

Different Data Futures: An Experiment in Citizen Data

Dr. Evelyn RUPPERT

Professor of Sociology, Goldsmiths, University of London, UK

Big data and citizens are inseparable: from smartphones, meters, fridges and cars to internet platforms, the data of digital technologies is the data of citizens. In addition to raising political and ethical issues of privacy, confidentiality and data protection, the repurposing of big data calls for rethinking relations to citizens in the production of official statistics if they are to be trusted. I argue for relations that involve co-producing data—or ‘citizen data’—where citizens are engaged in all stages of statistical production, from the design of a data production platform to the interpretation and analysis of data. While raising issues such as data quality, I suggest that in a time of ‘alternative facts’, what constitutes legitimate knowledge and expertise are major political sites of contention and struggle and require going beyond defending existing practices towards inventing new ones. In this light, the future of official statistics not only depends on inventing new data sources and methods but also mobilising the possibilities of digital technologies to establish new relations with citizens.

1. Introduction

Facebook data breaches and election influencing of Cambridge Analytica along with claims about alternative facts make it a challenging time to talk about a research experiment that involves designing an app for citizen data. Of course it is a challenge that extends to governments and other organisations that are experimenting with apps and platforms for producing data. But such moments also afford an opportunity or opening to imagine different data futures.

How Facebook data was allegedly used to interfere in the US election and UK referendum was joined by the disclosure that the personal information of up to 87 million users was harvested without their permission by an app designed by a Cambridge academic. A main lesson to draw is not that an academic, an internet platform, and a data company are culpable. Rather it is that data and politics are inseparable such that academics, statisticians or app developers cannot be naïve but must be

reflexive about how they may be implicated in the ways data is part of emerging forms of power relations. For data is not only shaping social relations but democratic politics.

That the proliferation of digital technologies and data have contributed to competing knowledge has fueled similar reactions about the threat of alternative facts. While some reactions are that this represents a democratization of knowledge and the erosion of the domination of experts, the separation between true and false is never straightforward. Such a dichotomy belies how all facts are produced and mediated by complex practices and technologies and are full of uncertainties [1]. The division between the real and fictitious is vexed—there are no truths and falsehoods independent of the knowledge regimes that produce them. For this reason, I doubt that the politics epitomised by Trump and his followers heralds a new era of post-truth; rather, it signals the emergence of new regimes of truth better characterized as ‘post-trust’ [2].

Yet, a prominent reaction has been the proliferation of expert practices to now authenticate facts in order to restore authority. For example, Open Europe’s Fact Check blog is where European experts distinguish ‘EU fact from EU fiction’. This has led to numerous challenges such as who will fact check the fact checkers. However, rather than restoring authority, these efforts only amplify the binary and truth making as a struggle between different gatekeepers, intermediaries and validators. It treats citizens as needing experts to validate facts for them.

I suggest that these reactions are openings for thinking about different data futures through what I call an experiment in citizen data. With a focus on data about individuals and populations, it is an experiment that reconsiders relations between states, citizens and digital technologies in the production of data for statistics by imagining a new political subjectivity, that of the data citizen. Before elaborating on these openings, in the first part of this paper I reflect on how these struggles are driven by imaginaries of big data that conceive of subjects as passive actors and individual privacy regulators. No doubt anyone who engages with digital technologies in the EU has experienced the call to be an individual privacy regulator through the implementation of the General Data Protection Regulation or GDPR. People are now given ever more fine-grained ways of regulating what, when and how data can be produced about them. While important, and satisfying to not opt for future communications, data rights are confined to consenting to the collection of data and the sending of emails. But how people might participate in the production and interpretation of data to which they agree or consent is not at a matter of concern.

This is an issue I take up in the second part of the paper where I describe an experiment that imagines subjects as data citizens with the right to shape how data is made about them and the societies of which they are a part.

This paper draws from research on a European Research Council funded project called ARITHMUS for short. The project is broadly concerned with the practical and political implications of methods for knowing the European population and, amongst other things, experiments with new digital technologies such as smart phones, tablets and web platforms to produce official statistics from big data such as that from mobile phones, search engine queries and social media. Methodologically, I have been studying these issues along with five researchers through what we describe as a multi-sited and multi-method collaborative ethnography of the data practices of EU national and international statistical institutes. My paper consists of reflections on a series of working papers and articles we have written and how that work informed an experiment in citizen data.

2. Sociotechnical Imaginaries of Big Data

What does it mean to reimagine relations between states and citizens in the production of data for official statistics? Philosophers and social and political theorists have argued that to understand what holds societies together requires understanding its institutions and the imaginaries they require to function. Benedict Anderson, for example, engaged with the force of imaginaries in his well-known definition of a nation as ‘an imagined political community’ [3]. He elaborated how shared imaginaries of technologies such as the census, the map, and the museum were organised historically and came to shape how colonial states governed their subjects and territories. Others have expanded on this to argue that in modernity, science and technology have been most significant and for this reason they refer to ‘sociotechnical imaginaries’ [4]. For our present time, some of the most forceful sociotechnical imaginaries concern those about digital technologies and big data. From the internet as both liberating and enslaving to autonomous yet murderous cars, one that persists is that of a ‘big data revolution.’ What exactly *is* big data remains a matter of some debate and it can include everything from that generated by social media, browsers and digital transactions to that from mobile phones, emails, text messages, electricity meters, sensors and travel cards. But my use of it here is not to accept the way it is being defined but to consider its imaginary force which is very much shaped by the predominant definition referred to as the ‘3Vs’: volume, velocity and variety. It is one of the most oft-cited, debated and vague and yet accepted definitions whose take up demonstrates two things.

First, what each of the ‘Vs’ means is hugely variable such that the force of big data imaginaries is not in its definition but how its functioning provokes valuations of qualities of data such as speed, real-time, quantity, granularity, flexibility, scalability, extensity and reach. Second, the focus on these qualities obscures the practices through which data is being produced and analysed by myriad organisations, corporations, institutions, and so on that have fueled imaginaries of it as a revolutionary force. That is, the force of big data imaginaries is to be found in the adoption of new ‘mindsets’ and ‘paradigms’ that take cues from how data is imagined and produced by private technology corporations and analysed by data scientists. It is to be found in how such imaginaries are simultaneously reconfiguring cultures and practices of data production within both universities and governments.

To speak of dominant imaginaries then is to underscore that they not only shape what is thinkable but also the practices through which actors perform them. So, while some commentators declare big data as ‘hype’, these pronouncements underestimate the material and political effects of imaginaries as they are taken up in practices through which new ways of thinking are propagated. One effect is what sociologist Pierre Bourdieu refers to as the power to ‘make things’ or in other words the authority over the making of statistics on economies or populations, or what comes to be legitimated as collective knowledge and truth [5]. However, it is the other side of this power – what I refer to as the power to ‘produce subjects’ - that concerns me here and which brings me back to what I highlighted in the introduction of this paper: who are the subjects of big data?

All methods are configured in ways that imagine who are their subjects and how they should, can and will likely perform. For example, the dominant method of statistical institutes—the modern paradigm of the census or survey questionnaire—typically imagines and engages with people as respondents and data subjects. While not without problems and not wanting to idealise questionnaires, they involve direct relations with subjects who are called upon to participate in their identification but who can thereby also exercise the capacity to not answer, subvert questions, challenge categories and so on. Historically, there are many examples of how people have variously influenced, interfered, or intervened in the ways questionnaires have imagined them as respondents and data subjects by, for example, challenging social categories such as on race, ethnicity, and gender.

Critical citizenship studies offers a way to interpret these as ‘acts of citizenship’ where being a citizen is understood as a political subjectivity that includes not only the possession of rights but the right to make rights claims.¹ With this conception subjects who perform in ways not anticipated by a

method and who demand identifications that are not recognised can be understood as performing as ‘data citizens’ by claiming the right to shape how data is made about them and the populations of which they are being constituted as a part [7].

Methods of data production such as questionnaires have enabled such contestations in part because of how they are technically and socially organized. From open text fields enabling the insertion of elective categories to skipping or refusing to respond to questions, the method, has variously enabled such contestations and reinterpretations. One condition of this possibility is that they involve more-or-less direct and explicit relations between statistical institutes and subjects. But it is also through such relations that citizens’ trust has also been secured about how data is generated and used for official statistics. This includes practices such as focus groups, the pilot testing of questions, and consultations with civic organisations about issues of consent, data protection, privacy, impartiality and professional standards. Through these and other means NSIs have sought to secure the trust of citizens [8]. Understood in this way, trust is not the result of one but myriad practices through which the trustworthiness of official statistics is accomplished.

How then does big data transform relations between subjects and methods of data production? Subjects are imagined as passive actors where technologies are one-way tools for extracting data about them. Through subjects’ digital interactions and transactions with platforms and devices such as social media, mobile phones and browsers, data is produced without their knowledge and through processes that work in the background. Furthermore, while that data is used for purposes such as the functioning and performance of a technology such as a platform, it can also be repurposed. This is one of the valuations promoted in big data imaginaries: the possibility of the circulation and reuse of data for purposes beyond that which they were originally generated. Data are imagined as independent of the relations of production that brought them into being and interpreted as simple reflections of who subjects are, what they think, and what they do.

The many exploitive consequences of the repurposing of big data in relation to the commercial agendas of technology corporations are now well documented. And, to return to my comments in the introduction, it is the repurposing of Facebook data by an academic to do psychological profiling and by a corporation to intervene in democratic elections that have fueled current struggles. Much critical attention is being paid to what this means for data protection, ownership, privacy and consent and effects such as profiling, the filtering of choices and influencing of opinions, and so on. However, what such criticisms underestimate are the implications of separating data from their conditions of

production. Instead, the deleterious effects of repurposing are resolved by reducing subjects to the role of individual privacy regulators in ways such as those being instituted by the GDPR.

What is ignored is that repurposing big data is implicated in the rationalities, assumptions, interests and norms of private sector professionals and technology corporations. Consequently, if repurposed for official statistics, this could undermine the trust that NSIs have relatively well achieved. As some statisticians have noted, '[o]f critical importance is the implication of any use of Big Data for the public perception of a NSI as this has a direct impact on trust in official statistics' [8]. Additionally, if authority for the production of what is deemed as 'official' knowledge of societies is indeed a stake in struggles over alternative facts, then it means delegating some of that authority to the private sector. It also means relegating to others relations to subjects as users, customers and data sources and which make the capacity of subjects to perform as data citizens in the ways I have expressed more difficult if not impossible.

The imaginary of citizen data that I outline in this paper is one possible response. It is founded on the principle of citizens as co-producers of data for official statistics rather than as ever more distant subjects whose impressions and confidence need to be managed. Co-production, as defined here, not only recognizes that the data of digital technologies is the data of citizens. It also means engaging citizens in how data about their digital actions, interactions, transactions and experiences are categorised, included and excluded and interpreted for policy and research. It is also based on the premise that co-production could lead to greater trust in the resulting statistics. Of course, such issues are not entirely new or limited to big data. Former Eurostat Director General Walter Radermacher has expressed this more generally as a gap between citizen experiences and official statistics which in turn calls for 'subjective statistics'.² In saying so he stressed the need for a more democratic debate between citizens and data producers and owners to achieve a 'more subjective, differentiated understanding of our world', instead of 'technocrats and politicians sitting together and confronting citizens in the end.' Digital technologies now afford the possibility of addressing such a gap, which the repurposing of big data potentially widens by detaching relations between states and citizens. This is evident in some NSI experiments that are seeking to better utilize digital technologies to enable more interactive and responsive relations to respondents such as in the design of digital censuses and surveys.³ The imaginary of citizen data extends these experiments to the yet-to-be realized interactive and inventive possibilities of digital technologies for engaging subjects as citizens rather than simply respondents. In the next section I describe how this imaginary has been

taken up in practice in the form of a collaborative workshop that experimented in the design of prototypes for a ‘citizen data app’.

3. A Citizen Data App

Four principles informed the development of a citizen data app, which were derived from key matters of concern statisticians have expressed about the future of official statistics: experimentalism, citizen science, smart statistics and privacy-by-design [9]. I will only elaborate on the second principle here. There are many examples of citizen science initiatives that involve people producing their own data on issues such as pollution, crime and urban change. Others involve citizens in various roles as co-producers with government, scientific or other organisations. In relation to statistics, these include an increasing number of government initiatives related to the implementation of the Sustainable Development Goals (SDGs) including the United Nations’ new Citizen Science Global Partnership that seeks to promote and advance citizen science for a sustainable world,⁴ Eurostat’s development of indicators to monitor SDGs based on data generated by a consortium of individuals and organisations⁵, and Statistics Canada’s crowdsourcing of citizen work to help fill in data gaps on geolocations⁶.

The principle of citizen science builds on these co-production models by extending them to consider how citizens might perform not simply as data collectors but as active contributors at all stages of data production. It is in relation to this principle that we have been conducting collaborative workshops to imagine different data futures through creative experimentation. Rather than summarise this work I will offer one key insight from the first workshop that involved ARITHMUS researchers, statisticians, other academic researchers, information and interaction designers, and facilitators. Four groups developed principles, designed prototypes and proposed road maps for developing four different apps. One group, for example, designed an app called ‘How we move’ to explore the different meanings of and relations citizens have to mobility that defy usual statistical categories of where people live and work. One proposition put forward was that existing statistical categories about what is called a subject’s usual place of residence do not capture the complexity of mobilities in contemporary societies. Amongst other issues, the group considered how these categories could be rethought through an app that mixed automatically collected GPS data along with citizens annotations, interpretations and categorisations of their and others’ mobility patterns. The

premise was that GPS data alone is not sufficient to understand the motivations, rationales, meanings and lived experiences of mobility.

A tension though existed between the introduction of design elements that created possibilities for citizens to engage with data in these interpretive and meaning-making ways versus those that aimed to control data collection, standardisation and quality. Not a surprising dynamic perhaps but rather than resolving the tension one solution offered was that co-produced data could be treated as a hybrid form based on different quality standards yet generative of unique and perhaps previously unimagined kinds of statistics. In subsequent conversations, statisticians spoke of co-produced data as complementary rather than a replacement of existing data, a term they often call forth when a new and unsanctioned form of data is innovated. That is, relegating co-produced data to a special status was a strategy of accepting it while at the same time retaining the authority of existing methods of data production.

However, it also signified another potential. It signified that there is not one set of standards through which data can be produced and made ‘official’. Indeed, if there is any one conclusion from our years of fieldwork it is that in practice such variability is a condition of all methods from how surveys are conducted to how administrative registers are organised. Adherence to standards, norms, conventions, rules and principles varies to the extent that what can become ‘official’ is not settled or measurable by any single standard, but as something that is *collectively* negotiated, instituted and maintained.

To note this is to underscore that the practices of different collectives may involve forms of reasoning that adhere to different principles and standards. Researchers have documented, for example, that standards such as measurement accuracy are not the only criteria for evaluating environmental data gathered through, for example, citizen sensing practices [10]. A ‘rough’ measurement to identify a pollution event when it is happening or when it has happened might be sufficient and ‘good-enough’. In other words, methods can be evaluated according to different norms, objectives and standards, and in the case of our experiment, such as the relations of production between citizens and states that bring data into being rather than their truth claims. To imagine complementary data then is to offer a different way of accomplishing what counts as ‘official’. However, and critically, this interpretation does not mean according complementary data the status of ‘alternative facts’. In the introduction, I argued against ‘fact checking’ as an answer to the evaluation of competing ‘facts’ in part because it disregards critiques of the epistemic authority and command of experts. What is at stake is not which experts win the authority to legitimise public

facts—or what counts as official data and statistics. Rather, it is the norms and values on which public facts are produced including the relations through which they are instituted and negotiated.

4. Final reflections

While examples such as the Facebook data breaches dominate the headlines, there are numerous initiatives that imagine different data futures such as DataCommons, an association of citizens that seeks the right to self-determination over their own data. Our experiment in citizen data is to imagine yet another that involves new relations between states, citizens, and digital technologies in the production of data for official statistics. We are now taking forward the prototype for a ‘How we move’ app in further workshops that will involve citizen groups not for the purposes of producing data on mobility but to probe the category of usual residence. In this way we will experiment with how design experiments can also contribute to reimagining statistical categories.

Proposals that NSIs need to defend the quality and legitimacy of official statistics through gatekeeping practices such as demonstrating their trustworthiness by making their statistical practices transparent and thus assessable, fact checking competing statistics, and ‘calling out bad numbers’ certainly have a role to play. However, they potentially play into the premise that what is at stake is winning a competition of ‘facts’. They ignore that what constitutes ‘public facts’ should be open to democratic contestation and deliberation because they inevitably involve normative judgements about social meaning and choices about which experiential realities matter [1]. In this view, NSIs have a role to play in fostering official statistics as democratic accomplishments where their legitimacy is derived from conditions of co-production that address data subjects as citizens.

The concept of citizen data raises many practical and political questions. For one, even if adopted, existing methods and their relations to citizens would not of course be superseded. However, methods such as surveys and questionnaires will likely change as digital technologies are increasingly adopted and a concept of citizen data can possibly inform those changes. That is, beyond big data sources, how data is produced by NSIs using various methods can be reconceived along the lines of citizen data. While online or digital surveys and censuses, for example, are being adopted they do not imagine the possibilities of co-production nor utilise the affordances of digital technologies through which data that more closely align with the experiences, insights and knowledge of citizens may be produced.

In sum, the authority and expertise to make statistics official are not founded in a single institution, but in processes of co-production and direct relations to citizens. In that regard, citizen data

approaches claims of ‘alternative facts’ as not matters of accuracy and standards but of the relations to citizens through which data and in turn statistics are made official. It entails a move from ‘data driven’ to ‘democratically’ driven data for statistics.

5. Acknowledgments

The project, Peopling Europe: How data make a people (ARITHMUS) is funded by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement no. 615588. Postdoctoral researchers are Baki Cakici, Francisca Grommé, Stephan Scheel, and Funda Ustek-Spilda and Doctoral researcher Ville Takala. We are grateful for the support and contributions of statisticians at seven field sites who made our research possible: UK Office for National Statistics, Statistics Netherlands, Statistics Estonia, Turkish Statistical Institute, Statistics Finland, Eurostat and UNECE.

6. Notes

¹ This understanding from the field of critical citizenship studies is summarised in [6]. Being a citizen is understood as a political subjectivity that includes not only the possession of rights but the right to make rights claims.

² Fieldwork notes, Eurostat Agility Conference, November 2016.

³ For example, Statistics Netherlands has been exploring data collection designs that introduce respondent interaction with sensor data as way to reconceive questionnaire design: Mussmann, B. O., J. Bakker, B. Schouten, and R. Warmerdam (2017) Dissolving Questionnaire Borders with Technology: The Paradigm Shift in Data Collection. Paper presented to the UNECE workshop on Statistical Data Collection, 10-12 October 2017, Statistics Canada.

⁴ See <https://bit.ly/2NB44Vc> for an overview of the launch of the ‘Global Partnership’ in December 2017.

⁵ SDG 15 concerns ‘Life and Land’ and the indicator is the ‘Common bird index by type of species’. For this indicator the collection of raw data and compilation of population indices are coordinated through the Pan-European Common Bird Monitoring Scheme (PECBMS), which has been developed through a consortium of individuals and organisations from many countries, cooperating through the European Bird Census Council (EBCC). The PECBM scheme permits a sharing of knowledge and know-how and encourages and supports the setting-up of new schemes. See: https://ec.europa.eu/eurostat/cache/metadata/EN/sdg_15_60_esmsip2.htm.

⁶ The pilot was organized by Statistics Canada in collaboration with OpenNorth, MapBox, City of Ottawa, OSM Canada. OpenNorth is a non-profit organization developing digital tools for civic engagement. Avail at: <http://www.statcan.gc.ca/eng/crowdsourcing>.

7. References

- [1] Jasanoff S and Simmet HR. (2017) No funeral bells: Public reason in a ‘post-truth’ age. *Social Studies of Science* 47: 751-770.
- [2] Lynch M. (2017) *Post-truth, alt-facts, and asymmetric controversies*. Available at: <https://bit.ly/2IPjgcp>.
- [3] Anderson B. (1991) *Imagined Communities: Reflections on the Origin and Spread of Nationalism*, London: Verso.
- [4] Jasanoff S. (2015) Future Imperfect: Science, Technology and the Imaginations of Modernity. In: Jasanoff S and Kim S-H (eds) *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, Chicago: University of Chicago Press, 1-33.

- [5] Bourdieu P. (1989) Social Space and Symbolic Power. *Sociological Theory* 7: 14-25.
- [6] Isin, E., & Ruppert, E. (2015). *Being Digital Citizens*. London: Rowman & Littlefield International.
- [7] Ruppert, E. (2018). Sociotechnical Imaginaries of Different Data Futures: An Experiment in Citizen Data. Erasmus School of Behavioural and Social Sciences, University Rotterdam.
- [8] Struijs, P., Braaksma, B. and Daas, P.J.H. (2014) Official statistics and Big Data. *Big Data & Society*, Apr-June, 1–6.
- [9] Ruppert, E., F. Grommé, F. Ustek-Spilda, and B. Cakici. Forthcoming. ‘Citizen Data and Trust in Official Statistics’. *Economie et Statistique*.
- [10] Gabrys J and Pritchard H. (2015) Just Good Enough Data and Environmental Sensing: Moving Beyond Regulatory Benchmarks toward Citizen Action. In *Workshop Proceedings, Infrastructures and Platforms for Environmental Crowd Sensing and Big Data*. Barcelona: European Citizen Science Association, 28-30 October.