ARITHMUS

Citizen Data and Official Statistics: Background Document to a Collaborative Workshop

Working Paper Number 2

Francisca Grommé, Funda Ustek-Spilda, Evelyn Ruppert, Baki Cakici

Goldsmiths, University of London

6 July 2017

ARITHMUS
www.ARITHMUS.eu

Centre for Invention and Social Process (CISP)

Department of Sociology
Goldsmiths | University of London
New Cross | London SE14 6NW
csisp@gold.ac.uk
www.gold.ac.uk/csisp/

The ARITHMUS project is funded by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 615588.

Citizen Data and Official Statistics: Background Document to a Collaborative Workshop

Working Paper Number: 2

Authors: Francisca Grommé[1], Funda Ustek-Spilda, Evelyn Ruppert, Baki Cakici

Author Affiliations: Goldsmiths, University of London

Page Count: 30

How to cite: Grommé, F., F. Ustek-Spilda, E. Ruppert, B. Cakici. 2017. "Citizen Data and Official Statistics: Background Document to a Collaborative Workshop." ARITHMUS Working Paper Series, Paper No. 2. DOI: 10.13140/RG.2.2.18532.88966

Abstract: This working paper was written in preparation for a collaborative workshop organised for statisticians, social scientists, information and app designers and other participants inside and outside academia. The autumn 2017 workshop aimed to develop the main principles for a citizen data app for official statistics. Through this work we sought to conceive of a new regime of data collection in official statistics through different devices. How can we capture citizens' meanings and intentions when they produce data? Can we develop 'smart' methods that do not rely on cooperating with, and data generated by, large tech companies, but by developing methods and data co-produced with citizens? Towards addressing these issues we developed four key concepts outlined in this document: experimentalism, citizen data, smart statistics and privacy by design. We introduced these concepts to facilitate shared *understandings* of their meaning, provide a background to discussions about them and the questions they raise for official statistics. Through then jointly working on the *practical* development of a citizen data app, the objective was to reflect on and reshape these concepts and to identify other concepts that might aid our understanding of the possibilities of citizen co-production of data for official statistics.

Keywords: smart statistics, practice-based research, Big Data, citizen science, experimentalism, privacy by design

[1] Corresponding author: Francisca Grommé, Goldsmiths, University of London (f.gromme@gold.ac.uk).

Preface

This working paper was written in preparation for a collaborative workshop organised for statisticians, social scientists, information and app designers and other participants inside and outside academia. The aim of the workshop is to collectively identify the main principles for the design of a citizen data app that would involve citizens in the co-production of data for official statistics.[2]

Collaborative design is a new phase in the ARITHMUS research project[3], for which we have been conducting an ethnography of the making of official statistics through interviews and various degrees of participant observation. In our research, we have been concerned with the practical and political problem of assembling multiple national populations into a European population and people through statistical practices. Broadly, our question is: how do data make a people? To answer this question, we have been following the work practices in five national statistical institutes (NSIs) in Europe[4] and two international organisations (Eurostat, the Statistical Office of the European Union and the Statistical Division of the United Nations Economic Commission for Europe (UNECE)). The ARITHMUS team has examined, with the insights of national statisticians, how new digital technologies are stimulating methodological diversification in how censuses are currently taken and how Big Data sources are challenging established official statistics methods, specifically those of census-taking.

One of the problems encountered and articulated together with statisticians during our fieldwork concerns a drawback of adopting Big Data methods in official statistics. This drawback mainly stems from the distance between citizens' actions, answers and identifications that produce data and the interpretation of that data for and by statistical methods. The workshop sought to address these issues through a collaborative process that involves identification of the main principles for the design of a citizen data app, which could be co-produced between NSIs and citizens. For us, collaboration is a method for exploring the possibilities of citizen co-production of data for official statistics.

The text that follows was written to support an initial meeting with the ARITHMUS project's Advisory Group of national statisticians to discuss the possibility and organisation of a collaborative workshop.[5] It introduces key concepts to facilitate shared *understandings* of their meaning, provide a background to discussions about them and the questions they raise for official statistics. We discussed these concepts with the Advisory Group and their applicability, usefulness, and urgency for official statistics. While the concepts may change, the current version outlined here was largely supported by the ARITHMUS Advisory Group.

[2] Whereas the previous ARITHMUS working paper discussed collaboration in a research team (see Scheel et al, 2016), in this working paper we discuss collaboration between academia and professionals.

[3] The ERC funded project, Peopling Europe: How data make a people (ARITHMUS) includes six team members: Evelyn Ruppert (Principal Investigator), Baki Cakici, Francisca Grommé, Stephan Scheel, Ville Takala and Funda Ustek-Spilda.

[4] These are Statistics Netherlands, Office for National Statistics (England and Wales), Statistics Finland, Statistics Estonia and the Turkish Statistical Institute.

[5] The Advisory Group consists of designated representatives from the international and national statistical institutes that the project is studying.

We have also produced this paper as a resource for other research projects concerned with similar questions and issues related to the production of statistics and data-related practices more generally. In addition, it may be read as an example of an experiment with collaborative methods through which we not only study our 'research subjects' but also conduct research through the practice of doing or making an app with them. Although at the time of writing we were only at the beginning of a longer and tentative process (a process that we recognise may not succeed in ways we intended), we aim to also adopt and develop elements of practice-based research as part of our ethnographic methods. Here, we understand practice-based research as a method to increase our understanding of the (often) hidden skills, collectives and infrastructures that are part of designing a new way of statistical data collection through living the messiness and failure that are part of designing an app (Jungnickel, 2017).

In what follows, we present the original document discussed at the Advisory Group meeting of February 2017 with only minor changes. As such, in terms of tone and style, it addresses a readership of professional statisticians. First, we give an overview of the main ideas and concepts informing the workshop: experimentalism, citizen science, smart statistics and privacy by design. Next, we briefly present example projects. Each section ends with a series of questions we formulated to open up discussion during the Advisory Group meeting.

Introduction

As part of the ARITHMUS research project, we propose a collaborative workshop with members of the Advisory Group and invited participants. This was originally part of the research design of the ARITHMUS Project, but the specific idea for a collaborative workshop emerged in previous Advisory Group meetings.

The aim of the workshop is to design a 'citizen data app' or a platform that allows different apps to communicate and share data for statistics for government and research. The interest in developing a citizen data app emerged from the context of producing and using statistics based on Big Data sources. Throughout our fieldwork, statisticians whose work we have followed have noted several drawbacks they identified with Big Data from the perspective of official statistics, such as issues of access, data ownership, privacy and ethics, representativeness of data, data quality and so on. In addition to these, throughout our fieldwork, we have also become cognizant of the distance between the citizens' actions, identifications and experiences that create the data, how they are categorised, included and secluded, and the interpretation of that data for statistics. This workshop provides an opportunity to think about how NSIs might generate statistics from new data sources (including Big Data) by co-producing data with citizens, instead of predominantly relying on data generated by companies in the private sector with all the aforementioned drawbacks. It is important to note here that using Big Data for official statistics is still at an experimental stage for most NSIs and some NSIs reject using it as a source altogether due to these and other drawbacks. However, the concerns that we seek to address extend beyond Big Data and include methods and data of statistical production such as questionnaires and registers. Against this background, the aim of the collaborative workshop is to identify the possible design elements of a 'citizen data app' or a platform that allows different apps to communicate and share data for statistics for government and research.

In this background we discuss key concepts that may serve as initial principles for an app/platform. Mainly, it should incorporate privacy by design; be based on a form of 'citizen science' and should be open source. The resulting design (and if successful, app development) would ideally be applicable to a wide range of NSIs in Europe and beyond.

The workshop will focus on the initial *concept work*, such as the problems to be addressed and the design components of a solution for a citizen data app/platform. This means that at this stage, the app will not be materialized/developed and most of its technical specifications will remain undefined. Instead, the outcome will be a description of design principles about how the app could be configured in terms of its aims, issue areas/topics and the type/topic of statistics produced. After the workshop, the ideas will then be used in a Proof of Concept application to the European Research Council (ERC). This proposal could include focus groups and workshops with citizen organisations amongst others to experiment, test, and trial alternatives.

The workshop offers a platform for collaborative work between national statisticians and the ARITHMUS researchers to explore and conceptualise common interests. Whereas we have done this throughout the ARITHMUS project through conversations, meetings and

conferences, in this workshop we aim to do this through the practice of developing a method and technology together.

To facilitate discussions on the development of design principles for a citizen data app, we introduce the four key concepts shared both by the ARITHMUS researchers and national statisticians. These are experimentalism, citizen science, smart statistics and privacy by design. We see the term 'citizen data' as an outcome of this conceptual framing; it makes explicit the evolving relationships between data users, NSIs and citizens.[6] What we understand by citizen data and why we would choose the citizen data approach will be the topic of the Advisory Group Meeting and the workshop.

This document serves as a background to and inspiration for further discussion during the 2017 ARITHMUS Advisory Group Meeting. It is structured as follows: First, it gives an overview of the main ideas and concepts informing the meeting: experimentalism, citizen science, smart statistics and privacy by design. Next, it briefly presents three example projects. We do not propose to immediately adopt the concepts and ideas reviewed here but to use them to facilitate a shared understanding of the background discussions and questions they raise for official statistics.

Experimentalism

We propose to conduct the workshop in a mode of research we refer to as experimentalism. Experimenting as proposed here does not involve controlled and scripted procedures in the closed environment of a laboratory, as might be found in empiricist and positivist fields such as psychology or economics. With experimentalism we refer to exploratory, trail-and-error work formats that do not rely on a distinction between research subject and objects (cf. CRESC Encounters Collaborative, 2013). Experimental work forms explore new topic areas and procedures, accept uncertainty about the outcomes[7] and stay clear from a language of absolute success or failure. Our general intention is to remain open to surprises, whether positive or negative.[8] We propose to adopt and reflect on experimentalism because it might be a suitable approach for collaboratively developing new methods for producing new statistics using citizen data.

In proposing a collaborative, experimental approach we join a trend in social science disciplines, such as Science and Technology Studies (STS) and anthropology, to adopt experimentation as a method (Lezaun, Marres, & Tironi, forthcoming). This is not entirely new. Various strands of social science have operated with degrees of experimentalism, one example is early 20th century Chicago School sociology in which society was researched and analysed as self-experimental (Gross & Krohn, 2005; cf. Guggenheim, 2012). Relatively new, however, is the adoption of experimenting as a method to open up areas of scientific and

[6] There is also a potential to explore the concept of subjectification as the relational process through which people engage in how they are identified in statistics with different degrees of self-identification and influence.
[7] See Rheinberger (1995) about the tensions between the requirement to articulate experimental outcomes and their unknowability.
[8] It has been argued that surprise is an achievement of experimental techniques, rather than an 'answer' of reality to experimental tests (Latour, 1993). Yet, this does not mean that all outcomes are foreseen or accepted by experimenters. See Gross (2010) on the relevance of surprise in experimental, collaborative learning.

technological expertise (Waterton & Tsouvalis, 2015). In areas as diverse as wheelchair design, Big Data and synthetic biology,[9] social scientists have adopted experimentalism in order to open up spaces for problem formulations, invite new actors and consider alternative futures. Academia has not been unique in these efforts, as government agencies and corporations have adopted similar methods. Official statistics is a good example of a government domain openly embracing experimentalism, as attested by innovation laboratories, sandboxes, hackathons and exploratory research projects.

The adoption of experimentation as a method to open areas of scientific and technological expertise is most commonly undertaken to achieve social benefits of various sorts. Broadly speaking, we can distinguish two formats by which collaborative experiments seek to achieve this. The first is through participation, in order to achieve a degree of democratisation by opening up scientific and technical debates and processes to publics (or rather: transforming technical issues into public issues, and generating publics (Marres, 2012). The second, which is the format we propose here, is to experiment collaboratively with professionals in order to develop and explore new problem formulations, transcend ingrained styles of reasoning, disrupt existing hierarchies and critically examine how objects of study come into being and what they exclude (Rabinow & Bennett, 2012; Ruppert et al., 2015). This is the model of a 'collaboratory' (or, co-laboratory) in which participants engage in the common exploration of a topic. In the fore-runner of the ARITHMUS project a collaboratory was organised to 'socialise' Big Data through discussion and provocations (Ruppert et al., 2015).[10] An example of one of the issues that experimenting with citizen science can provoke is a shift from 'users' (ministries and government agencies) to citizens as key stakeholders for NSIs.

The workshop adds a new element to the collaboratory. Adopting ideas emerging from conversations between the disciplines of design and STS, and related fields, it proposes to collaborate on a joint end product. The practical focus, therefore, is not an object of knowledge, but a 'thing': a tangible end product that requires bringing together technical, social, and organisational components. The aim is not to produce a finished and stable end product. Rather, working on a common product (a proposal for an app/platform, in our case) makes "issues experimentally available to such an extent that 'the possible' becomes tangible, formable, and within reach" (Binder, Brandt, Ehn, & Halse, 2015, p. 12). Through 'thinging' mutual problems and issues emerge. The focus is on a common enquiry, or in the words of Tomás Sanchez Criado and Adolfo Estalella, a "joint epistemic exploration" (2016).[11] Working on a common object thus forces participants to make future modes of working explicit (Muniesa & Linhardt, 2011),[12] in this case to consider the future of data collection and analysis in official statistics through a variety of devices.

Experimenting in collaboration also demands a critical and caring attitude. With regard to experiments in other fields, critics have argued that experimentation has become an

[9] For these three examples, see: https://entornoalasilla.wordpress.com/english/; Ruppert et al. (2015); and http://www.anthropos-lab.net/about.

[10] Another inspiration for this project is the model of the ethnographic 'para-site', see http://www.socsci.uci.edu/~ethnog/theme3.htm.

[11] Also see Callon et al.'s notion of 'co-researchers' (2011).

[12] A related research approach is practice-based research, in which participants from different disciplines develop a common project together. This would not necessarily result in a material object, but could also be an event or participation in a common activity.

institutionalised mode to avoid accountability and regulation (Levidow & Carr, 2007). Furthermore, experiments and experimenters are places where power is exerted and where citizens can, unwillingly, be made into experimental subjects (Rottenburg 2009). In many cases, the terms of assessment and evaluation are not made transparent and accountable. Generally, from the social studies of science we learn that experiments are not impartial and objective processes of discovery. Instead, they require a reshaping of relations between participants, objects of knowledge and things. They bring into being new entities, agencies and problematisations (Haraway, 1988; Latour, 1993) that may pose unexpected risks and problems to their environment, as has been shown for experimenting in the domain of finance and biotechnology (Millo & Lezaun, 2006).

What we suggest, therefore, is a 'care-full' approach to collaborative experimentation (Grommé 2015). Collective experiments do not so much require strict protocols, yet care and caution should be part of risk-taking (Latour, 2006). Elements of such an approach are to monitor and document who and what are (unavoidably) in – and excluded; avoid ambiguity about our terms of evaluation (when do we think something is 'good enough'?); avoid attributing failure solely to perceived local circumstances; avoid separating normative elements from scientific fact (Latour, 2006); and to ensure documentation. The elements of a care-full experimentalism could be one of the outcomes of this project. For example, a particular issues identified in advance by the ARITHMUS Advisory Group are citizen privacy, anonymity and consent, which we will treat this as a separate section in this document.

Based on the foregoing, we highlight the following questions:

1. What are the possible advantages and disadvantages of experimentalism as a learning-by-doing method for developing applications and data co-produced with citizens?
2. What are the opportunities and challenges of embedding experimentalism in the organisational structures of NSIs? (For instance, in organisational cultures, in organisational timelines or in management hierarchies?)
3. What rules, norms and values should be part of an experimental attitude in NSIs?
4. What examples of experimentalism as a working mode do you know in your organisation? What were the benefits and problems?

From 'Citizen Science' to 'Citizen Data'

A second concept informing the workshop is that of citizen science. Like experimentalism, it is a concept that has arisen as a matter of interest within NSIs. Citizen science has attracted more attention in relation to the uptake of Big Data in official statistics, which is generating new relations between citizens, statisticians and users. From utilising mobile phone geo-location data generated by telecommunication companies to measure cross-border migration or commuting trajectories, to generating consumer price indices based on publicly available data on supermarket prices, recent technological developments have opened new ways in which data is produced, processed, disseminated and imagined. Such technologies also make it possible to think of citizens as not mere research subjects, but as actively involved in the production of data as opposed to traditional methods where they have been otherwise 'passive' subjects. In this workshop we would like to discuss this difference between Big Data

and traditional methods of data production and the ways that citizens and users of statistics might be involved in the processes of statistical data production.

The involvement of citizens in the production of data and knowledge is a key concern of those practicing, engaging with and writing about 'citizen science' (CS). We therefore present key definitions and debates in the academic literature about CS below. Our suggestion is that a particular form of CS is required in the production of official statistics and propose to use the term 'citizen data' for the aims of this workshop.

Definitions of Citizen Science

There are many definitions and interpretations of citizen science in literature, with a range of responsibilities and roles attributed to citizens in scientific data production. Goodchild defines *citizen science* (CS) to describe communities or networks of citizens who act as observers in some domain of science (2007, 218). According to Kullenberg and Kasperowski, this is the most commonly accepted definition of CS, as illustrated by the significant momentum the concept has gained in the natural sciences in recent years (2016, 2). Kullenberg and Kasperowski, however, acknowledge that the practice of engaging people in collecting and submitting data for scientific purposes is not in itself a new venture and goes back at least to the 1960s, though the term itself was not used until the 1990s. The authors highlight that the emergence of CS is sometimes even extended to include the National Audubon Society's Annual Christmas Bird Count in early 1900s, where citizens were asked to participate in the observation and enumeration of bird species in their areas during the Holidays. This was to ensure the health of the bird populations in geographically diverse areas, while at the same time providing an enjoyable Holiday activity for citizens (see Figure 1 below).

As Kullenberg and Kasperowski (2016) note observation, however, is only one of the ways in which citizens can be included in scientific data production. As much as citizens might be mobilised to fulfil scientific tasks, CS can also be imagined in a way where citizens co-design scientific studies to reflect their own concerns, needs and questions. They cite Alan Irwin's book *Citizen Science: A Study of People, Expertise and Sustainable Development* where CS is defined as "(…) a science which assists the needs and concerns of citizens (…) a form of science developed and enacted by citizens themselves" (1995, 13). This definition expands the approach taken up by natural scientists and transforms the role of citizens from mere observers or data subjects to stakeholders in scientific data production.

Figure 1: National Audubon Society's Annual Christmas Bird Count in early 1900s

Source: Video still taken from "Christmas Bird Count" by Chan Robbins. http://www.audubon.org/news/11-reasons-christmas-bird-count-rocks

These two major strands of understanding CS in academic literature do not exhaust all forms taken up and used by researchers interested in co-producing science with non-scientists. Kullenberg and Kasperowski list a plethora of concepts that have been coined to describe primarily local and activist-oriented forms of CS, such as: 'community based auditing', 'civic science', community environmental policing', 'street science', 'popular epidemiology', 'crowd science', 'Do It Yourself Science' (2016, 2) and other research projects also coined their own interpretations of this method such as 'Doing it together science'[13]. These notions and concepts range from citizens seeking close alliances with scientific and knowledge institutions to citizens seeking to produce independent knowledge together with scientists.

One of the important difficulties with CS studies is that they could be more difficult to trace via scientometric methods as their results are not published in peer-reviewed literature (Kullenberg and Kasperowski, 2016) or not published at all. Nevertheless, social scientists conducting research on CS or using CS methods help these local and activist oriented forms of CS to gain visibility. Moreover, these studies push the border between science, activism and political, legal, environmental and other humanitarian causes. For instance, whereas some CS studies engage in taking legal action against the polluting industries, others use geo-information systems to promote the rights of indigenous people living in particular localities. Yet other studies utilise 'civic technoscience' with the intention to develop affordable instruments for monitoring oil spills and green urban areas (Ibid).

[13] togetherscience.eu.

CS has also been utilised to test new ideas, new technologies and new approaches to 'disrupt' current thinking and established methodologies. For example, in the context of smart cities, Henriquez (2016) argues that computer technologies have the potential to foster a type of collaborative, interdisciplinary citizen science that would potentially facilitate mutual learning between experts and 'lay people' (author's own terminology). Correspondingly, Gabrys et al. (2016) describe their collaborative development of an environmental monitoring kit for air quality monitoring in shale gas extraction sites in the US. The researchers note that although citizen-collected air quality data might be qualitatively different than the data produced by scientific communities, they have the potential to transform the way we look at air quality, as it involves not pure numerical results confirming or not confirming to set benchmarks, but what poor or good air quality means for those breathing it on a daily basis. In addition, Goodchild's examples of *mash-up*, the ability to superimpose geographic information from sources distributed over the Web, many of them created by amateurs, also reveal how CS might be an interesting way to produce science for not only the scientists but also users (2007, 211–213) and how this new way of looking at data production might also shift the way we interpret data.

Why do People Engage in CS?

"Science is not just for scientists these days" reads the first line of the article on the "Rise of the Citizen Scientists" in the journal *Nature* (2015). This quotation nicely sums up the many interests non-scientists might have in CS. From documenting concerns over environmental issues to creating an online archival map of their local historical sites, there may be many reasons for why people might want to engage in scientific data production. Goodchild notes that there are at least two reasons why people might be motivated to participate in CS: self-promotion and personal interest (2007, 219–220). He argues that *self-promotion* is clearly an important motivator for internet activity, as increasingly citizens spend their lives online and engage in online social networks. *Personal interest* can also be a good motivator, if citizens believe that the information might be useful for them personally or their communities at some point in future. For instance, Web 2.0 sites have been a convenient way of making information available to users' friends and relations, even though by engaging in Web 2.0, users make the same information available to all (219).

In addition, Jasanoff notes that CS facilitates meaningful interaction among policy-makers, scientific experts, corporate producers and the public (2003, 235–236). She argues that the pressure for accountability in expert decision-making manifests itself in the demand for greater transparency and participation by stakeholders. However, participatory opportunities cannot by themselves ensure the representative and democratic governance of science. She underscores that the attention of modern states has focused on refining the 'technologies of hubris' which are designed, on the whole, to facilitate management and control, even in areas of high uncertainty (238). What is lacking is not just knowledge gaps, but ways to bring uncertain, unknown processes and methods into the dynamics of democratic debate (239-240). Hence, there is a need for an interaction between different stakeholders in the production of science, which might be possible through CS. Subsequently, she suggests 'technologies of humility', which are *social* technologies that would give combined attention to substance and process, and stress deliberation as well as analysis (243) by all stakeholders.

So, beyond personal and individual reasons for engaging in CS; users, institutions, data producers and scientists might seek CS for a more inclusive production of science.

Can Scientific Principles Be Adhered to in CS?

The role of non-scientists in the production of science seems to be the most contentious issue in CS. Scientific research builds on standards and scientists are expected to demonstrate that their methods and results adhere to the established principles in their field.[14] However, as Goodchild demonstrates, although CS might not necessarily fulfil scientific criteria, it would open up new ways of thinking and approaching to data. He gives the example of traditional mapping agencies which have elaborate standards and specifications to govern the production of geographic information and employ cartographers with documented qualifications (2007, 219). The quality of cartographic information produced by these cartographers stems from their scientific expertise. Google maps, on the other hand, has no such reputation in the domain; its database having been mostly fed by users. Yet, Google Maps is used every day by millions around the world with little or no scepticism about its performance. While Google's authority in mapping might be arising from its success in other areas of computation (and that the mapping platform itself is based on principles of computer science and a previous cartographic software called Earthviewer), the wide acceptance of Google's mapping reference systems demonstrates that quality in data can be ascertained in many ways, especially in light of new technologies (Ibid, 219-220). Goodchild calls this "the democratisation of GIS" because by integrating straightforward capabilities of GIS to the general public, the platform enables citizens to access data, use it but also influence it, through marking newly developed areas or identifying wrong addresses. Providing an easy to use platform that does not require coding knowledge or a scientific understanding of geo-mapping further facilitates users' engagement with the platform.

Similarly, in 'geodesy', a field of mathematics concerned with mapping the shape, size and gravity field of the Earth, several studies have shown how CS can assist scientists. For instance, when digging foundations for building and trenches for pipelines, construction workers often work together with civil engineers to track soil quality, understand the properties and types of soil they will be dealing with as well as engage in soil mapping across their sites (Nature, 2015).

Some studies, however, do find that volunteers are less able to identify, record and document scientific information than scientists or they are less efficient (Ibid). Even for the Christmas Bird Count, studies indicated that participants still required a fairly high level of skill for identifying distinct bird species and individual birds they have observed. In fact, over the years of the bird count, a number of protocols were established for citizens to comply with, so as to ensure that the resulting data have high quality (Goodchild 2007, 218). Other studies sought different modes of citizen engagement to ensure data quality. For instance, the 'escalator model of engagement' involves people entering at a level of participation that matches their needs, interests and abilities, while also encouraging them to move beyond.[15] These studies also stressed that the role of citizens in CS does not need to be limited to data production.

[14] Also see Gabrys et al. (2016) and Freitag et al. (2016) for discussions about data quality and credibility.

[15] https://uclexcites.wordpress.com/2016/04/29/introducing-doing-it-together-science-an-eu-citizen-science-project/.

This means that the interests and abilities of citizens might also be useful in other aspects of data production. Some possibilities include creating spaces of observation and interaction (exhibitions), spheres of collaboration (seminars and conferences), safe, inviting and immersive arenas for discussion (science cafes), opportunities for experimentation and deliberation (workshops) which include planning and coordinating the scientific project's legacy, in terms of intellectual property, innovations management, knowledge transfer and ownership of assets (Ibid). Depending on their interest and background, citizens could be recruited to actively engage in any of these aspects of a scientific project.

We conclude by noting that traditional methods in social sciences and official statistics have a long tradition of engaging closely with citizens as *objects* of study, especially in survey-oriented research and the census. Engaging in CS would require a shift to active participation or contribution from citizens as research *subjects.* This is more closely aligned with a model of CS that conceives of citizens as co-producers. Given the requirements of the scientific principles cited above, together with the call for more inclusive production of science, we suggest that a particular form of CS is required for citizen involvement in the production of official statistics. This form would attempt to provide a combination of statistical science and citizen science. This could also entail that the produced data becomes more representative and inclusive of citizens' concerns, needs and experiences, as well as their own identifications. This is why we propose to use the term 'citizen data' for the aims of this workshop.

Based on the foregoing, we highlight the following questions:

1. What form of citizen engagement (assistants/observers, co-producers or producers) would be most suitable for citizen data, and why?
2. What would be the potential data quality and credibility issues in each form of citizen engagement?
3. What should be the principles of data ownership in each form of citizen engagement?
4. What topics would best address user needs for citizen data that could be collected via a citizen data app?

Smart Statistics

Active participation and contribution from citizens for their own benefit has also been extensively conceptualised and practiced by civil society organisations and foundations. De Waag Society in the Netherlands, Nesta in the UK, and the FabLab-Medialab Prado in Spain are just some examples. What these organisations have in common is their aim to use digital technologies for social, bottom-up innovation. In this line of thinking, innovation occurs in an 'innovation ecosystem' consisting of various communities, such as innovation labs, open hardware and software communities and open data communities (Bria, 2015).

With regard to the production of data by citizens, these initiatives build on the concept of smart cities, understood here as the use of Big Data, urban sensors, Internet of Things and other forms of data collection and data integration to streamline municipal governance and transportation infrastructures, rejuvenate local economies, transform the urban environment, make it more sustainable, liveable, socially inclusive (see, for instance,

Henriquez 2016). While smart cities have been defined in various ways, the concept generally refers to 'how cities are increasingly composed of and monitored by pervasive and ubiquitous computing and, on the other, whose economy and governance is being driven by innovation, creativity and entrepreneurship, enacted by smart people' (Kitchin, 2014). In this view, Big Data enables real-time analysis of city life, new modes of urban governance, and provides for envisioning and enacting more efficient, sustainable, competitive, productive, open and transparent cities. When citizens actively participate, monitor and use data in relation to smart cities they are often referred to as 'smart citizens'.

What then is the potential role of citizens in the development of 'smart' official statistics? We understand smart statistics and its conceptualisation are still in the making, and perhaps permanently ambiguous. Not intending to follow a new 'hype' unreflexively, we propose that the workshop makes explicit one of the directions in which smart statistics could be developed and thereby open it up for discussion. Below, we will first discuss the concept of smart statistics and discuss various ways in which NSIs have recently used various forms of Big Data.

Smart statistics draws on a genealogy of 'smart systems,' such as smart cities, smart energy, smart meters, smart transport, and so on. As a way of thinking, it builds on the massive proliferation of electronic devices and sensors connected to the internet that generate and communicate huge amounts of data.[16] Related to this are developments such as the Internet of Things, smart objects, data exchange, cloud computing, greater automation, artificial intelligence and cognition, which are part of what also is referred to as Industry 4.0 and the Fourth Industrial Revolution. How these might be embedded in statistical production systems such that statistics could be generated in real time and automatically is what smart statistics could be. In this view, data capturing, analysis and processing are envisioned as embedded in activities that generate and simultaneously analyse Big Data. While the focus of thinking and investments over the past few years has been on Big Data, smart statistics would move attention to the automatisation of analysis. The adoption of such smart systems would dramatically transform the production system for official statistics. They would not only transform technical infrastructures, but also facilitate rethinking business processes and architectures, laws and regulations, ethics, methodologies, and so on.

A 'smart' (and existing) example in official statistics is the growing use of paradata for digital or online censuses. Paradata is a type of metadata that is not descriptive as in traditional metadata but consists of 'process' data on respondent's digital actions. Paradata is also sometimes referred to as Big Data because it is generated in 'real-time' and in large volumes and requires processing by algorithms. In relation to online censuses, it includes data about which devices are being used; which buttons (help, back, forward) are being clicked and when; duration spent on each question, questionnaire, and explanatory note; which plausibility checks are triggered; and changes subjects make to answers.[17] For each, inferences are made about the types of questions or groups of respondents more affected than others by a particular element of the questionnaire design; what different answer durations might mean; the relation between responses and the reading of explanatory notes; and how respondent edits and their frequency signal a problem of question wording. Paradata thus involves the tracking of the relation between the census and the respondent through parameters and

[16] Eurostat Big Data Task Force, 'Smart Statistics'. Draft document, Oct 2016.
[17] Statistics Austria (2015).

metrics of data collection[18] and is part of a 'data driven approach' which informs strategies for increasing response rates.[19] It is a by-product of digital technologies that is then put in the service of better calibrating responses.

Smart statistics has also been extended to user interactivity and the dissemination of data in a way that allows users to design their own data tables and combinations. An example is the experimental StatMine application by Statistics Netherlands that aims to facilitate data exploration by users.[20] Smart statistics can furthermore refer to a continuous integration of and adaptation to user demands in the production of statistics. Finally, Statistics Canada offers an example of including citizens as knowledge producers and stakeholders. A pilot on the use of crowdsourcing to map buildings is framed as not only extracting information from citizens, but as a way of providing citizens with insight into their communities (also see section 'Example Projects').[21]

Models of Smart Statistics

In relation to 'smart statistics', how might pervasive and ubiquitous devices and computing be used to compose and monitor populations, businesses and economic phenomena in new ways? Two models have been suggested: using third party systems that exist for other purposes than statistics but from which statistical information can be extracted or developing entirely new data gathering approaches such as sensors and devices exclusively for generating statistical information.[22] In either case, data could be joined up with that from a variety of other sources generated by statistical offices. The latter approach pertains to the objective of the workshop: to develop a citizen data app that could incorporate elements of smart systems but additionally address the concepts of experimentalism, citizen science, and as developed in the next section, privacy by design. In sum, we would call this an approach to smart statistics that facilitates and fosters citizenship.

Based on the foregoing, we highlight the following questions:

1. What citizen feedback and interactive mechanisms could be included? For example, enable citizen interventions (add questions or categories), capacity to add subjective evaluations about their data, provide feedback on the workings of the app?
2. Could a citizen data app have the capacity to change and adapt in relation to citizen input and/or activity as an ongoing feature built into the app? What would be gained/lost?
3. Could a citizen data app provide automated and periodic feedback to individual citizens about his or her data and its relation to groups and populations?

[18] Paradata is identified as a standard of statistical modernization in the Generic Statistical Business Process Model adopted by the High Level Group on the Modernisation of Statistics of the Commission of European Statisticians. Economic Commission for Europe Conference of European Statisticians. 2014. Sixty-second plenary session. Paris, 9-11 April 2014. Fieldnotes.

[19] Not only response rates to online digital census forms might be improved. Statistics Australia had field officers use their personal handheld devices for the conduct of questionnaires (Statistics Australia, 2016).

[20] See http://research.cbs.nl/smp/#/intro.

[21] Statistics Canada (2016).

[22] As proposed in Eurostat Big Data Task Force, 'Smart Statistics'. Draft document, Oct 2016.

4. How could a citizen data app be responsive to both the 'front end' subject and 'back end' user and if so, through what mechanisms?

Privacy by Design

Privacy by design is the fourth concept informing the workshop. It can be understood as the embedding of privacy protection at the software design stage of data collection platforms, devices or applications. It entails designing privacy protection with citizens in mind at the outset and the implementation of these designs in a transparent manner. As such, privacy by design is a response to the problem of privacy, consent, and confidentiality through software. It can be used in tandem with other tools, such as privacy impact assessments. By employing privacy by design, software designers tackle privacy issues at the beginning of the process, in contrast to other approaches that aim at solving privacy issues after software development is complete or leave privacy considerations to legal or regulatory frameworks.

Cavoukian, Taylor, and Abrams (2010) define privacy by design through seven foundational principles:

1. Proactive not reactive; preventative not reactive
2. Privacy as the default
3. Privacy embedded into design
4. Full functionality—positive sum, not zero-sum
5. End-to-end lifecycle protection
6. Visibility and transparency
7. Respect for user privacy

These principles are formulated to assist practitioners in tackling privacy issues before they turn into problems. They require the design to commit to privacy from the beginning and limit data collection to ways that are respectful of individual expectations of citizens. The principles also require the design of any data collection software to take into account that data may exist after the software stops functioning. They also emphasise that the lifecycle of the software must be considered when deciding on how to best protect privacy, including making plans for retiring data once the software reaches the end of its lifecycle. Finally, the principles compel organisations dealing with personal data to be transparent in their goals and to remain accountable to citizens.

Understanding Privacy

Collection and processing of personal data present many other challenges for privacy in addition to individual privacy. Nissenbaum (2004) argues that privacy norms need to be tied to specific contexts. She describes three principles that have dominated debates around privacy throughout the 20th century, namely, limiting surveillance of citizens by governments, restricting access to private information, and curtailing intrusions into private places. She suggests a new term, ‘contextual integrity’, to deal with the new challenges introduced by information technologies. Contextual integrity demands that information gathering is kept

appropriate to the context and obeys the governing norms of distribution within it. The key insight is that norms of distribution vary across culture, historical period, locale, and other factors, and in order to establish contextual integrity, there is a need to be aware of not only the case itself, but also the surrounding social institutions (Nissenbaum, 2009).

Additionally, approaches that aim to protect individual privacy may still lead to undesired outcomes in large-scale data collection efforts. When individually anonymised data are joined to create profiles, the individuals who fit the profile would still experience the effects of the system even when they are not identified individually. For example, Graham (2005) discusses how software can be used to assign different categories to different parts of the city based on school performance, house prices, crime rates, etc., which might potentially orchestrate inequalities and discriminate inhabitants, even when they are not personally identified. Similarly, Zwitter (2014) has also identified and problematised the potential discriminatory 'group effects' of anonymised data.

The use of Big Data also introduces new privacy challenges (ENISA, 2015). Barocas and Nissenbaum (2014) argue that anonymity and consent are often fundamentally undermined in Big Data applications, and that we need other approaches to protect integrity, such as policies based on moral and political principles that serve specific contextual goals and values. Instead of focusing on anonymity in Big Data applications, they instead emphasise securing informed consent, not only as a choice for subjects to waive consent or not, but a requirement that data collectors justify their actions in relation to norms, standards, and expectations. In sum, privacy is not a single thing but depends on the context of production, accountability for group effects and mechanisms of informed consent.

Blockchain Privacy

Recently, scholars working to overcome technical challenges of privacy in Big Data have proposed a method of privacy protection by taking advantage of blockchain technology (De Montjoye, Shmueli, Wang, & Pentland, 2014; G. Zyskind, Nathan, & Pentland, 2015; Guy Zyskind, Nathan, & Pentland, 2015). Blockchain is a distributed computing method where a large number of devices communicate with one another over a shared network, without requiring a central server to authorise the participation of each member or to keep a list of currently connected members. By applying the blockchain technology to privacy, it becomes possible to encrypt and distribute private data over a large network without requiring a trusted central server.

Blockchain is a method for ensuring that different parties connected to the same network agree that the calculations that they perform are valid. The calculations are represented as a "block", and the block can contain any kind of data. A well-known example is the bitcoin[23], which relies on a blockchain where each block represents a financial transaction. In a blockchain implementation, new information is encoded as blocks, and all participants automatically verify its validity at regular intervals[24]. As the description suggests, keeping a

[23] https://bitcoin.org.

[24] A successful update to a blockchain means that all copies of the blockchain have verified the history of operations that came beforehand. In other words, every copy of a blockchain contains its whole history, and this

blockchain is extremely computation intensive even for the simplest of operations. Therefore, while this technology does not require trust in a third party, it does generate a high amount of computational overhead.

The blockchain privacy platform proposed by Zyskind et al. (2015) involves installing and signing up for the service on an individual machine, such as a mobile phone. All information is encrypted and sent to the blockchain, which retains only the location of the data on the public ledger, while the data itself is moved to an 'off-blockchain key-value store', which is not stored on the blockchain, but the transfer can still be verified by the chain. Blockchain privacy methods are intended to solve underlying privacy challenges using a technical framework during software development. They do not stand on their own as the sole solution to preserving privacy, but rather supplement the legal policy-oriented considerations such as contextual integrity, and software design methodologies such as privacy by design.

Based on the foregoing, we highlight the following questions:

1. What forms of consent would be appropriate and relevant, for example, blanket consent for statistical purposes or capacity to limit consent on the part of the citizen for a time period?
2. Does a citizen data app introduce new data security vulnerabilities and if so, what are they?
3. Would blockchain privacy be useful for statistical data production? Why/why not?
4. What other technical options are there to secure privacy by design?

Example Projects

Here we highlight three examples that might inspire and inform the workshop. We selected these examples from a review of 11 projects performed by organisations such as Nesta, De Waag, the MIT Center for Civic Media and the Goldsmiths Citizen Sense research project (see Appendix for the list of projects). The review presented here does not aim to be comprehensive but rather focuses on learning about a variety of methods and organisational perspectives. That said, the three examples discussed below might not all be directly applicable to official statistics. We offer them here to raise ideas, questions and concerns. After discussing each example we list broader considerations following from our initial review.

Example 1: Promise Tracker by MIT Centre for Civic Media

The first example includes a questionnaire designed and conducted by citizens and led by the MIT Centre for Civic Media. This centre aims to develop applications (apps, websites, and portals) that foster civic engagement and political action. The projects need to be developed and adopted by community organisations and local governments. Consequently, the aim of

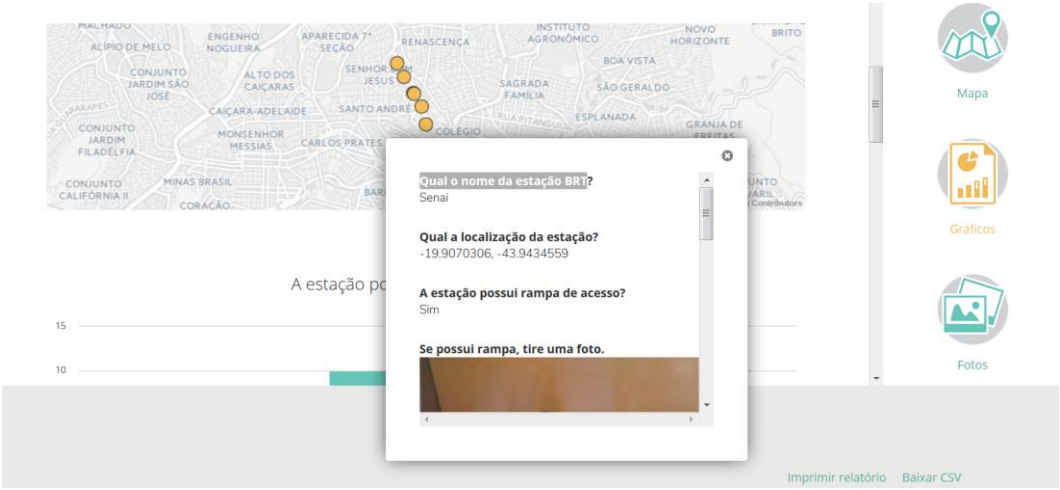
history cannot be altered. If an individual party changes this history, for example in an attempt to create bitcoins where none existed, at the next verification step it would be rejected, as it would differ from all the other copies.

the Centre is not to produce data and statistics. Yet, as we will see, the data collected may contribute to this process.

The incentive for the Promise Tracker was recent Brazilian legislation obliging politicians to publish the promises they make during election rounds in specified lists. The Promise Tracker allows app users to keep track of whether or not politicians fulfil these promises by monitoring the specified promise lists after the elections. Citizens are engaged in data production to the extent that they monitor politicians' promises in their neighbourhoods and feed them back to key stakeholders. The issues they report on are at least as diverse as the promises given to them, such as monitoring of the construction of elementary schools in their districts and the accessibility of bus stations.

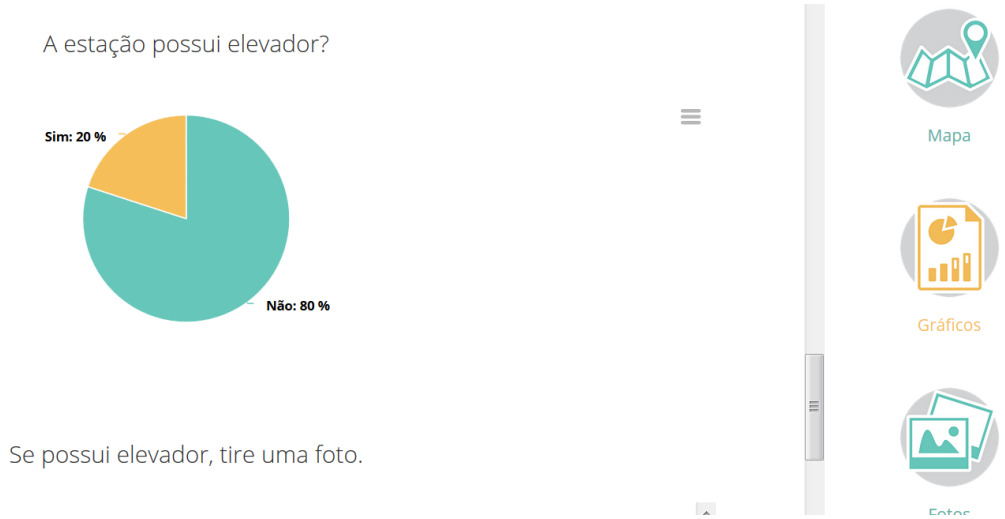
The mobile app allows citizens to create mobile surveys about the solution of a problem and to disseminate the surveys to the phones of other community members. This enables citizens to track progress in real time, and to expose areas and issues where politicians have not adhered to their promises. The app also allows other data to be collected and connected to geo-location data, such as matching qualitative descriptions or photos with quantitative data produced through surveys and other tracking methods. Citizens can visualise the data through a web platform as illustrated below. The collected data are released under a Creative Commons Zero license so they can be used by other advocacy groups (DataShift).

Figure 2: Screenshot of the Promise Tracker for Belo Horizonte, Brazil



Source: http://monitor.promisetracker.org/pt-BR/campaigns/25/share (date of access: 9 January 2017).

Figure 3: Screenshot of the Promise Tracker for Belo Horizonte, Brazil



Source: http://monitor.promisetracker.org/pt-BR/campaigns/25/share (date of access: 9 January 2017).

We have chosen to discuss this example because it provides a number of potentially inspirational ideas. Firstly, the Promise Tracker produces data on new indicators considered relevant by citizens, rather than traditional survey methods, where categories, their definitions as well as the formulation of questions are predominantly decided by statistical authorities. This means that citizens can be actively engaged in data production, not only as *respondents,* but as *survey developers* and *analysts.* Secondly, the project proposes a radically different take on surveys in terms of purposes, uses and methods. Citizen designed surveys may not always adhere to established survey methodologies and analyses but can introduce more descriptive and interpretive approaches. Thirdly, the project enables citizens to combine different data types (qualitative, quantitative, visual, textual and numerical) when seeking answers to their questions. Again, while departing from traditional survey methodologies, data can be richer in detail but statistical interpretation less straightforward.

For the aims of our workshop, relevant learning points offered by the organisers and participants are the difficulties of translating election promises into clear monitoring goals and indicators. This is an important point that raises questions such as 'Is everything quantifiable?' and 'Do the monitoring goals and indicators developed by the app developers match those meant and understood by citizens?' Another point we highlight here is that the project documentation does not offer information about privacy and the ownership of data. Given that geo-locations are identified based on users' mobile data use, we suggest that privacy is a concern.

We thus learn that citizen surveys raise a number of statistical and data quality issues as they do not satisfy criteria of traditional survey methods. The co-production of surveys could possibly address some of these issues but would require rethinking methods and forms of data analysis and interpretation that could address the richness and complexity of this data.

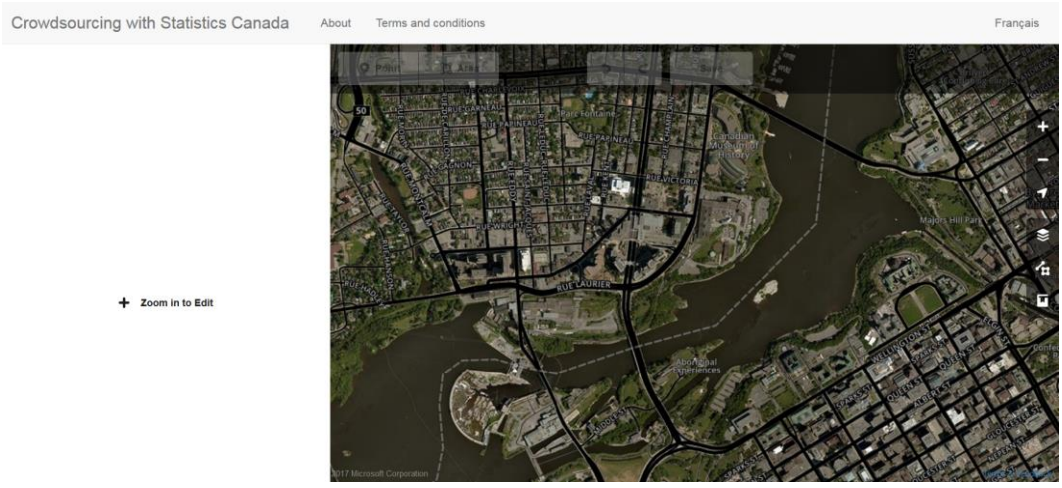
Example 2: Crowdsourcing by Statistics Canada

Statistics Canada initiated a crowdsourcing pilot in October 2016 to address a lack of accurate national level statistics on buildings and their attributes that can be used to compare specific local areas (Statistics Canada, 2016).[25] The aims of this pilot are to fill current data gaps and to learn about the possibilities of crowdsourcing for statistical data production.

The data are collected via a third-party website: Open Street Map (OSM). OSM is a collaborative project designed to create a free and editable map of the world. The application for Statistics Canada allows users to select a geolocation and edit, for instance, the name of a street. Users may also label a road a 'one-way' or mark set of buildings as a pool. According to the OSM website, 97 people had made edits by 13 December 2016.[26]

Interestingly, users may present themselves on the website through a user name or an alias. At times they share personal information (for example, sharing that you have lived here all your life), at other times, they may remain entirely anonymous. User information is publicly shared, as well as a history of edits. Sharing the GPS information of the user device is optional, which provides some level of privacy to the users.

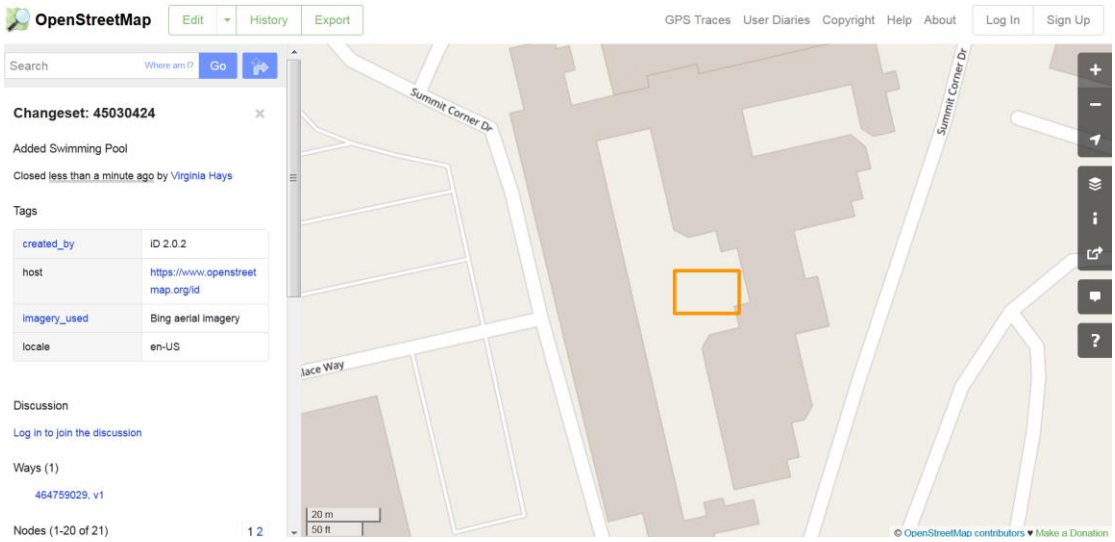
Figure 4: Screenshot of Statistics Canada crowdsourcing application (Ottawa)



Source: https://crowdid.osmcanada.com/#background=Bing&map=15.00/45.4241/-75.7106

[25] The pilot is 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.
[26] https://crowdid.osmcanada.ca/#background=Bing&map=12.00/45.4241/-75.7106.

Figure 5: Screenshot example of user modification



Source: https://www.openstreetmap.org/changeset/45030424

Given that this is a very new project, Statistics Canada has not yet made statements about the reliability and usability of the data produced. Regarding reliability, Statistics Canada states that all responsibility falls on the third-party website. As more data are fed to the website, it will become clearer whether the reliability of data will be a concern for its usability.

In terms of privacy, at this early stage of the pilot it is also unclear whether publicly shared user information (such as 'lived there all my life') will be sufficiently anonymous for sparsely populated areas. Statistics Canada does not comment on other potential privacy issues that might arise from sharing personal information. A final issue we address in relation to this project is data ownership: what will the users get out of correcting and editing OSM?

In sum, this example is inspirational for our workshop not only because it attempts to use crowdsourced data for official statistics, but because it uses existing platforms to collect the data and engages users at various levels with both the data and the platform.

Example 3: Citizen Sense, Goldsmiths

The final example draws on the use of sensors by citizens to collect information about their environment. Citizen Sense (ERC project led by Jennifer Gabrys, Goldsmiths) investigates the relationship between practices of environmental sensing and citizen engagement.[27] With the help of smart phones and other networked devices, users are able to engage with similar modes of environmental observation and data collection. This project aims to understand the effectiveness of these practices in not only crowd-sourced data, but also in giving rise to new modes of environmental awareness and practice.

[27] http://www.citizensense.net/

The project comprises different types of sensing: 1) Pollution sensing: for examining the use of digital environmental sensors to monitor and report on environmental pollution; 2) wild sensing: for investigating the use of digital environmental sensors in citizen-sensing projects to study and record flora and fauna activities; 3) Urban sensing: for investigating the use of digital environmental sensors in urban environments to promote sustainability (see figure 6). It uses low-tech sensor kits distributed to participating citizens and these kits transmit data to both the researchers' and users' devices, allowing them to track air quality in real time.

The data were collected in addition to Citizen Sense 'logbooks' that asked for input on key environmental and health concerns related to natural gas infrastructure. In addition, citizens were able to analyse their own data using a Data Analysis Toolkit. In order to anonymise the exact monitoring locations and protect the privacy of the citizens, the monitoring locations have been labelled with township locations instead.

Figure 6: An example of the Dustbox hardware for measuring urban air pollution

Source: http://citizensense.net/kits/dustbox-hardware/

The project thus provides examples of using sensors. We highlight two aspects that might be of interest for the workshop. First, this project generates and combines data generated by different devices. Second, this data type does not adhere to the quality criteria of official measurements, yet is able to raise new issues and concerns in ways that official quality criteria do not make it possible (Gabrys et al., 2016).[28]

One difference between Citizen Sense and the aforementioned examples is that Citizen Sense used township locations, rather than the exact geo-locations of citizens to protect their privacy. This is an important difference because it highlights that exact location data might not always be needed (or desired) depending on the research aims and privacy concerns of the citizens participating in data production.

Other Social Science Examples

There are several examples of engaging citizens as assistants/observers, co-producers or producers of data in social sciences. Although attempting to include them here would be an ambitious undertaking, we summarise recent examples presented at the *Smart Ethnography Workshop*, held at Goldsmiths, University of London (December 6, 2016). One example study was conducted by anthropologists Mark Johnson, Maggy Lee and Mike McCahill. Their research "Smart Ethnography and Surveillance Subjectivities among Transnational Migrants in Hong Kong" aims to enable research participants to be actively part of data production throughout their fieldwork. In order to facilitate this, the researchers distributed smart phones to domestic workers and asked them to record their day-to-day lives, experiences and feelings on a specially designed app. By using the app, study participants not only provided timely data to the researchers, but they also had a say on what kinds of data came to be produced. While some workers just typed notes about their experiences, others took photographs, recorded songs or videos to document their lives. The smart phones enabled the researchers to immediately respond to the data they received from the research participants, for instance, asking them to clarify what they meant by a poem they sent or a picture. This way, the research shifted both the *active* researcher, *passive* participant dynamics of research project, and subsequently, enabled the co-production of rich, contextual and informative data.

Johnson, Lee and McCahill's study presents important points about the co-production of data with research participants. While their participants might not be familiar with ethnographic research methodologies, they were familiar with keeping a diary. In other words, although research skills and knowledge of methodologies might be limited to a small numbers of scientists and professionals, citizens also possess important everyday skills and knowledge that can become part of data collection. But, as this project demonstrates, this requires that researchers be open to working with new forms of data.

[28] In a project measuring air pollution in Amsterdam, experts were invited to comment on citizen produced data. It appeared data quality was lower than the data generated by municipality operated sensors. One expert stated that lower data quality was acceptable in relation to the aims of the project. A municipality affiliated expert, however, dismissed the results. His concern was that imprecise measurements might cause unnecessarily distress to citizens, leading to citizen dissatisfaction with local government (Waag Society Amsterdam, 2014).

Another study is by Paola Prieto López and Raluca Mariana Pinzari. In their study, "Mapping Urban Engaged Encounters: Alternative Ways of Sensing the City", the researchers have developed an app with their study participants instead of relying on pre-designed applications available for smart phones. In this study, researchers and participants recorded events they have heard of or have attended on a multi-platform mapping software so as to create an 'alternative' cultural map of their cities. While still at an early stage, the study has already begun to raise important questions for data production, which are also relevant for this background paper. For instance, should this co-produced app be open to public access or only limited to those who record events and details on the app? Should it, otherwise, be limited to only those who have been trusted by at least a current member of the app? Researchers have expressed their concerns about online trolls who might destroy the already fragile software structure of the app or they might come and delete all previously stored data. In that case, who would be liable and accountable for data protection and privacy of all the people who have openly shared the details of the 'alternative' events they have attended in the city? Presumably, in smaller and more conservative cities, even attending these events might potentially carry certain risks. How could the app continue to exist (and grow) and yet solve these privacy issues, while at the same time being more inclusive (open access or otherwise)?

Conclusion

In this document, we have provided a background for a workshop conducted in the spirit of experimentalism. The aim is to specify design principles for an app/platform co-produced with citizens that generates citizen data for official statistics. This enables us to conceive part of a new regime of data collection in official statistics that includes various devices and technologies. Here we referred to this regime as 'smart statistics'. Considering insights from the fields of citizen science and civic engagement, citizens might not only be included as subjects, observers or data producers, but also as co-producers, stakeholders and citizens.

For official statistics, we propose that a citizen data approach has the potential to produce new statistical variables desired and identified by users and citizens, and increase identification with official statistics which might facilitate trust in official statistics. However, these approaches also challenge the principles of data quality, reliability, privacy, security and anonymity. Moreover, they raise important privacy and ethical questions. In this regard, the issues that the workshop will address are not dissimilar to those that arise in relation to Big Data sources. However, designing a bespoke data generation app/platform has the advantage of potentially measuring, documenting and mitigating data principles and issues. It could moreover incorporate citizens' demographic data, their subjective evaluations and feedback. Furthermore, like Big Data sources, a citizen app/platform does not need to replace existing sources. It can also be auxiliary and supplementary to more traditional and longstanding statistical sources.

Considering the novelty of this approach and the (unknown) challenges of the project, the workshop may or may not lead to applicable outcomes. At the very least, the applied work of conceptualising a citizen data app/platform may contribute to specifying the role(s) of citizens in a smart statistics regime.

References

Barocas, S., & Nissenbaum, H. (2014). Big Data's End Run Around Anonymity and Consent (pp 44-75). In J. Lane, V. Stodden, S. Bender, & H. Nissenbaum (Eds.), *Privacy, Big Data, and the Public Good*. New York: Cambridge University Press.

Binder, T., Brandt, E., Ehn, P., & Halse, J. (2015). Democratic Design Experiments: Between Parliament and Laboratory. *CoDesign*, *11*(3–4), 152–165. https://doi.org/10.1080/15710882.2015.1081248

Bria, F. (2015). *Growing a Digital Social Innovation Ecosystem for Europe: DSI Final Report*. European Commission. Retrieved from https://waag.org/sites/waag/files/public/media/publicaties/dsireport.pdf

Callon, M., Burchell, G., Lascoumes, P., & Barthe, Y. (2011). *Acting in an Uncertain World: An Essay on Technical Democracy*. Cambridge, Mass.: MIT Press.

Cavoukian, A., Taylor, S., & Abrams, M. E. (2010). Privacy by Design: Essential for Organizational Accountability and Strong Business Practices. *Identity in the Information Society*, *3*(2), 405–413. https://doi.org/10.1007/s12394-010-0053-z

CRESC Encounters Collaborative. (2013). (Un)doing collaboration: reflections on the practices of collaborative research. CRESC: University of Manchester and The Open University.

DataShift. (Not dated). *Case Study: Promise Tracker* (pp. 1–4). DataShift. Retrieved from http://civicus.org/thedatashift/wp-content/uploads/2015/07/Promise-tracker-case-study.pdf

De Montjoye, Y.-A., Shmueli, E., Wang, S. S., & Pentland, A. S. (2014). OpenPDS: Protecting the Privacy of Metadata through SafeAnswers. *PLOS ONE*, *9*(7). https://doi.org/10.1371/journal.pone.0098790

ENISA. (2015). *Privacy by Design in Big Data: An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics*. European Union Agency For Network And Information Security. Retrieved from https://www.enisa.europa.eu/news/enisa-news/privacy-by-design-in-big-data-an-overview-of-privacy-enhancing-technologies-in-the-era-of-big-data-analytics

Freitag, A., Meyer, R., & Whiteman, L. (2016). Strategies Employed by Citizen Science Programs to Increase the Credibility of Their Data. *Citizen Science: Theory and Practice*, *1*(1). https://doi.org/10.5334/cstp.6

Gabrys, J., Pritchard, H., & Barratt, B. (2016). Just Good Enough Data: Figuring Data Citizenships Through Air Pollution Sensing and Data Stories. *Big Data & Society*, *3*(2), https://doi.org/10.1177/2053951716679677

Goodchild, M. F. (2007). Citizens as Sensors: The World of Volunteered Geography. *GeoJournal*, *69*(4), 211–221.

Graham, S. (2005). Software-Sorted Geographies. *Progress in Human Geography* 29(5). 562–580.

Gross, M. (2010). *Ignorance and Surprise: Science, Society, and Ecological Design*. Cambridge MA and London: MIT Press.

Gross, M., & Krohn, W. (2005). Society as Experiment: Sociological Foundations for a Self-Experimental Society. *History of the Human Sciences*, *18*(2), 63–86. https://doi.org/10.1177/0952695105054182

Guggenheim, M. (2012). Laboratizing and De-Laboratizing the World Changing Sociological Concepts for Places of Knowledge Production. *History of the Human Sciences*, *25*(1), 99–118. https://doi.org/10.1177/0952695111422978

Haraway, D. J. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, *14*(3), 575–599.

Henriquez, L. (n.d.). *Amsterdam Smart Citizens Lab: Towards Community Driven Data Collection*. Amsterdam: De Waag Society and AMS Institute. Retrieved from https://www.waag.org/sites/waag/files/public/media/publicaties/amsterdam-smart-citizen-lab-publicatie.pdf

Hickman, T. (2016). *Smart Participation*. Presented at the Smart Ethnography Workshop, London, 16 December 2016.

Irwin, A. (1995). *Citizen Science: A Study of People, Expertise and Sustainable Development*. London and New York: Routledge.

Jasanoff, S. (2003). Technologies of Humility: Citizen Participation in Governing Science. Minerva, 41(3), 223–244.

Johnson, M., Lee, M., & McCahill, M. (2016). *Smart Ethnography and Surveillance Subjectivities among Transnational Migrants in Hong Kong*. Presented at the Smart Ethnography Workshop, London, 16 December 2016.

Jungnickel, K. (2017). Making Things to Make Sense of Things: DIY as Research Subject and Practice. In J. Sayers (Ed.), *The Routledge Companion to Media Studies and Digital Humanities*. Oxon: Routledge.

Kitchin, R. (2014). The Real-Time City? Big data and Smart Urbanism. GeoJournal, 79(1), 1–14. https://doi.org/10.1007/s10708-013-9516-8

Kullenberg, C., & Kasperowski, D. (2016). What Is Citizen Science? – A Scientometric Meta-Analysis. *PLOS ONE*, *11*(1), e0147152. https://doi.org/10.1371/journal.pone.0147152

Latour, B. (1993). *The Pasteurization of France*. Harvard University Press.

Latour, B. (2006). Which Protocol for the New Collective Experiments? *Boletín CF+S*, (32/33). Retrieved from http://habitat.aq.upm.es/boletin/n32/ablat.en.html

Levidow, L., & Carr, S. (2007). GM Crops on Trial: Technological Development as a Real-World Experiment. *Futures*, *39*(4), 408–431. https://doi.org/10.1016/j.futures.2006.08.002

Lezaun, J., Marres, N., & Tironi, M. (forthcoming). Experiments in Participation. In C. Miller, U. Smitt-Doer, & R. Fouche (Eds.), *Handbook of Science and Technology Studies* (Vol. 4). Cambridge: MIT Press.

Marres, N. (2012). *Material Participation: Technology, the Environment and Everyday Publics*. Basingstoke: Palgrave Macmillan.

Millo, Y., & Lezaun, J. (2006). Regulatory Experiments: Genetically Modified Crops and Financial Derivatives on Trial. *Science and Public Policy*, *33*(3), 179–190. https://doi.org/10.3152/147154306781779046

Muniesa, F., & Linhardt, D. (2011). Trials of Explicitness in the Implementation of Public Management Reform. Critical Perspectives on Accounting, 22(6), 550–566. https://doi.org/10.1016/j.cpa.2011.06.003

Nature. (2015). Rise of the Citizen Scientist. Nature News, 524(7565), 265. https://doi.org/10.1038/524265a

Nissenbaum, H. (2004). Privacy as Contextual Integrity. *Washington Law Review*, *79*(1), 119-158.

Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.

Prieto López, P., & Pinzari, R. M. (2016). *Mapping Urban Engaged Encounters: Alternative Ways of Sensing the City*. Presented at the Smart Ethnography Workshop, London, 16 December 2016.

Rabinow, P., & Bennett, G. (2012). *Designing Human Practices: An Experiment with Synthetic Biology*. University of Chicago Press.

Rheinberger, H. J. (1995). From Experimental Systems to Cultures of Experimentation. In *Theories and Rationality in the Biological Sciences: The Second Pittsburgh-Kanstanz Colloquium in the Philosophy of Science* (pp. 107–123). Konstanz: UVK, Universitätsverlag Konstanz.

Ruppert, E., Harvey, P., Lury, C., Mackenzie, A., McNally, R., Baker, A. B., Lewis, C. (2015). Socialising Big Data: From Concept to Practice. Retrieved from http://www2.warwick.ac.uk/fac/cross_fac/cim/research/socialising-big-data/sbd_wp_2015.pdf

Sánchez-Criado, T. (2016). *The Ethnographic Experiment, Revisited: Experimental Collaborations, or the 'Devicing' of Fieldwork for Joint Problem-Making*. Presented at the Workshop: The New Experimentalisms, London, 9 September 2016.

Scheel, S., Cakici, B., Grommé, F., Ruppert, E., Takala, V., & Ustek-Spilda, F. (2016). *Transcending Methodological Nationalism through Transversal Methods? On the Stakes and*

Challenges of Collaboration (ARITHMUS Working Paper No. 1). Goldsmiths College, University of London.

Statistics Australia. (2016). *Register-Free Evolution and Revolution in Australia*. Presented at the Eurostat Towards More Agile Social Statistics Conference, Luxembourg.

Statistics Austria. (2015). *Integrating the Web Mode in the Austrian Household Budget Survey 2014/15*. Presented at the Eurostat New Technologies and Techniques in Statistics Conference, Brussels.

Statistics Canada. (2016, August 23). Crowdsourcing. Retrieved January 9, 2017, from http://www.statcan.gc.ca/eng/crowdsourcing

Waag Society Amsterdam. (2014). *Eindrapportage Smart Citizen Kit Amsterdam: Meten is Weten?* (pp. 1–21). Amsterdam: De Waag Society. Retrieved from https://waag.org/sites/waag/files/public/media/publicaties/eindrapportage-sck-asd.pdf

Waterton, C., & Tsouvalis, J. (2015). On the Political Nature of Cyanobacteria: Intra-Active Collective Politics in Loweswater, the English Lake District. *Environment and Planning D: Society and Space*, *33*(3), 477–493. https://doi.org/10.1177/0263775815594305

Zwitter, A. (2014). Big Data Ethics. *Big Data & Society 1(2). doi:10.1177/20539517145590253*.

Zyskind, G., Nathan, O., & Pentland, A. (2015). Decentralizing Privacy: Using Blockchain to Protect Personal Data. In *2015 IEEE Security and Privacy Workshops (SPW)* (pp. 180–184). https://doi.org/10.1109/SPW.2015.27

Zyskind, G., Nathan, O., & Pentland, A. (2015). Enigma: Decentralized Computation Platform with Guaranteed Privacy. *arXiv Preprint arXiv:1506.03471*. Retrieved from http://arxiv.org/abs/1506.03471

Appendix: Reviewed Citizen Science and Citizen Data Resources and Examples

- Amsterdam Smart Citizens Lab, De Waag Society: https://www.waag.org/sites/waag/files/public/media/publicaties/amsterdam-smart-citizen-lab-publicatie.pdf
- MIT Center for Civic Media: https://civic.mit.edu/
- Doing It Together Science: http://togetherscience.eu/
- CitizenScience.org: http://citizenscience.org/
- Commons Lab, Wilson Center: https://wilsoncommonslab.org/
- Digital Social Innovation Reports and Mapping, De Waag Society, Nesta and SUPSI, https://waag.org/sites/waag/files/public/media/publicaties/dsi-2nd-interim-report.pdf
- Citizen Sense, Goldsmiths, University of London: http://www.citizensense.net/
- Health-Lab, Amsterdam Region, Zorg en ICT: http://waag.org/sites/waag/files/public/Publicaties/healthlab_2013-en-lr.pdf
- Impfpolitik mit Citizen Science, University of Vienna: http://citizenscience.univie.ac.at/mitforschen/
- OpenNorth: http://www.opennorth.ca/
- Promise Tracker, DataShift: http://civicus.org/thedatashift/learning-zone/case-studies/