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

Rainer Diaz-Bone* and Kenneth Horvath
Department of Sociology, University of Lucerne, 6002 Lucerne, Switzerland

Abstract. The rise of big data and ongoing political and social transformations confront official statistics with important questions regarding its self-understanding and its role in public debates. These questions imply serious tensions that will very likely increase in the foreseeable future. This article introduces a specific sociological perspective for thinking and talking about these developments. Building on the "economics of convention", this perspective challenges currently dominant conceptions of official statistics which do not adequately mirror the plurality of possible representations of the social world and the variety of justifiable ways of assessing the quality of these representations. Taking these pluralities into account allows to develop a fuller picture of the actual practices and institutions involved in the production of statistical knowledge and, especially, of their unavoidable entanglement with normative orders, epistemic values, and political formations. The notion of "data worlds" is presented as a means for tackling this problem of pluralities. On this conceptual basis, it becomes possible to link methodological questions to analyses of how statistical data and knowledge are embedded in wider political, economic, and social contexts. Problems of "data quality" thus appear in a different light: their reflection involves more than the usually discussed issues of the institutional independence (secured by public funding and by law) and the high scientific standards of official statistics. Instead, an institutionalist theoretical approach is needed that offers blueprints for linking the production of "official statistical facts" to (always contested and contextual) conceptions of the common good. Such a conception would allow to conceive new forms of public participation and democratic control of processes of quantification, measurement, and datafication. In sum, we believe that the specifically sociological approach outlined in this article would support official statistics in dealing with the variety of critical interventions and challenges it currently faces in a proactive and coherent manner.

Keywords: Economics of convention, big data, data worlds, data quality

1. Introduction

Following up on the questions raised by Radermacher [20], this article asks what role sociology might play for understanding the challenges that official statistics faces in times of big data and profound social and political transformation. It proposes a specific theoretical perspective for thinking about the nature and role of statistics. The key characteristic of this perspective is that it discusses methodological issues and questions of data quality in their interplay with political, economic, social, and moral contexts of data and knowledge production. In contrast to still widespread positivist understandings, the approach we introduce starts from the assumption that statistical facts are always intimately interrelated with norms and values. The numeric representation of social realities cannot be separated from normative orders and political values. Norms and values themselves need to be treated as "facts": they ef-

*Corresponding author: Rainer Diaz-Bone, Department of Sociology, University of Lucerne, Frohburgstrasse 3, 6002 Lucerne, Switzerland. E-mail: rainer.diazbone@unilu.ch.

fectively inform the definition of indicators as well as measurement strategies, the assessment of data quality, or the analysis and presentation of findings. In other words: Both the production and the evaluation of statistical data necessarily depend on epistemic and political *conventions*. These conventions provide rules and criteria for producing and evaluating (statistical) knowledge about the social world that qualifies as trustworthy and relevant information in a given societal context. At the same time, these conventions are the outcome of socio-historical developments, scientific debates, and political negotiations; accordingly, they can always be contested, criticized, transformed, or (in the long run) replaced.

It is exactly such a kind of "dispute" that official statistics currently faces in relation to big data and commercial data production on the one hand and civil society actors on the other. This dispute has many faces and takes different forms. It is evident in alleged revolutions of social knowledge production that proclaim the coming of a new age of data-driven social physics (Pentland, 2015) as well as in open science initiatives[1] or in the activities of "data NGOs" that lobby for the usage of new data sources to address pressing social issues.[2] These disputes are not necessarily hostile. To the contrary, the initiative to discuss coming challenges for statistical knowledge production often comes from within the field of public statistics itself (as is evident from the very SJIAOS discussion that this article contributes to).

In the following, we introduce the concept of "data worlds" as a means for thinking and talking about these developments. This notion shall allow to investigate how data infrastructures (historically formed cognitive and organizational frameworks and resources that structure the production, distribution, and usage of statistical data) are anchored and embedded in social rationalities ("conventions") and for understanding how data are produced, analyzed, evaluated and, eventually, applied as an informational resource for collective action. The concept of "worlds" is informed by the approach of "economics of convention" (EC) which emerged in France in the mid-1980s with the involvement of sociologists and economists (many of whom actually worked at the "Institut national de la statistique et des études économiques"/INSEE) [10]. EC scholars have promi-

nently applied the concept of "worlds", be it in earlier works to problems of economic coordination [25], be it later on to questions of methodological coordination and data production in survey worlds [11,26]. In an EC understanding, different worlds of data production vary in their methodological cultures, their epistemic values, their quality criteria and their collective understanding of and relation to the "common good". As will be discussed in more detail below, this implies that official statistics marks only one of a plurality of co-existing ways of producing and using numeric representations of social phenomena. We argue that official statistics is increasingly questioned by two more recent data worlds which we refer to as the big data world and the civil data world. We present a framework for making sense of the relations and tensions between these three data worlds. The intention of this framework is not to criticize these different data worlds or even to advocate one of them at the expense of the others. Rather, the aim is to offer a basis for thinking about how these different data worlds are currently becoming related to each other, what tensions might follow from these interplays and what social and political implications these new constellations might have.

2. Economics of convention

The French approach of economics of convention offers a contemporary conceptual framework that is closely tied to the field of quantification and official statistics [4,9]. One of EC's birth moments actually lied in the empirical analysis of official statistics and early representatives of EC such as Alain Desrosières or Laurent Thévenot were affiliated to the French national statistical institute INSEE. EC is informed by a neopragmatist understanding of science: statistical facts are seen as the result of concrete practices of knowledge production that take place in given institutional contexts and are structured by existing orders of (scientific, political, cultural …) knowledge which they build on, contribute to, and transform. Correspondingly, any over-simplifying understanding of data as pure representations of "social facts" is rejected. Instead, data are regarded as resulting from a complex interplay of conventions, actors, and technologies. One of the notions used to decipher these social processes of data production is the concept of "statistical chains" [5,8]. Statistical chains are best conceived of as institutionalized social processes which allow to generate data. In this sense, data need to be seen as so-

---

[1]See for example the Data Justice Lab at https://datajusticelab.org/ [retrieved 23/10/2020].

[2]See for example the https://flowminder.org/ [retrieved 23/10/2020].

cial artefacts and social constructions [20]. Desrosières pointed to the many tensions which arise in and along such statistical chains because of the different views that the involved actors hold on data and their quality. "Methodological statisticians" are usually highly aware of these processes and of the resulting "conventional nature" of data. However, many users of statistical data – including media experts, "subject-matter-specialists", politicians and the public – expect data that "hold together", data that can be treated as true and unambiguous "facts" [7]. Thus, contradictory understandings of the ontology of data are confronted with each other, with widely prevalent expectations to "transform" data from its actually construed and conventional foundations into true and indisputable representations of social reality [7,9].

The concept of statistical chains points to a decisive point. Statistical data are based on a division of labor between actors involved in the definition of categories, actors and technologies mobilized for and in measurement processes, situations in which data are distributed, communicated, analyzed and interpreted and decisions that are taken on the basis of statistical data. These different actors and stages involved in data production and usage are all involved in their own social and institutional contexts, engaged in different sorts of situations, and faced with different kinds of problems and challenges. As a correlate to the idea of chains, we may hence think of these various actors as embedded in their respective "worlds". These worlds are "data worlds" [11] in and through which the production and interpretation of data evolves along the statistical chain.

Following Desrosières [6], the production of quantitative data in and across these different worlds is only possible on the basis of conventions of measurement. The definition of such conventions of measurement always happens in concrete social circumstances and is always deeply interrelated with political and normative orders. What counts as relevant, acceptable, and fair process of quantification depends on social and political contexts. Categories and concepts that inform statistical measurement are anchored in public and political discourses and are always linked to specific forms of defining and understanding uncertain and tension-ridden situations. Salais et al. [22] give the example of the measurement of unemployment to demonstrate how institutions of official statistics (and the categories they employ) co-evolve in parallel with other social institutions, in this case the industrial organization of labor relations. The emergence of long-term industrial labor contracts allowed employees and new social classes to develop the expectation that employers provide permanent work and salary. As a complement to this expectation, unemployment as a new social category emerged, a category which actors referred to in order to make sense of their own situation in case they lost their industrial employment. In the course of the 19th century, official statisticians step by step included this category into their procedures of data production by implementing conventions for categorizing human beings as unemployed [22]. Unemployment, in other words, was never a "natural state of being" but a new way of interpreting a novel kind of biographical and social situation. Only by being routinely encoded by official statistics, it turned into a social fact. It is in this sense that EC scholars argue that statistical categories and indicators are expressive not only of social reality, but also of moral and political orders. Facts and values, in other words, are always already intertwined in statistical data.

The identification and exploration of this pragmatic, yet inevitable link between data production on the one hand and political and social normativities on the other may be seen as the crucial contribution of EC scholarship. Conventions are involved in the statistical chain at all stages. They serve as logics of coordination (e.g. between different actors along the statistical chain) and, at the same time, as criteria for evaluation (e.g. of the adequacy and validity of indicators and categories). EC hence applies a normative position to the question of data generation, its application and interpretation. The meaning of data as well as their concrete value for collective action are structured by conventions. Without conventions, data would be just meaningless figures without any anchoring in criteria and logics for assessing their validity, quality and relevance.

The need for a "conventional underpinning" of statistical measures entails that the data produced in order to represent social realities be related to commonly accepted understandings of the social world and established conceptions of justice and the common good. The crucial point again is that there are several different such conceptions that are widely accepted as legitimate and relevant in modern societies; in EC terminology, there is a plurality of "orders of justification". EC scholars have identified numerous such overarching conventional orders that are commonly established in current societies as basis for governing, justifying, and criticizing the social world and how it is represented in statistical information [2,3,14]. Table 1 presents the eight most important of these conventions. The key point is that each of these conventions has a history (it has been developed and established over time) and is mirrored

Table 1
Eight important conventions

Convention	Worth/quality	Evaluation criteria	Information format	Persons' qualification	Interpersonal relation
Domestic	Tradition, handcraft	Esteem, reputation	Oral, exemplary	Authority and flexibility	Trust
Market	Demand-orientation, free exchange	Price	Money units	Desire, purchasing power	Exchange
Industrial	Planning and standardization	Efficiency, productivity	Measurable criteria, statistics	Professional, expertise	Functional link
Inspired	Grace, nonconformity, creativity	Originality, innovative capacity	Newness, emotionality	Creativity, ingenuity	Passion
Opinion	Renown	Amount of recognition	Semiotic	Celebrity	Recognition
Civic	Collective interest	Relevant for collectivity	Formal, official	Equality	Solidarity
Green	Ecology (its integrity)	Environmental compatibility	Narrative	Ecological knowledge	Responsibility
Network	Activity, self-management	Successful projects	Meetings	Capacity for teamwork	Project orientation

Boltanski and Thévenot [3], Boltanski and Chiapello [2], Lamont and Thévenot [14], Diaz-Bone [10].

in everyday practices, processes and institutions. For example, the domestic convention is deeply anchored in our understanding not only of family life, but is also mirrored in understandings of how companies or other organizations should care for their employees, take responsibility for the community, and best be built on a kind of natural trusted authority. The industrial convention, in contrast, focuses on the role of standardization and quantification to ensure fair, transparent, and efficient procedures. These different conventions can be differentiated analytically; in real life, actors and institutions face the challenge of combining elements from various conventions and of finding compromises between tensions that arise between these worlds.

These conventions can be regarded as "always already present" cultural resources which structure institutions, including data infrastructures or the different worlds in which data are generated, analyzed and used. Since they have been developed, established, and implemented over the course of decades and centuries, different conventions may be more or less prominent depending on the concrete political and social circumstances. Their relative relevance may vary – both across historical periods and between different political and social contexts (e.g. across the West or between the global North and the global South). But overall, these conventions serve as globally available repertoire of evaluative logics that actors or institutions can draw on when they face the need to legitimize or criticize the relevance and validity of statistical data. Every statistician has an implicit recognition of the presence of this plurality of conventions as normative "substructure" for the data generating process, when acknowledging that there is a legitimate and adequate "definition" of categories, that there is self-evident "space for interpretation", that data are always (only) more or less appropriate for the

representation of a social issue at stake, not because of a lack of accuracy but rather, for example, because of a lack of relevance for collective and public action.

It is evident that the industrial convention is of outstanding importance for *any* modern institution, situation or "world" in which numerical information is generated and used on a wider basis. But the industrial convention will usually be forged into (sometimes complex) compromises with other influential conventions. This results in changing roles and meanings of the industrial rationality depending on the different compromises that are formed in a given situation. In many scientific contexts, the industrial convention and the convention of inspiration are the most important quality conventions used to legitimize data quality. Other organizations such as official statistics institutes often have to mobilize further conventions to demonstrate their societal legitimacy and relevance. In this regard, the civic convention and the convention of opinion have become more and more important for official statistics, to justify the need for independent official statistics in contemporary societies. This development is mirrored in debates on the future of public statistics. For example, the nascent notion (and job profile) of "information stewards" can be read both as a sort of "interface" between different data worlds and as a "conventional compromise" that combines elements from different worlds, including the market world and the network world.

3. The plurality of data worlds and the rise of tensions

In empirical data worlds, all of these eight quality conventions are (even if "virtually") present as logics

that actors can draw on to design and implement research infrastructures and to successfully perform tasks of coordination. Even if in most empirical cases the number of conventions that become effective is small, quality conventions are nonetheless usually combined; it is their interplay that marks given hegemonic institutional logics. Different such "coalitions" of quality conventions enable and structure different data worlds. By implication, these data worlds need to be seen as the outcome of historical developments. To a certain degree, they are always arbitrary; other configurations would have been possible and could have evolved.

3.1. The world of official statistics

The data world of official statistics has developed in Western societies from the beginning of the industrial era onward, roughly over the past two centuries. Modern institutions of official statistics were designed as dispositives for the governance of populations and societies. Nation-states invested in the cognitive form of numerical data in order to produce knowledge on their territory and their population, on national wealth and health. Subsequently, official statistics internationalized and statistics became a general dispositive for economics and politics. Up to this day, official statistics institutions are mainly conceived of as part of the political governance by nation-states, although their actual social basis is broader than that; after all, they are based on democratically deliberated and negotiated laws and offer statistical information for the wider public (e.g. in the form of tables, press releases, or publications for download). In line with their purpose and history, categories, topics and reports proceed slowly in official statistics and are coined in statist terms using categories such as demography, education, production, economical organizations and others. The "statist" orientation of public statistics is among others evident in the prominent role that administrative data play in this field. These are a source of relevance and mark an enormous methodological potential – but at the same time mirror the deep historical and institutional entanglements between state bureaucracies and official statistics.

In this data world, data are often considered as objective representation of facts. As mentioned before, it has mainly been the industrial convention which made official statistics possible and structured the development of its modes of coordination and its criteria for evaluation. The lasting influence of the industrial convention results in a reliance on a positivist understanding of science, which considers facts as givens and as based on unprob-

lematic and somehow "preexisting" statistical chains. The conventional work of producing data and facts is blinded from view. The specific relevance of the industrial convention in this data world is linked to the needs and technologies of statist planning and governance. At the same time, the industrial convention empowers scientists and statistical specialists to act as legitimate experts for data production at the detriment of other social actors. While these other actors have for a long time hardly been regarded as relevant stakeholders, it is noteworthy that official statistics lately have started to reach out to such other interest groups and organizations. These initiatives mirror the increasing pressure and demands for new kinds of data services and public expectations to serve a common good in other ways than those related to concerns and interests of efficient government. The dominant position of official statistics is thus increasingly questioned by new social agents such as NGOs and social movements. Official statistics are criticized for not providing relevant data for public purposes – and as a reaction they are also reconsidered from within the field of public statistics itself.

3.2. The big data world

Today, all spheres of society are equipped and entrenched with numerical data and numerical representations. With computerization, digitalization and the Internet, the datafication of societies has accelerated. The buzz word "big data" signals fundamental changes in the discourse about data, but also in discourses on science and society. The idea entailed in this concept is that data are produced by and through interlinked and dynamic digital technologies and as such are ready to be exploited for business or political decisions, in many cases in real time. The notion of big data implies not only a decentering, but also an invisibilization of processes of data production and data exploitation, for the reason alone that the majority of data producing technologies are owned by private companies. Proponents of big data at the same time question classical statistical concepts and quality criteria and claim that big data are superior in comparison to established conceptions of statistics [19]. The privatization that defines this data world points to a dramatic change: the missing link of data production to democratic legitimization and public visibility. Data here is seen first and foremost as an economic resource, which offers profit exactly because of the asymmetry implied by private data property [17]. In this context, the market convention becomes highly influential, which credits worth and quality to immediate

and temporary exchange and therefore regards value as necessarily unstable. The differences between the market convention and the industrial convention (including their methodological consequences) can be demonstrated by the examples of "deep learning" or "predictive analytics", which aim to predict individual behavior. These strategies are not valued for the truth of their modeling of causal relations, but for their predictive success alone [27]. The inspired convention provides the basis for crediting data scientists, for example for their capacity to identify behavioral patterns that could be exploited to achieve profit out of data. Data analytics as a field of "social research" of course also relies on scientific expertise and is therefore also oriented towards the logic of the industrial convention. The industrial convention, however, is only marginally relevant in this data world, and mainly because it allows to justify scientific standards and techniques which are applied in big data analytics. But because of the nontransparent character of methodological standards and criteria, big data mostly refrain from explicit legitimization based on quality conventions. Predictive success justifies big data strategies for managers and stock owners, without further questioning of the validity of algorithms.

The big data world is linked to contemporary forms of governance in many ways [11]. A natural link is provided by the trend towards neoliberal policy instruments such as benchmarking and governing every aspect of society by indicators. These processes have already led to situations in which official statistics and the big data world have become intertwined – for example in initiatives to use new data sources (such as mobile phone or social media data) as a basis for calculating indicators demanded by the political field (see for example the UN's Sustainable Development Goals, MacFeely and Nastav [18]; Fraisl et al. [12]). In these coalitions, big data are used for specific methodological reasons. For example, there have been several proposals to move from traditional population forecasting towards datafied "nowcasting", i.e. for projecting the development of statistical estimates from administrative, census, or survey sources on the basis of mobile phone or social media data. However, important tensions remain. One logical line of critique is that the data production in the big data world is not linked to any transparent or controllable form of pursuing a common good. These tensions materialize in disputes concerning the legitimacy and feasibility of procedures of datafication. The field of migration statistics provides an interesting example: After first and rather enthusiastic contributions that highlighted the potential of using new data sources

for monitoring and forecasting international mobility, recent assessments based on first hands-on experiences show a lot of skepticism regarding fundamental ethical issues, from data privacy to potential harm to migrant groups.[3]

3.3. The civic data world

Over the past few years, we have witnessed the emergence of additional data worlds that may be characterized as "civic" insofar as they are mainly embedded in civil society. Social movements, civil society initiatives and lay organizations, NGOs, media enterprises, as well as scientific networks and organizations are engaging in growing numbers and increasingly professionally in initiatives that aim at the construction of novel data infrastructures which allow new forms of data, enabling the detection and analysis of current social problems, including ecological issues or reliable and transparent data about pandemics (as is the case for the COVID-19 pandemic). Phenomena such as citizen science or open science [12] form an important part of this data world: they demonstrate the specific forms of compromises that are forged between the civic and other conventions in this data world. Of course, this emerging data world is not (yet) as strongly integrated in or shared by institutions as the world of official statistics is. What we are currently witnessing is an inherently heterogeneous and many-faced data world. The motivation of actors engaged in this world is mainly based on a critique of the inadequacy of current state information and official data, its categories, missing data access and the long-term horizon of its production. A well-known example is the questioning of the gross domestic product as an adequate indicator for economic wealth by Stiglitz, Sen and Fitoussi (see also Karabell [13]). In this context, critical social movements and also scientific networks mobilize new data sources and infrastructures to strengthen civil society knowledge about social issues, which are neglected or inadequately measured by private companies or state institutions. For example, the French movement "statactivisme" [14] is actively involved in disputing and criticizing the neoliberal (mis)use of statistics in public realms.

The civic data world is not principally skeptical about numerical data, but skeptical about the adequacy of ex-

---

[3]See e.g. the recent report on an expert workshop jointly organized by the German Federal Foreign Office and the International Organization for Migration (IOM): https://displacement.iom.int/reports/workshop-report-forecasting-human-mobility-contexts-crises [retrieved 26/10/2020].

Table 2
Comparing data worlds

	Official statistics	Big data	Civic society
Quality conventions	Industrial convention, civic convention	Market convention, convention of inspiration, industrial convention	Civic convention, industrial convention, domestic and network convention
Infrastructure	State centered, financed by tax payers, but independent (on the basis of law)	Mainly owned by private companies, nontransparent and driven by private interest	Aligned and mobilized by different kinds of actors to generate data fit for public action; conceived as "owned" by the public and accessible for engaged civilians
Engagement for common good	Aiming in a long-time perspective for provision of neutral data, categories are related to interests and tasks of governments and ministries	Aiming in a real-time perspective for profit-generating knowledge and knowledge to influence consumer behavior	Aiming for middle-term knowledge related to social issues, conflicts and problems. and related to public action and empowering civic agencies (as social movements, NGOs)
Common good	Democratizing knowledge about societal "facts", enhance effectiveness and transparency of governance; enhancing voters' political knowledge by providing objective data about society	In most cases no engagement for a common good	Providing numerical representations which empower civic agencies and countervail governmental or entrepreneurial representation of "social facts"; bringing in civic participation into political decisions based on self-generated data
Mode of governing by numbers	As provider of numerical presentation for governmental institutions, media and public actors; in many countries official statistical institutes are independent from political intervention and have an institutional autonomy	Embedded mainly in economic decision making, marketing and consumer behavior analysis; governance effects are mainly invisible; big data infrastructures and big data are also mainly owned by private companies	In alliance with mass media and social media, civic data world mobilizes political support to influence governmental agencies, companies and populations.

isting indicators and categories and dominant interpretations of data. The core argument is that criteria how to define indicators and how to organize measurement need to be coherent with normative orders and forms of public action. Data is seen as a resource for civil societies, to be controlled by scientifically skilled and engaged citizens who reflect at once normative decisions ("what and how to measure") and the form of representation of social facts implied by different kinds of data. The combination of the industrial and the civic convention is a prerequisite for this data world. It is in this data world that the definition of indicators and categories is currently reflected most intensively. One important demand is that data are expected to be related to one's personal regional and local context and to the concrete social problems actors are concerned with and about. The domestic convention and the network convention therefore are also influential in this data world; they are drawn upon to justify new forms of data production and to assess data quality in novel ways. Further, the network convention is mirrored in the project-oriented and problem-oriented realignment of data infrastructures that marks this data world [15,16]. The discontent with "slow" or nontransparent infrastructures demonstrates the need to adjust data production quickly to emerging social needs, in a way that can be evaluated by the public t as Lane [15,16] has insisted in the face of the current need for flexible data infrastructures regarding the COVID-19 pandemic.

3.4. Comparing data worlds

To characterize these different data worlds in a comparative manner, Table 2 applies a set of criteria to them. Among others, it illustrates that the claim to provide "quality data" to society is not restricted to official statistics.

The plurality of data worlds introduced here by aligning three of their most influential current variants is in fact more complex and not restricted to the ones presented here. (For example, we did not discuss the social sciences as a data world in its own right, see Vogel [26] and Diaz-Bone et al. [11]) Still, Table 2 presents the de-centered situation of data-infrastructures in contemporary (Western) societies. We argue that the state and official statistics are about to lose their formerly dominant position of data production and distribution. Also, the comparison emphasizes the contradictory nature of the three data worlds, which is the main cause for the tensions and critiques which currently arise.

4. Tensions and critiques

The Corona crisis can be mentioned as a typical example of the tensions which can arise due to the plurality of data worlds. Interestingly, the most influential provider of latest data on COVID-19 infections and mortality is not an agency anchored in the field of offi-

Table 3
Mutual critiques between different "data worlds"

	Official statistics	Big data	Civil society
Official statistics		Inefficient, state dependent, applying old-fashioned methodologies, too slow in data proceeding and publication	State centered and not providing data, relevant for contemporary social issues and civic concerns
Big data	Opaque procedures, not relying on methodological standards, not pursuing a common good		Profit-oriented and opaque, not linked to a common good, ignoring the imperative to justify measurement
Civil society	Methodological amateurs, particular and politically biased interests of social movements and specific parties	Naïve, because engaging for a common good and not for profit; limited by lack of access to data generating technologies	

cial statistics, but the Johns Hopkins University, a private US American university which gathers and publishes relevant data for countries worldwide, but also on a regional level. The data of this university were used to criticize information published by national institutions such as the Robert Koch Institute (RKI) in Germany or the Federal Health Department (BAG) in Switzerland, based on the argument that state-driven institutes were too slow or that the data produced by their infrastructure was not efficiently organized. Vice versa, institutes such as the German RKI insisted on the profound and ex-post proven quality of their data. In Switzerland, it was the national media agency (SRF) which implemented a data unit that gathered data on infections and deaths from Swiss cantons to be published as quickly as possible.

These developments mirror important tensions between data worlds that become effective in the form of mutual criticism. Table 3 represents the fundamental critiques which each of the outlined three data worlds voices (in the columns) in relation to the other two worlds. It would be misleading to portray the tensions and critiques as based on controversies about single – and for many evaluations too simple – criteria such as adequacy, timeliness, and so on, or in fact any fixed and finished "criteria catalog". Instead, we argue that different data worlds are based on quality conventions that express different institutional rationalities, which again are linked to varying deeper logics of evaluation, valuation and interpretation, including forms of methodological reasoning. Investigating these different underlying quality conventions allows the identification of the interlinked normativities and epistemic values, which again articulate different conceptions of public action and the common good.

The plurality of data conventions as well as their complex entanglement with institutional forms, political rationalities, and concrete data practices have important implications. In a nutshell, official statistics needs to actively discuss methodological questions in relation to how statistical data and knowledge production is em-

bedded in political, economic, and social formations and situations. It is not sufficient to criticize other data actors for their evidence lacking rigor or quality or for being "unsound". Instead, we need to think about how the production of data as well as of "facts" always depends on contextual conceptions of the common good.

Related to the problem of plurality, official statistics also needs to move beyond its state-centered identity which conflicts more and more with the increasingly neoliberal character of societies on the one hand and with an emergence of influential social movements and NGOs on the other hand. Against this background, the political definition of the common good as well as the concrete instruments of "governing by numbers" are themselves becoming increasingly contested. This development has two important consequences. First, official statistics are less and less employed to just represent social realities, but increasingly to consciously transform and govern the social world. Political approaches such as evidence-based policies, benchmarking, and indicator-based regulation of social practices and organizations indicate a profound shift in the role that official statistics play. In this context, facts can never be considered as neutral, pure, or innocent. Second, the function of providing the numbers by which societies are governed is in many cases actually already fulfilled by private actors. For example, the German "Bertelsmann-Stiftung" has become, as a private agent, deeply involved in the public representation and governing of social affairs. In other contexts, official statistics are actually already exploring options for coalescing with the "big data world"; for example, the obligation to deliver numbers on completely new social indicators defined in the UN's Sustainable Development Goals has led statistical agencies to promote and actively launch "data alliances" in fields such as migration (see e.g. the "Data for Migration Alliance", data4rmigration.org). The world of official statistics is thus already deeply entangled with other data worlds. As a result, it needs to reflect upon conceptions of the common good as well

as of "good data" that prevail in these other worlds. In this context, scientific standards have to be conceived as normative, and not in a positivist manner as self-evident and given, and "data" as unquestioned mirrors of social reality. In this sense, the discussion of "fake news" versus "scientific truth" is misleading. Any claim that official statistics is soundly based alone on science threatens to become a false front as long as it is not linked to opening the "normative black box" underlying any process of quantification or datafication.

Further profound questions arise from these observations. To give an example, the production of data on the social world is necessarily linked to practices of classification. Over the past centuries, the state was in control of implementing and of legitimizing social categories (on the basis of the law, but also reflecting public debates and social movements). With the rise of big data, other actors become involved in the "business" of classification. The states' power to define categories is vanishing and social categories emerge faster and in a more decentralized fashion in other contexts. This leads to the serious issues of who controls these processes of classification, who monitors their social consequences, and who organizes spaces and possibilities for discussing and criticizing them. The question is, how to include social deliberation into the foundation of national statistical institutes and official statistical categories? The task of the social sciences in this context is to increase the awareness of the convention-based nature of measurement and statistics.

In our view, the key strategical question which will decide about the future role and legitimacy of official statistics is how official statistics engages with struggles for common goods and debates about their definition. Will and can official statistics act as an ally for new political actors such as NGOs, social movements, media agencies, but also for those active in the context of citizen science or open data? Only if and in so far as this perspective is taken, one can truly speak of "public statistics".

As this contribution has made evident, the important opposition for official statistics is not "fake news" against "scientific data" [21]. The important challenge for the data world of official statistics stems from the opposition that is currently evolving between the big data world "versus" the civic data world. One key future reality test (in the sense of Boltanski [1]) will lie in how official statistics deals with the challenge of big data, which is mainly driven and controlled by big tech companies. What is an adequate institutional response to develop data infrastructures that (1) serve some clearly defined public interest (or common good) but that also (2) can prevail in the face of asymmetrically powerful privatized big data infrastructures owned by big Internet and other data companies? Can official statistics become more "fluid" to form coalitions with new public movements and structures? Which (experimental) institutional principles could provide a basis for organizing these coalitions?

References

[1] Boltanski L. On critique. A sociology of emancipation. London, Polity. 2011.

[2] Boltanski L, Chiapello E. The new spirit of capitalism. New York, Verso. 2005.

[3] Boltanski L, Thévenot L. On justification. Economies of worth. Princeton University Press, Princeton. 2006.

[4] Bruno I, Didier E, Prévieux J. Statactivisme. Comment lutter avec des nombres. La Découverte, Paris. 2014.

[5] Desrosières A. The politics of large numbers. Harvard University Press, Cambridge. 1998.

[6] Desrosières A. Pour une sociologie historique de la quantification. L'argument statistique I. Mines ParisTech, Paris. 2008.

[7] Desrosières A. How to be real and conventional. A discussion of the quality criteria of official statistics. Minerva. 2009: 307–322.

[8] Desrosières A, Thévenot L. Les catégories socioprofessionnelles. 5th ed. La Découverte, Paris. 2002.

[9] Diaz-Bone R. Convention theory, classification and quantification. Historical Social Research. 2016(2): 48–71.

[10] Diaz-Bone R. Die "Economie des conventions". Grundlagen und Entwicklungen der neuen französischen Wirtschaftssoziologie. 2nd ed. Springer VS, Wiesbaden. 2018.

[11] Diaz-Bone R, Horvath K, Cappel V. Social research in times of big data. The challenges of new data worlds and the need for a sociology of social research. Historical Social Research. 2020: 314–341.

[12] Fraisl D, et al. Mapping citizen science contributions to the UN sustainable development goals. Sustainability Science. 2020, 15(6), 1735–1751.

[13] Karabell Z. The leading indicators: A short history of the numbers that rule our world. Simon & Schuster, New York. 2014.

[14] Lamont M, Thévenot L. Rethinking comparative cultural sociology. Cambridge University Press, Cambridge. 2000.

[15] Lane J. After Covid-19, the US statistical system needs to change. Significance. 2020(4): 42–43.

[16] Lane J. Democratizing our data. MIT Press. Cambridge. 2020.

[17] MacFeely S. The big (data) bang: Opportunities and challenges for compiling SDG indicators. Global Policy. 2019, 10(S1), 121–133.

[18] MacFeely S, Nastav B. "You say you want a [data] revolution": A proposal to use unofficial statistics for the SDG global indicator framework. Statistical Journal of the IAOS. 2019, 35(3), 309–327.

[19] Mayer-Schönberger V, Cukier K. Big data. A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt. New York. 2013.

[20] Radermacher WJ. Governing by the numbers. Statistical governance: Reflections on the future of official statistics in a digital

and globalized society. Statistical Journal of the IAOS. 2019: 519–537.

[21] Radermacher WJ. Official statistics 4.0. Springer, Cham. 2020.

[22] Salais R, Baverez N, Reynaud B. L' invention du chômage. 2nd ed. Presses Universitaires de France, Paris. 1991.

[23] Savage M, Burrows R. The coming crisis of empirical sociology. Sociology. 2007: 885–899.

[24] Stiglitz J, Sen A, Fitoussi JP. Mismeasuring our lives. Why GDP doesn't add up. The New Press, New York. 2010.

[25] Storper M, Salais R. Worlds of production. The action frameworks of the economy. Harvard University Press, Cambridge. 1997.

[26] Vogel R. Survey-Welten. Eine empirische Perspektive auf Qualitätskonventionen und Praxisformen der Umfrageforschung. Springer VS, Wiesbaden. 2019.

[27] Zuboff S. The age of surveillance capitalism. The fight for a human future at the frontier of power. Public Affairs, New York. 2019.