup>9</sup>Institutions (economy, government, education system, etc) contribute to its construction through public actions and communication. But their intervention is not necessary. Inter-subjective communicative action, transfer of knowledge over generations, and a complex interplay of social realities would naturally establish a directory of terms for a social dialogue about the reality of poverty.

this distinction,<sup>10</sup> I aim to focus on public statistics, using it as a benchmark for examining the frontiers and functions of official statistics and national statistical systems.

2. A simplified model (as illustrated in Fig. 2) outlining essential inputs and outputs in the production process of public statistics<sup>11</sup> is used to describe how public statistics fit in the social reality construction model.

*Inputs:* For statisticians, it is commonplace that individuals and institutions directly contribute to data and observations mostly as primary units of enumeration/measurement, or sometimes as enumerators. Nevertheless, the cognitive contribution of subjective and institutional entities to the production of statistics for public use is not widely recognized and neither is usually facilitated by statistical systems. For public statistics to be meaningful, underlying parameters and classifications must be defined to adequately reflect social realities [21]. Parameters can objectify subjective experiences and capture the essence of social realities only when they are grounded in the social stock of knowledge pertaining to the phenomena at hand. This is only possible when individuals and institutions have the agency to contribute to the production of public statistics. The cognitive contribution can manifest at least in three forms: at the design phase (e.g. defining reference population, formulating variables, defining parameters, decisions on reference time and periodicity of data collection and dissemination, and considering alternative data sources), during the instrument development process (e.g. questionnaire design, enumerator recruitment, and training), and in the form of technical contribution and quality audits (research, analysis, and use). Individuals can contribute through feedback loops, focused group discussions, public consultations, and co-production. However, these inputs are often channeled through institutions, such as academic societies, civil society organizations, and special interest groups. These institutions serve as conduits through which individual and collective

cognitive insights are streamlined into statistical design and production.

There are many examples wherein individuals or social groups, when granted the opportunity, have shaped not only “what” but also “how” the information about them is collected. Public consultations for the 2021 census of England and Wales started in 2015, with the participation of 279 organizations and 816 individuals [22]. Sexual orientation and gender identity were among the five topics that users and interest groups proposed to be added to the census and the Office for National Statistics (ONS) considered further research on the feasibility and impact of adding these questions to the census questionnaire. Despite initial concerns about the sensitivity of the questions and the potential impact on responses to other questions (what prevented adding the same topic to the 2011 census [23]), the pilot testing defied official assumptions. The new (voluntary) questions were well-received by the public and did not have a significant impact on the remainder of the census content. A similar story applies to the question about religious affiliation which minority groups leveraged their political negotiation power to re-introduce to the census in England and Wales in 2001, after 150 years [24]. This was not just another statistical table but signified a shift from officials determining how people are counted to people having the agency to decide how they wish to be counted.

*Outputs:* A coin has two sides. The cognitive exchange between statistics and society must be two-way and dialectical. The role of statistics in the construction of social realities is indispensable, binding different parts of society and creating social cohesion. It provides the informational basis essential for public reasoning and development [18]. Each of the four generic outputs of the statistical process illustrated in Fig. 2 holds a distinct mandate in contributing to the knowledge base for reality construction. It is important to note that social realities are co-constructed, with or without statistics, through the inter-subjective exchange of knowledge. However, quantification facilitates a shared understanding of the extent, change, and pattern in the social life of the concerned reality. It establishes a common language and reference point for meaningful public discourse that would otherwise have to rely on perceptions, opinions and claims.

<sup>10</sup>I acknowledge that this distinction is not a clear-cut dichotomy and there are many applications of statistics in science that could yield significant social benefits and potentially be viewed as part of public statistics too.

<sup>11</sup>For simplicity, this representation excludes many steps in the process such as field operations, processing, analysis, communication and use, focusing only on immediate inputs and outputs.

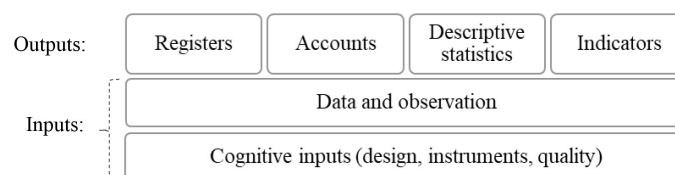


Fig. 2. Inputs and outputs in the process of producing public statistics. Note: Examples of statistical outputs are i) registers: business, population or address registers; ii) accounts: system of national accounts (SNA), balance of payments (BOP), or system of environmental-economic accounting (SEEA); iii) descriptive statistics: summary measures of variables such as median income, average household size, and total population; and iv) indicators: poverty rate, maternal mortality ratio, and CO2 emission per unit of value added.

The incorporation of statistical outputs into the Fig. 1 framework could be explained by the following example. Consider the four realities of child identity (subjective), marriage (constructed), family (institution), and child marriage (complex). Except for the last one, all these realities are naturally constructed in any society, without a need for statistics. However, different statistical outputs could significantly influence how social entities debate and act upon each of them. For instance, population registers are important mechanisms to provide legal identity to a child, which in turn qualifies him/her for access to public services such as education and health. Demographic statistics derived from population registers, censuses and surveys serve as common knowledge of family structure and dynamics. They also empower society with a nuanced understanding of trends and patterns in marriage based on real data. At this point, social realities have established a reflexive relationship (exchange of knowledge) with statistics and perhaps a social discourse on an appropriate age of marriage for girls has also emerged, albeit without “consensus”. For public statistics to be relevant to the discourse it must provide real knowledge on inter-related realities. In this case, cultural elements, religious beliefs, related laws, and economic factors are all important parts of the knowledge set needed for a meaningful dialogue about the “appropriate age of marriage for girls”. But without objectification and quantification of complex realities, any meaningful social discourse is impossible. It was in response to this ongoing discourse in modern societies that institutions started to objectify the intricacies of the issues by constructing complex realities and developed “indicators”, such as child marriage in this instance, to quantify these realities by way of parameterizing their complexities. The absence of statistics on these indicators would render debates about any complex reality a mere clash of subjective anecdotes and authoritative claims: a non-conclusive and non-cohesive discourse. It is especially the norm-setting feature of complex realities that highlights the distinctive role of indicators [25] in quantifying the quality of life.

## 2.2. Public statistics as social stabilizer

### 2.2.1. Quantification

Okrasa [26] lists four institutional contributions of quantification: public administration, democratic rule, economic audit and action, and datafication of everyday life. Our social reality construction model clearly adds a fifth element to the list, the cognitive contribution. Public statistics, and in particular indicators, provide a common reference point for meaningful public discourse, empower groups to negotiate their identity and rights recognition, and influence public policies.

The issue of identity provides a clear example of public statistics’ contribution to the social reality construction. An individual’s sense of self is constantly shaped by assigning significance to various memberships within different groups of society [27]. Statistical classification plays a vital role in officially recognizing and facilitating such membership prioritization. In contrast, the lack of statistics on memberships, attributes, and perceptions within social groups become political instruments for suppression of identities and entitlements, causing social divide and violence. Many social groups (e.g., children engaged in child labor, women and girls experiencing domestic violence, internally displaced populations, individuals trafficked, and ethnic and religious minorities) and institutions advocating for their rights would not be able to negotiate their entitlements without standard classifications and quantification of their attributes. In 2000, respondents were given the option to choose more than one race category from the list provided in the population census of the United States. It was clear from the beginning that this seemingly small change in the census questionnaire has far-reaching effects on the role of race and ethnicity in American society [28].

Recognizing the influence of classification on individuals and institutions and acknowledging that every quantification inherently involves some form of classification can shift our perspective on indicators. Instead of viewing them merely as sets of numbers used by au-

thorities for policy-making and monitoring, we can perceive them as essential sociological factors in making better societies. Starr [29] lists four different ways in which statistical classification could have a sociological impact on groups:

- a) *Domain definition*: To what extent do domains (such as ethnicity) truly exist? And to what degree does the classification merely represent the “state’s mapping of society”, driven by the need for simplification or political motivations?
- b) *Grouping*: Does classification group individuals together who consider themselves part of the same group?
- c) *Labeling*: What is the social response to the label associated with a group in classification?
- d) *Ordering*: Does classification apply across the entire society? To what extent do the defined boundaries between groups reflect the complex and fuzzy realities of social life?

Quantification involves three steps: *abstraction*, *parameterization* and *measurement*. Abstraction is the process of using human language to construct conventions that simplify complexities introduced by unpredictable subjective experiences of reality and find a common denominator that the majority of the subject individuals can connect with their experience. Parameterization is the act of expressing objectified reality in mathematical terms. It may involve formalization, standardization, normalization, codification, and classification. Finally, measurement consists of summarizing the variation in subjective experiences of the objectified reality in the form of statistics.

### 2.2.2. E-O-I mechanism

A woman has everyday experience as a member of society (subjective reality). Through generations, a *common language* is also created to describe roles, responsibilities, and expectations from the category of “women population” (a constructed reality) to which she can relate and use the *stock of knowledge* about the social norms and value system (other constructed realities), family and law (institutions) which she uses to relate to men, women and other groups in her society. However, some women in the same society experience violence, sometimes in different ways, which can be described in different words and not always referred to as violence. To objectify this common “thing” among all women subject to such phenomenon the complex reality of “violence against women” is created by in-

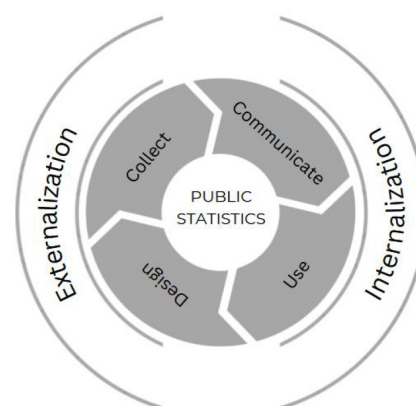


Fig. 3. Public statistics facilitates objective externalization internalization.

stitutions.<sup>12</sup> The model in Fig. 1 illustrates this process and depicts where statistical outputs contribute to it. Nevertheless, a deeper understanding of the mechanism through which statistics contribute to the formation of the common language and stock of knowledge is necessary to grasp the social function of public statistics as a key element in this process.

An individual member of society simultaneously externalizes her/his own being into the social world and internalizes it as an objective reality [16]. It is through the *externalization-objectification-internalization (E-O-I)* mechanism that individuals define their “self” as members of society and public statistics can only deliver its social mandate by facilitating E-O-I. Public statistics does this through quantification. And the individuals’ contribution to quantification must start from abstraction. It is through externalization that individuals contribute to the abstraction of social realities and through internalization that their identity is influenced by the constructed conventions. The extent to which individuals and institutions have the entitlement to engage in, and the capacity to influence the outcome of the abstraction process determines the degree of externalization [30]. Clearly, the dialectical nature of E-O-I and the imperative of representing social diversity in the process requires the full participation of society in the production of public statistics to ensure it is fit for purpose in facilitating the E-O-I.

As depicted in Fig. 3, the process of producing public statistics extends beyond mere data generation ma-

<sup>12</sup>At the global level, the reality was officially recognized in 1993 in a Declaration on the Elimination of Violence against Women ([https://www.un.org/en/genocideprevention/documents/atrocities-crimes/Doc.21\\_declaration%20elimination%20vaw.pdf](https://www.un.org/en/genocideprevention/documents/atrocities-crimes/Doc.21_declaration%20elimination%20vaw.pdf)).

chinery and concurrently serves as an enabler for the externalization and internalization of subjective and social realities. Considering four broad phases of the data value chain, design and data collection phases provide opportunities for the externalization of real-life experiences and communication and use facilitate an objective internalization of social realities. At the design phase, individuals and institutions contribute to abstraction (what to collect), while at the collection phase, they can contribute to decisions on data collection procedures and instruments (how to collect) or directly engage in the collection/reporting process. Additionally, any data collection is a channel for respondents to externalize their opinions, choices, and experiences, actively contributing to the dialogue for constructing social realities. Communication and use of public statistics enable citizens and other social entities to engage in an objective social dialogue on what matters to them. Public statistics work as a stabilizer for the social fabric through this process. Conversely, citizens' exclusion from the process denies their role in this social dialogue and undermines their externalization and internalization rights.

### 2.2.3. *Putting pieces together: The case of violence against women*

The issue of violence against women provides a good example to describe how the quantification process has worked in the social reality construction framework to facilitate a social discourse leading to recognition and change in an unfortunate reality. This complex reality would probably be constructed based on individual experiences and without statistical evidence. Nevertheless, without quantification, there is no reference point for meaningful dialogue, making consensus on the extent and nature of this phenomenon impossible. With statistics on standard and agreed indicators, and participation of citizens in the quantification process, women in society can better internalize and relate the social phenomenon to their personal experiences. The quantification of this complex reality facilitates the process of E-O-I as a social fabric and establishes a web of reflexive relations between individual experiences, social realities (values, traditions, laws, etc.), and institutions leading to a social object commonly understood as violence against women.

In this quantification process, women must be the primary source of knowledge for designing data collection methods, instruments, and standards that accurately reflect their everyday lives. They can provide cognitive input either directly, or through their representative in-

stitutions. Globally, only since 2000 have women had this opportunity to share their experiences through violence against women surveys after the global recognition of this ancient reality in 1993. Cognitive testing of these surveys has shown a great contribution of citizens in the design and conduct of the surveys [31]. It is hard to imagine the same level of success for global campaigns (such as 16 Days of Activism Against Gender-Based Violence, or #MeToo) without public statistics on violence against women. The centerpiece for these campaigns is the personal stories combined with statistics, both based on globally agreed definitions provided by quantification efforts.

## 3. Reimagining conventions

It is inspiring that the General Assembly resolution on the Fundamental Principles of Official Statistics (FPOS) [32] distinguishes between official statistical systems and national statistical systems (NSS). When it highlights the importance of public trust, it refers to an official system, but to underscore legislative and institutional requirements, it refers to NSS. Moreover, the first principle states that “*Official statistics provide an indispensable element in the information system of a democratic society, . . . to be compiled and made available . . . by official statistical agencies to honour citizens' entitlement to public information*”. The first principle carries three implicit points crucial to our following discussion: (a) the information system required for a democratic society expands beyond, yet inclusive of the official statistics, (b) official statistical agencies are part of, but not entire NSS, (c) official statistics are a part of public interest which are recognized and honoured by official statistical agencies. The public information system is the system that produces public statistics introduced in our sociological model. As such official statistics may also be explained by its function in the reality construction model and the national statistical system by its role in facilitating institutional and legal requirements for public and official statistics to contribute to the E-O-I process.

### 3.1. *What are official statistics for?*

The qualities of life, air, and oceans were once considered out of the scope of official statistics but are now being officially compiled and reported through surveys, remote sensing data, and accounts [33,34,35]. *The official statistics are a subset of public statistics*



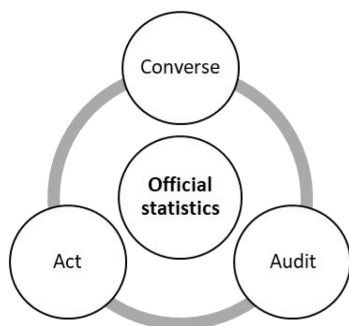


Fig. 4. Use of official statistics by individuals and institutions.

quantifying social realities that the state, institutions, and members of society have agreed to produce following officially endorsed standards and procedures. There are realities (like illegal economic activities) that members of society and private institutions have no incentive to share with the state. On the other hand, states may refrain from quantifying some other realities (such as public safety, public satisfaction, corruption, and public perception of politically sensitive issues). Quantification allows for detecting any deviation from reality and permits auditing the state's actions. It provides a reference point for social dialogue, mobilizes citizen agency, and creates social forces for or against change. The boundaries of official statistics are fluid and constantly redefined based on the new tripartite agreements (between state, institutions, and society) on the "realities to be officially recognized and quantified". This sociological understanding of official statistics, as opposed to *statistics being produced and used by official agencies*, has crucial implications for assessing, reviewing and transforming national and official statistical systems. Perhaps rarely is there an adjective used for decision-making more than evidence-based. As an official statistician, one should perceive this as a positive attribute, provided that the terms "evidence" and "decision" are applied in a broader social context. However, concerns arise when, in practice, *providing evidence for decision-making* is translated to *official statistics for official actions*.

As a subset of public statistics, official statistics carry similar features and functions Fig. 4. The first and foremost use of official statistics is in social reality construction through providing information basis for objective dialogue (converse). Secondly, members of society and institutions use them as reliable metrics and benchmarks for assessing the performance of themselves and one another in fulfilling commitments to adhering to social realities (audit). Thirdly, they are used for taking public and private actions upon realities

(Act). Decision-making by state is a public action [9], a policy instrument to act on and influence social realities for the public good. The "evidence-based decision-making" language describes an excellent quality for decision-making but provides a very narrow scope to discuss the development of official statistics and assess the performance of national statistical systems. It does not recognize the role of official statistics as a social stabilizer, leaves limited scope for private and institutional action, and only acknowledges self-audit by the state. While publicly accessible official statistics are crucial for participatory auditing and auditable governance as integral components of democratic societies.

In this sociological definition of official statistics, official agencies only have an executive role in the quantification of social realities, but not the full authority over reality construction. In an ideal scenario where members of society and institutions have the agency to provide cognitive inputs to the production of statistics, the boundaries (what) and the production and use (how) of official statistics are defined through a communicative process. On the contrary, in authoritarian governance systems, official statistics reflect the official interpretation of reality, providing a monopoly over the social narratives. And evidence-based decision-making by the state, if used to describe the primary role of official statistics, could become a facilitator for data tyranny as opposed to information society.

### 3.2. Are the national statistical systems fit for their purpose?

An immediate implication for sociologically defined official statistics is that it poses a challenge to the existing conception of the NSS. According to the United Nations definition, Official statistics are constrained to what is produced (not only used) by state and state-certified agencies, and membership in the NSS is similarly confined to the same agencies [36]:

- Official Statistics: Statistics produced in accordance with the Fundamental Principles of Official Statistics by a national statistical office or by another producer of official statistics that has been mandated by the national government or certified by the national statistical office to compile statistics for its specific domain.
- NSS: Comprises the national statistical office and all other producers of official statistics in the country.



These diverge from the FPOS and clearly provide a framework inadequate for official statistics and the NSS to deliver their social roles. To fulfill the first principle, a distinction must be made between two pairs of inter-connected concepts: official vs national statistical system and official vs public information.

A NSS must be established to produce public information for society. The official statistics, as an element (subset) in public information, are compiled and made available by a sub-system, an official statistical system, consisting of official agencies. As with any other “democratic” institution, the NSS must provide agency (and not only facilitate participation, as mostly recommended/encouraged) to all non-state institutions and members of the public to practice their rights for cognitive contribution to the design of public statistics.<sup>13</sup> The agency will only be guaranteed if they have formal membership in the system with formal procedures for their contribution. As is currently defined, members of society and non-state institutions will only be able to contribute when official agencies recognize their role and facilitate their participation (as if they do a favour). As a result, the approach often taken by the international community is to encourage the NSS (the governments) to engage with the private sector, academia, civil society, and the public, as “stakeholders”. In essence, the NSS is not considered a contributing factor to democratic governance; rather, it is hoped to be governed democratically when the state “happens to be” democratic. Two immediate implications of the current definition of the NSS are evident in frameworks developed for assessing the performance of the NSS ([37] and Statistical Capacity Monitor<sup>14</sup>) and the guidelines for developing National Strategies for Development of Statistics (NSDS).<sup>15</sup> In both, non-state players are recognized as stakeholders to be consulted and mostly as “users”. On the production side, if recognized at all, they are listed as providers of “new data sources” (e.g. big and citizen data) which the NSS (the government) is encouraged to embrace, and never as formal contributors to the design and production of conventional official statistics. Therefore, NSSs are not currently assessed against their role to provide the informational basis necessary for the democratic and objective co-construction of social realities, but against their mandate to “provide official statistics for evidence-based decision-making”. In other words, the international community is primarily con-

cerned with the development of official (and not national) statistical systems. And this must be the starting point for a paradigm shift. Bridging the gap between citizens and official statistical systems is not merely an issue of communication and stakeholder engagement [38]. The need for public involvement in the co-design and co-production of public statistics [8] calls for a different solution, a complete shift in the statistical system and epistemic understanding of public statistics and its social function. The European Statistical System provides a good example where practical steps are taken to facilitate public engagement in the co-production of statistics.<sup>16</sup>

### 3.3. Has the revolution transpired?

The concept of “data revolution”, introduced to the political discourse in 2013 by High-Level Panel of Eminent Persons on the Post-2015 Development Agenda [39], is often understood as a rapid growth in volume, speed, diversity of sources, and advanced data applications. One could assess if this change could be considered a revolution or is merely a data explosion or a rapid evolution in the data landscape as a natural feature of the modern world [40]. But regardless of the term used, the question is whether the statistical systems have transformed enough to mobilize the new data landscape for more transparent and accountable governance, empowered people, and better societies. Alex Cobham speaks about two aspects of what is called data revolution: technical and political aspects [41]. He argues that most of the change has taken place on the technical side but to complete the revolution, we need a political transformation:

- ‘The data revolution that is needed must be a revolution indeed: not (just) a gradual process of dealing with technical difficulties, but a radical and deeply political challenge to the power structures that lie behind the uncounted’ [41, p. 58].

The current conception of the NSS and official statistics does not seem to provide adequate language for international discourse on the actual change needed in data politics. To shift political discourse, a real revolution is needed to replace the “data for decision-making” paradigm with “data as social stabilizer”. A sociological model for statistical development could turn our utopia upside-down.

<sup>13</sup>Public statistics and public information are used interchangeably.

<sup>14</sup><https://statisticalcapacitymonitor.org/indicator/>.

<sup>15</sup><https://nsdsguidelines.paris21.org/nsds-lifecycle>.

<sup>16</sup><https://ec.europa.eu/eurostat/web/european-statistical-system>.

Take citizen data as an example. At the conceptual level, it is a progressive initiative taken by the statistical community. Nevertheless, due to mostly systemic challenges<sup>17</sup> the focus at first steps may be narrowed only to the contribution of data generated by citizens to complement official statistics.<sup>18</sup> This is mainly because our NSSs do not provide a fair playground for both state and non-state players. They are, in practice, facilitating the production of official (not public) statistics by official or officially certified agencies (not recognizing membership of non-state actors). A new understanding of statistics as the informational basis for democratic deliberation and public reasoning [42] could transform NSSs from embracing citizen data to enabling “informed citizen” and “information society”. An information society is not merely a society with open access to data or a society in which citizens are generating data and waiting for the official system to embrace it. It is a society that co-creates information and knowledge and co-constructs its shared realities based on informed social discourse. In this dialectic process, humans need agency in shaping and changing social realities Fig. 1 and it requires creating opportunities to participate in the distribution and creation of the social stock of knowledge through the production and use of public statistics. An information society uses evidence to *converse* (determine what meanings are ascribed to social realities and what is important and relevant to quantify), *audit* (the ability to objectively assess the sufficiency and efficacy of public actions in addressing important issues), and *act* (influence (re-) construction of social and subjective realities) in the public sphere and not only demand and hope for evidence-based decision-making by the state. Unless the NSSs are designed to foster an information society, we will continue living with cognitive ambiguity and factual ignorance around the realities of our societies [43].

#### 4. Conclusion

The discipline and language we employ to describe a phenomenon influence how we conceptualize and understand it and formulate related problems and solutions. The first fundamental principle clearly positions

official statistics as social elements vital for fulfilling public interest. A meaningful discourse on the development of official statistics requires an interdisciplinary conceptual framework in which official statistics are more than a product of a state-led data generation machinery but an integrated part of the social process of reality construction. This paper has proposed a sociological framework to describe the interaction between statistical outputs and the process of constructing social realities. Examining the conventional views about statistical systems and official statistics within the proposed framework reveals a deficiency in both language and relevant conceptual framework for transforming statistical systems that grant agency to people and non-state institutions for co-creation and cognitive input into data governance structures. The preceding discussions could have at least three practical implications:

##### i. Statistical systems

National Statistical Systems, as currently defined, do not provide the required institutional environment for unlocking the power of statistics for public action. Given their restricted membership, they may be called official statistical systems at best. Even for the latter, the state-led governance structure of the systems poses a serious challenge to fulfilling the FPOS. Public statistics, if seen as social constructs, call for statistical systems that provide formal membership to non-state social institutions; a system broader than, but interlinked with the official statistical system.

##### ii. Statistical business process

The sociological framework broadens our perspective of the uses of official statistics. It shifts the focus from “evidence-based decision-making” (mostly by the public sector) to the use of official statistics for establishing objective social dialogue on what matters (*converse*), forming a participatory auditing and auditable governance (*audit*), and adapting to and influencing the construction of social realities (*act*). This would not be achieved without the systematic inclusion of social institutions in the data lifecycle. From defining the data need, to cognitive input to the design, data collection and exchange, and the data analysis and use, the statistical business process should provide standard mechanisms for the participation of non-state actors and facilitate all three types of data utilization.

##### iii. Global discourse

Official statisticians can't run the global agenda for statistical development single-handedly and

<sup>17</sup>Report of the Expert Group in 2022: [https://unstats.un.org/sdgs/files/meetings/harnessing-data-by-citizens-for-public-policy-and-SDG-monitoring/CDG\\_EGM\\_report\\_final\\_public.pdf](https://unstats.un.org/sdgs/files/meetings/harnessing-data-by-citizens-for-public-policy-and-SDG-monitoring/CDG_EGM_report_final_public.pdf).

<sup>18</sup>Report of the Expert Group in 2023: <https://unstats.un.org/UNSDWebsite/capacity-development/events-details/650>.

expect the systems to work for a broader society. As an interdisciplinary topic, issues of statistical development must be addressed in collaboration with other science communities, particularly social and political science. What may be identified in statistical terms as a “*lack of capacity and resources in the statistical system to respond to the demand for data*”, may be described by experts from other fields as “*members of society partially denied of their rights to be heard (externalize their subjective stories), and adapt (internalize social objects), leading to social disconnect and instability*”. The two groups would come up with different solutions for what we understand to be the same issue. In the same way, the conceptual framework and language we use could impact the evaluation of statistical systems, implementation of the FPOS, quality assessment [44], data governance, and many initiatives taken in response to the data revolution.

One obvious example is the implementation of the SDG indicators. It is striking that after over eight years and despite remarkable progress made in filling the data gap, still indicators related to individual safety (crime, violence, and discrimination), public trust in governance (satisfaction, trust, participation, and corruption), and human rights (gender equality, access to justice, etc.) are among indicators with the largest data gap [45]. Understanding the role of indicators in the quantification of complex realities and providing a common reference for democratic deliberation and public reasoning (fulfilling the first FPOS) could significantly impact the approach toward adopting global indicators at the national and sub-national levels. In assessing the relevance of indicators, the focus could shift from the state priorities and official demand for data to the information requirements for social interaction on what society believes to be important. In the process of indicator selection, a sociological mindset could redirect our attention from the size of the indicator set and the correlation between indicators (which inspires reductionist approaches that aim to reduce the list to a minimum set of indicators that explain the most variation) to the fact that each indicator facilitates objectification of a social reality. Given the limited resources faced by statistical systems, it is natural and practical to prioritize a set of indicators. Nevertheless, the selection criteria may vary, for instance, from the

extent to which an indicator contributes to the total variance (purely statistical approach) to the cognitive ambiguity and social disconnect caused by the lack of an indicator (sociological perspective). Striking a balance would certainly require a broader conceptual framework, a new language, and interdisciplinary collaboration.

## Acknowledgments

I gratefully acknowledge the valuable feedback provided by four reviewers, which greatly improved the quality of this paper.

## References

- [1] PARIS21. The marrakech action plan for statistics: better data for better results- an action plan for improving development statistics. 2004.
- [2] PARIS21. Statistics for transparency, accountability, and results: A busan action plan for statistics. 2011.
- [3] United Nations. Cape town global action plan for sustainable development data. 2017.
- [4] World Bank. World development report 2021: Data for better lives. Washington, DC. 2021. doi: 10.1596/978-1-4648-1600-0.
- [5] Desrosières A. The politics of large numbers: A history of statistical reasoning. Cambridge, MA: Harvard University Press. 1998.
- [6] Hacking I. How Should We Do the History of Statistics? In *The Foucault Effect – Studies in Governmentality*, ed. Graham Burchell, Colin Gordon, and Peter Miller. Chicago: University of Chicago Press. 1991.
- [7] Porter TM. *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ, Chichester: Princeton University Press. 1995.
- [8] Radermacher WJ. *Official statistics 4.0. verified facts for people in the 21<sup>st</sup> century*. Cham, Switzerland: Springer. 2020.
- [9] Desrosières A. Words and numbers. For a sociology of statistical argument. In *The mutual construction of statistics and society*, Ann R. Saetnan, Mork Lomell and Svein Hammer, ed., London: Routledge. 2011; 41-63.
- [10] Diaz-Bone R, Emmanuel D. The sociology of quantification – perspectives on an emerging field in the social sciences. *Historical Social Research*. 2016; 41: 7-26.
- [11] Demortain D. The politics of calculation: Towards a sociology of quantification in governance. *Revue D’anthropologie Des Connaissances*. 2019; 13(4): 973-990.
- [12] Espeland WN, Mitchell LS. A sociology of quantification. *European Journal of Sociology, Archives Européennes de Sociologie/Europäisches Archiv für Soziologie*. 2008; 49: 401-36.
- [13] Henneguelle A. Socio-economics of quantification and value: the perspective of convention theory. In: Diaz Bone R, de Larquier G, (eds). *Handbook of Economics and Sociology of Conventions*. Springer, Cham.; 2022.
- [14] Saltelli A, Di Fiore M. From sociology of quantification to ethics of quantification. *Humanit Soc Sci Commun*. 2020; 7: 69.

- [15] Hacking I. *The social construction of what?* Cambridge, Mass: Harvard University Press. 1999.
- [16] Berger PL, Luckmann T. *The social construction of reality; a treatise in the sociology of knowledge.* Garden City; N.Y, Doubleday; 1966.
- [17] Habermas J. *The theory of communicative action Vol. I: Reason and the rationalization of society.* Translator in English Thomas Mccarthy, Boston: Beacon Press. 1984.
- [18] Sen A. *Development as freedom.* Oxford: Oxford University Press. 1999.
- [19] Durkheim É. *The rules of sociological method.* Translated by Solovay Sarah A, Mueller John H. Glencoe, IL: The Free Press. 1950.
- [20] Nguyen CT. *The limits of data.* *Issues in Science and Technology.* 2024; 40(2): 94-101. doi: 10.58875/LUXD6515.
- [21] Norwood JL. *Data policy and politics in a democracy.* *Journal of Economic Education.* 1994; 25: 213-217.
- [22] Office of National Statistics (ONS). *The 2021 census assessment of initial user requirements on content for England and wales response to consultation.* 2016.
- [23] Office of National Statistics (ONS). *Sexual orientation and the 2011 census – background information.* 2006.
- [24] Southworth JR. *Religion in the 2001 census for England and wales.* *Population, Space and Place.* 2005; 11 (2), 75-88. Cross-ref. ISI.
- [25] Radermacher WJ. *Guidelines on indicator methodology: A mission impossible?* 2021 Jan 1; 205-217.
- [26] Okrasa W. *Sociological aspects of the statistical research process: Toward a sociology of public statistics.* *Polish Sociological Review.* 2020; 211(3): 323-344.
- [27] Sen A. *Identity and violence: The illusion of destiny.* W W Norton and Co. 2006.
- [28] Perlmann J, Waters MC. *New Race Question, The: How the census counts multiracial individuals.* Russell Sage Foundation. 2002.
- [29] Starr P. *The sociology of official statistics.* In Alonso W and Starr P, (Eds.), *Politics of Numbers,* The Russell Sage Foundation. 1987; 7-58.
- [30] Desrosières A. *The economics of convention and statistics. The paradox of origins.* *Historical Social Research.* 2011; 36(4): 64.
- [31] Paletta A, Karen M. *Cognitive testing of questions to measure family violence.* Ottawa: Statistics Canada. 1998.
- [32] United Nations Statistics Division [Homepage on the Internet]. Available from: <https://unstatsun.org/unsd/dnss/gp/fundprinciples.aspx>. [Accessed 15 March 2024].
- [33] OECD. *How's Life?: Measuring well-being.* OECD Publishing; 2011. doi: 10.1787/9789264121164-en.
- [34] Shaddick G, et al. *Data integration model for air quality: a hierarchical approach to the global estimation of exposures to ambient air pollution.* Royal Statistical Society. 2016. arXiv: 1609.0014.
- [35] Colgan S. *A guide to creating core ocean GDP accounts.* Global Ocean Accounts Partnership; 2022.
- [36] United Nations. *Handbook on Management an Organization of National Statistical System: 4th Edition of the Handbook of Statistical Organization.* 2022.
- [37] Dang HAH, Pullinger J, Serajuddin U, Stacy B. *Statistical performance indicators and index: A new tool to measure country statistical capacity.* *Sci Data.* 2023; 10: 146.
- [38] Eyraud C. *Stakeholders involvement in the statistical value chain: Bridging the gap between citizens and official statistics.* *Power from Statistics: Data, Information and Knowledge. Statistics in the Digital Era;* 2017.
- [39] United Nation. *A New Global Partnership: Eradicate poverty and transform economies through sustainable development – The Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda;* 2013.
- [40] MacFeely S. *In search of the Data Revolution: Has the official statistics paradigm shifted?* *Statistical Journal of the International Association of Official Statistics.* 2020; 36(4): 1075-1094. doi: 10.3233/SJI-200662.
- [41] Cobham A. *The Uncounted.* 1st ed. Wiley; 2020.
- [42] Sen A. *Human rights and capabilities.* *Journal of Human Development and Capabilities.* 2005; 6(2): 151-66.
- [43] Salais R. *Deliberative democracy and its informational basis: What lessons from the capability approach.* SASE (Society for the Advancement of Socio-Economics) Conference. Paris, France. 2009 Jul.
- [44] Diaz-Bone R, Horvath K. *Official statistics, big data and civil society. Introducing the Approach of “Economics of Convention” for Understanding the Rise of New Data Worlds and Their Implications.* 2021 1 Jan; 219-228. doi: 10.3233/SJI-200733.
- [45] United Nations [Homepage on the Internet]. *Global SDG database;* 2024. Available from: <https://unstats.un.org/sdgs/dataportal/analytics/DataAvailability>. [Accessed 14 March 2024].