Methods as forces of subjectivation: experiments in the remaking of official statistics

Baki Cakici & Evelyn Ruppert

Published online: 05 Nov 2019.

Submit your article to this journal ⬈

View related articles ⬈

View Crossmark data ⬈

Routledge
Taylor & Francis Group

Methods as forces of subjectivation: experiments in the remaking of official statistics

Baki Cakici [a] and Evelyn Ruppert [b]

aTechnologies in Practice, IT University of Copenhagen, Copenhagen, Denmark; bDepartment of Sociology, Goldsmiths, University of London, London, UK

ABSTRACT

We develop the concept of methods as 'forces of subjectivation' in relation to experiments we have encountered in a study of government methods for generating official population statistics. These experiments problematise the subjects of traditional methods based on paper questionnaires and offer new digital technologies and data sources as possible solutions. We reflect on these experiments in relation to recent work on sociological and digital research methods as inventive and live. What this work identifies in relation to questions of research methods we take up to think about government methods in two ways. One concerns how government method experiments offered as solutions to problematic subjects, once put into action, change initial problem formulations and are inventive of new ones. Secondly, they are also inventive of their subjects who do not pre-exist but come into being through the agential capacities that methods configure. Both aspects of methods, we argue, are the result of the interactions and dynamics between human and technological actors, the outcomes of which cannot be settled in advance.

Introduction

There is much attention to the valuation and uptake of new digital technologies and new sources of data such as that generated by social media and search engines or government administrative databases. For National Statistical Institutes (NSIs) this is taking the form of experiments with big data and digital devices such as smartphones and online platforms as alternatives to their traditional methods that have long relied on data collected via paper questionnaires.[1] NSIs typically focus on the technical challenges of these experiments and their validity and veracity to generate population statistics. However, our interest is different. A critical driver of these experiments is how subjects are problematised in relation to how they answer and respond to paper questionnaires and how new digital technologies and data are offered as solutions. That is, solutions are founded on conceptions of and relations to subjects that seek to direct and configure their agential capacities in the making of population knowledge. Yet, experiments also engage other actors including different technologies and digital platforms and, as we explore in this article, are inventive of new problematic subjects when methods come into play.

It is in relation to this understanding that we attend to how methods of constituting populations are simultaneously about identifying and knowing through different forms of subjectivation (which we define below) that involve both humans and technologies. There are many practices of identification such as a passport (e.g. see (Torpey 2000)) but our interest is in the numerous dispersed government

sites involved in the practice of identification for the purposes of enumerating and knowing populations.[2] Enumeration is unique in the quest to identify and know whole populations and provide a base and reference for other methods such as surveys. But there is another difference that is central to our focus. It builds on the understanding that subjects do not pre-exist but are brought into being by methods (Ruppert 2011). For us, one thing at issue in experiments with new digital technologies and data is that subjects are being constituted in new ways because methods are made up of different socio-technical arrangements that form what we refer to as forces of subjectivation. As we will develop here, it is in relation to and acting through these forces that people are subjectified and the possibilities and potentials of their agency and how they are identified are configured.

It is along these lines that we consider how methodological changes in the pursuit of solutions to practical governing problems involve the invention of new ways of thinking and knowing that enact both subjects and populations (Osborne *et al.* 2008). It is by identifying and categorising subjects that the referent object of governing – that is, population – is enacted and its nature is then known, managed, regulated, maximised and governed (Foucault 1997). However, this requires not only a language to describe the object of government, but also 'the invention of devices to inscribe it' (Rose 1996, p. 70). Methods are such devices and enable a population to be known, evaluated, diagnosed and acted upon. But, rather than simply reflections, different methods enact different populations as objects of concern and intervention (Ruppert 2011). Correspondingly, methods do not simply reveal subjects as already formed but produce them and the particular capacities and agencies required for them to become part of a population. In this article, our interest is in the latter and how methods can be understood as forces of subjectivation. We first outline this understanding and next consider recent work on sociological and digital methods as 'inventive,' 'live' and having a 'social life' to interpret experiments in government methods for knowing populations. We then connect this work to how subjects are problematised in relation to traditional government methods that involve paper questionnaires and illustrate this in the example of the introduction of a voluntary question on religion in the 2001 census of England and Wales. Next, we examine how this problematisation was related to the design of two solutions we encountered in an empirical study of changing government methods across Europe.[3] 'Social media solutions: inferring subjects' focuses on experiments with social media as an alternative data source for generating live data about the dynamics of student internal migration and 'Digital census solutions: calibrating subjects' examines digital and online censuses as methods inventive of new relations between governments, technologies and subjects. In conclusion, we reflect on these analyses to consider the consequences of methods for subjectivation and for the object of methods, that is, the populations that are enacted.

The forces of subjectivation that make up identification

Our conception of subjectivation draws on Michel Foucault's formulation, which recognises subjects as being capable of reflection, self-formation and engaged in struggles against direct domination (Foucault 1982). While criticised for his assumed dismissal of political agency, our understanding comes from his later writings where he engaged with the concept of autonomy. As Cremonesi *et al.* (2016) note, he thought of autonomy in relation to techniques of the self through which subjects act in relatively autonomous ways and in his writings on enlightenment where he addressed how individuals work to transform themselves. Because these were relatively minor references, understandings of Foucault's conception of subjectivation have tended to focus on the asymmetrical production of the subject in relation to techniques of power. This is an argument advanced by Rossi (2016) in her account of Foucault's under-theorisation of resistance. However, Rossi, as well as many other researchers such as Rose (1999), Balibar (1991) and Isin (2012), have highlighted Foucault's insistence that the condition of possibility of techniques of subjection is the potential to resist, challenge and shape power relations. In feminist thought, Butler (1997), for example has also highlighted the potential for subjects to subvert and resist through practices of recitation and reiteration that challenge and resignify norms. Taking up this understanding, Cremonesi *et al.* (2016) refer to

this as the coupling of subjection-subjectivation to capture that power is not possessed but is a relation and process. This relation and tension between governing and technologies of the self are well captured in Foucault's conception of subjectivation:

> On the other hand, a power relationship can only be articulated on the basis of two elements which are each indispensable if it is really to be a power relationship: that 'the other' (the one over whom power is exercised) be thoroughly recognized and maintained to the very end as a person who acts; and that, faced with a relationship of power, a whole field of responses, reactions, results, and possible inventions may open up (Foucault 1982, p. 789).

Balibar (1991) builds on this formulation by drawing a distinction between being a subject to power and a subject of power. He argued that being a subject *to* power means to be dominated by and obedient to a sovereign. However, when a subject submits to power this opens the possibility to be subversive and be a subject *of* power. For Balibar, it is through submission to power that the possibilities of being subversive are made possible. This, he argues, is what distinguishes citizens from subjects: they are simultaneously subjects of obedience, submission, and subversion, and each is an always-present potential.[4] It is with this understanding of power that we consider subjects and their agential capacities in relation to methods. Rather than thinking of subjects as free or controlled, we seek to understand the complexities of acting in relation to methods and the possibilities for being subversive. That is, subjects act in relation to the sociotechnical arrangements of both humans and technologies that make up methods and format how they are identified. The question for us then is how are subjects identified in relation to different methods and what are the forms of power at work in their subjectivation?

We situate this question in relation to recent work on sociological methods and the making and knowing of social worlds raised in three edited collections. Lury and Wakeford's (2012) collection provides an inventory of methods for doing research-oriented to the inventiveness and 'open-endedness of the social world' (2). One line of inquiry they note concerns the recognition that humans are not the only source of agency in the working of methods and that all actors are active participants. A consequence they identify, and which we specifically address, is that while methods configure sociotechnical arrangements to address specific problems – what Foucault called problematisations – when put into action, the interactions and dynamics between human and technological actors can change initial problem formulations. That is, problems cannot be secured in advance because of the inventiveness of methods. As Neyland and Milyaeva (2016) note in relation to market interventions, problems are not settled and given but often reworked, transformed or lead to further problems. From climate change to vaccines, they explore how problems, solutions and interventions are entangled and dynamically reformulated. Back and Puwar (2012) take this up differently in their call for 'live sociology'. Amongst some eleven propositions, they highlight that one meaning of live sociology is to engage with the possibilities of adapting and repurposing data generated by online platforms for sociological research that are 'real-time' and 'live' investigations of social worlds. And finally, Law and Ruppert (2013) explore research methods as devices and the 'social work' that they do. In 'The Social Life of Methods: Devices' they argue for being reflexive about the work being done by devices and attending to normative and political choices and their consequences for how methods intervene in the social worlds of which they form a part.

While the foregoing work addresses a variety of sociological methods, a related strand concerns digital research methods and their relative instability and indeterminacy as a result of their entanglement with the sociotechnical arrangements of digital platforms. For some researchers, this reliance can result in methods being 'compromised' because platforms configure what is collected and made into data, and, in turn, the forms of analysis and knowledge that are possible (Langlois *et al.* 2015). Thus, the development and deployment of digital research methods not only face regulated or monetised access to data, but are entangled with the hidden assumptions, different objectives and biases of platforms (Bruns and Burgess 2015). These are issues Marres and Gerlitz (2016) address in their consideration of the resemblances and affinities between digital research methods and the techniques and digital analytics of platforms such as Twitter. Rather than critique such entanglements and

resemblances, Marres and Gerlitz suggest that digital social research methods can 'interface' with the technical arrangements of platforms. This, they argue, can be achieved through methodologies that adjust and adapt to digital platforms by assembling different techniques and practices and, in turn, make platforms productive for social inquiry.

What does this work on sociological and digital methods mean for the question we posed above? For one, the methodological issues they raise resemble and have affinities with the government experiments with digital methods for knowing populations that we analyse in this article. As we will argue, those experiments can be similarly conceived of as interventions that respond to and are inventive of problematisations; rely upon but adapt to and repurpose data generated by digital platforms; and are real-time and live investigations that interface with and adjust to the sociotechnical arrangements of platforms to meet their knowledge and governing objectives. At the same time, our question raises an additional issue: how do these features of experiments also invent subjects and constitute different forms of subjectivation? As we will develop in our empirical analyses, it is through a series of relations between subjects and sociotechnical arrangements or 'method agencements' that the possibilities of action are being configured in particular ways. Methods require the actions of subjects – whether through the selection of a tick box or the entry of a location in a free-text field – who participate in their identification in relation to their forces of subjectivation. Changes in methods thus constitute reconfigurations of agencements, and, in turn, the possibilities and potentials of agency.

As Lury and Wakeford (2012) argue, methods are designed in response to problems. For government methods, this involves numerous problematisations such as the cost, efficiency and timeliness of methods, but our interest concerns their problematisations of the 'right' or 'truthful' identification of subjects, or the anticipation of problematic subjects. While no method can direct subjects to one and only one way of acting, they are arranged to manage, guide or calibrate what subjects do. In other words, methods anticipate how a subject might *potentially* act and identify and seek to manage, direct and channel those possibilities. As other researchers have elaborated, such anticipatory logics underpin both governing and technical practices and are speculative regimes and forces (Adams *et al.* 2009) that tend to get internalised by subjects (Mackenzie 2013). Yet, how subjects perform when methods are in action is uncertain and can be generative and inventive of unexpected outcomes.

In the following sections, we explore these issues first in relation to how subjects are problematised in census methods based on the technology of paper questionnaires. We then examine how digital technologies and data generated by platforms are offered as solutions but then give rise to new problematisations.

The problematised subject

The problem, however, is to get the respondent to answer these questions.[5]

Amongst the principles of official statistics that guide the configuration of national census methods is that of 'non-excessive burden on respondents' (United Nations 2014) and choosing data sources that 'consider the burden on respondents' (Eurostat 2011). Quality, cost, timeliness of data and so on are also key, but it is in relation to burden that subjects are problematised.[6] While the conception of burden can be questioned or challenged, what is of interest to our argument is how the principle has been taken up and shaped statistician's conception and anticipation of subjects and has been translated into methods.

Respondent burden is most often articulated in relation to the traditional paper questionnaire. How then are burdened subjects imagined?[7] They are conceived as problematic because they do not respond because they are over-burdened by numerous state data collection activities. Burden is blamed, in turn, for a decline in response rates to paper questionnaires. But other problems are also at fault such as subjects' concerns about privacy and confidentiality in the face of ever-increasing volumes of data that states collect. Others include 'hard-to-count' populations such as transient

students or renters who are problematised as a source of undercounting and potential bias as they are difficult to capture with paper questionnaires. But even when subjects respond to questionnaires, their responses can be further problematised. While subjects are expected to reveal themselves truthfully, they are also understood, in some cases, to answer strategically and subversively. As we note later, subjects subvert questions by claiming unrecognised or unauthoritative categories of identification. This is just one problem of the self-eliciting subject whose answers are also influenced by how questions are worded or whether the questions are self-completed or involve an enumerator.[8] Such effects are usually attributed to the method or mode of asking a question and constitute a significant challenge but also problem of the subject.

Many efforts are thus directed at improving the reliability of responses, which often involve a tension between opening and closing the possibilities of how a subject can respond to a question:

> The value of open-ended questions is that they offer the respondent the right of total self-expression. The disadvantage is that the subsequent coding of responses and their allocation into a meaningful classification for output becomes more difficult and costly.[9]

Such problematisations of subjects were epitomised in the 2001 census of England and Wales when the Office for National Statistics (ONS) introduced a voluntary question on religion, which prompted much public debate,[10] including a widely circulated email that encouraged respondents to write in Jedi as their religion in protest.[11] More than 390,000 respondents wrote 'Jedi' in the free-text field of 'Any other religion' on the census questionnaire, making Jedi the fourth largest reported religion in the 2001 census, after Christian, Muslim, and Hindu. While the phenomenon may appear to be idiosyncratic, there are numerous examples of how subjects have historically challenged and demanded categories on census questionnaires to claim new identifications or to subvert the forces of subjectivation (Kertzer and Arel 2002).

The question was posed in the form of seven tick boxes and a free-text field. The tick boxes anticipated the most commonly expected categories (including 'No Religion') while the free-text field made

10 What is your religion?
♦ This question is voluntary.
♦ ✓ one box only.
- ☐ None
- ☐ Christian (including Church of England, Catholic, Protestant and all other Christian denominations)
- ☐ Buddhist
- ☐ Hindu
- ☐ Jewish
- ☐ Muslim
- ☐ Sikh
- ☐ Any other religion, *please write in*

Figure 1. Religion Question, 2011 Census Questionnaire for England and Wales.

it possible for subjects to write in their own answers (Figure 1). The format thus required subjects to submit to their identification by ticking boxes or filling in the free-text field. But it was through their submission to the question that it was then possible for 390,000 subjects to subversively repeat an unanticipated answer, that of Jedi Knight. While subjects likely choose this category for different reasons, subjects, as well as the category, became problematic and its repetition forced a response.

ONS reacted to the unanticipated answer by categorising it as 'No Religion' whereas other unanticipated answers were categorised as 'real' religions under the Other Religion category.[12] In response to a Freedom of Information request for the exact number of people who answered Jedi and their eventual categorisation, ONS pointed to the publicly available results, and added that the purpose of the classification of Jedi as 'No Religion' was to allow 'direct statistical comparability between the two censuses' (ONS 2013). However, while the responses of problematic subjects were technically resolved to serve the aim of comparability, their subversive acts were not so easily dismissed. Much media coverage of the result followed, and through repetition, the subversive act gained visibility. John Pullinger, the ONS Director of Reporting and Analysis at the time, in an effort to simultaneously acknowledge and resignify the controversy, issued a statement that noted the Jedi campaign had encouraged 'those in their late teens and twenties' to return valid census forms, leading to a statistically significant increase in response rates from a group routinely underrepresented in census returns. That is, a typically hard-to-reach population may have subverted the census but had to, in the first instance, submit to it. Through this resignification of the subversive act, the problem was thus transformed into a solution.

The Jedi campaign arose from an interplay between the legitimate categories of statistical authorities and those performed by subjects. It epitomised a recursive process of the ONS resignifying the campaign by turning subversive acts into ones of submission. It is plausible that the set-up of the paper questionnaire could be reconfigured by removing the free-text field to curtail this interplay and unrecognised categories, for example. This is not to say that it would render the method immune to subversion, but that it would direct, constrain and alter the possibilities of acting. Indeed, between censuses, ONS engages in myriad activities to calibrate its methods to address problematic subjects, maximise coverage and increase reliability. These activities include optimising the content of questions and possible answers, testing variations of question phrasing, conducting focus groups to explore reactions, as well as researching how advertising, outreach and promotion engage respondents.

The Jedi campaign was concomitant with a rethinking of questionnaires that went beyond adjusting its paper form. This was highlighted at the pre-appointment hearing for the office of Chair of the UK Statistics Authority in 2011 when a member of parliament, who was sceptical of the current methodology, brought up the Jedi responses as a source of 'distortion':

> Would you regard it as a failure if you do not get rid of the 10-yearly census and replace it with something likely to be more accurate and would not introduce, as the census does, its own distortions, when the forces of the Jedi knights and Pastafarianism decide to register some point by describing themselves as Jedi knights or Pastafarians? The whole thing is a nonsense. It should have stopped. It is a waste. It is there because of inertia, surely. You will stop it, won't you? (House of Commons 2012)

The statement echoed a growing debate more generally about paper questionnaires and their limitations including the slow process of feedback and adjustment spread over a 10-year inter-census period noted above. It is in this context of the problematised and not reliably submissive subject that digital technologies have been offered as solutions such as government administrative data registers, which are more commonly used in several European countries to conduct censuses (UNECE 2014). These registers hold detailed and historical data of transactions between citizens and the state on national insurance, taxation and social welfare, amongst others:

> Usually questionnaires about income are prone to socially desirable answers and suffer from imprecise answers caused by memory gaps of the respondents. Registers give full information on all incomes linked to a respondent, even extreme ones that are not desired to be reported or small and irregular income components that are easily forgotten. (Demunter and Reis 2014)

These issues have been central to the ONS 'Census Transformation Programme', which has been underway since 2015 to assess the 'government-stated ambition that "censuses after 2021 will be conducted using other sources of data"' (ONS 2017). One undertaking in that regard has been the Administrative Data Census Project, which has involved experimenting with various registers to produce key census outputs. Notwithstanding the assumption that subjects are more truthful in answering questions in relation to a government administrative register, in practice, registers – like questionnaires – are also used strategically. In register-based countries, NSIs note that the place of usual residence category in the population register is sometimes strategically used to gain access to services or to increase the representation of marginalised groups.

However, administrative registers are one amongst several other sources of data that are conceived as solutions to problematic subjects. In the next two sections, we exemplify two experiments with digital solutions specifically targeted to reduce respondent burden, capture hard-to-count populations and direct the acts of subjects: social media platforms and digital or online censuses. We show how they reconfigure methods in ways that not only anticipate but, as we argue, recalibrate agencements and forces of subjectivation. At the same time, we demonstrate that rather than resolving the existence of problematic subjects, the interactions and dynamics of human and technological actors that make up these methodological solutions are inventive of new forms of subjectivation and problem formulations.

Social media solutions: inferring subjects

There are many examples of experiments both within NSIs and academic research that engage with social media platforms such as Facebook profiles and Twitter posts to infer statistics on geography, language, and sometimes even gender and ethnicity (Mislove *et al.* 2011, Liu and Ruths 2013, Mocanu *et al.* 2013, Nguyen *et al.* 2013, Sloan *et al.* 2015). These method experiments, which involve digital technologies, big data and new analytics, diverge most significantly from paper questionnaires in that subjects do not self-identify. Rather than data from 'registers of talk' such as those of traditional methods, these experiments use data generated by platforms that are 'registers of action' (Marres 2017). Subjects' identifications are inferred from data traces of their actions and collected for other purposes and constitute a different form of subjectivation. For one, subjects can neither opt-out or subvert inferences, but, as we detail below, through various adjustments to how they interact with platforms, they can generate new problematisations. In this section, we explore this in relation to one experiment, an ONS pilot project that sought to use Twitter data to investigate how populations move within the country.

The ONS pilot project used aggregated data retrieved from publicly available Twitter profiles to identify patterns in when and where tweets were posted by different users. The assumption was that if tweets originated from different places at different times throughout the year, it would be possible to identify a pattern, and infer underlying reasons for why people moved from one place to another. As a method, this differs from subjects declaring their movement patterns thus avoiding the common problematisations of false reporting and underreporting, where respondents either provide a wrong address, or provide only one address when they occupy several. Moreover, it held the promise of increasing the timeliness of statistics, or in our words, that are closer to live data about how populations move between addresses throughout the year. In contrast to the population census, which assigns subjects to fixed addresses, the Twitter method constituted subjects as bodies in motion. For statisticians, it held the promise of providing more timely statistics that reflected lived experiences and which did not rely on elicited (and unreliable) responses from subjects. However, while potentially a solution to one problem, it was generative of new ones because of the acts of subjects in relation to the sociotechnical arrangements of the platform.

This was evident when statisticians initially investigated the free-text location field included in Twitter profiles. However, they quickly found it unreliable as subjects, echoing the problem of the religion question in paper questionnaires, used it in very different ways, sometimes leaving it

blank, or even writing in fictional places. The free-text field provided the potential for subjects to act in ways that subverted and was not compatible with the strict geographical definition of location necessary for the pilot project. As an alternative, the statisticians decided to concentrate solely on tweets that included GPS coordinates as these messages, also known as geolocated tweets, provided standardised data about the location the tweet was posted from, which made them much easier to analyse using existing statistical methods. Moreover, they were less prone to the kinds of uncertainty subjects introduced. In short, the complications that could arise from using GPS data for population data were easier to anticipate and handle for the statisticians. At the same time, GPS coordinates were linked to a much broader sociotechnical arrangement consisting of satellites, sensors, and mobile devices and generated a new set of unanticipated issues and different problematisations of subjects as we outline below.

Although each tweet in the chosen subset could be linked to a geographical location using GPS coordinates, the stream of tweets for each user still needed to be translated into significant locations, namely work and home. To perform this translation, the statisticians formulated an algorithmic method that organised the stream of tweets for each user into dominant clusters[13] (a grouping of nearby points into a single point). These clusters were compared to the borders of local authorities, and if any clusters appeared in different local authorities from month-to-month, they were considered to represent instances of internal migration of Twitter subjects.

Using this analysis, the statisticians detected a 'strong signal' coinciding with the cycle of the academic year, indicating that in local authorities with high proportions of students the volume of tweets would decrease in June and increase again in September and October. Based on this finding, they concluded that the data could be used as an indicator of student mobility, as this movement was not possible to detect using any existing data sources. It is worth noting that students are often problematised in population statistics as they move between universities and other residences within the academic year, and counts made at different times in the same geographic area can display high variations if the size of the student population is sufficiently large (Mitchell *et al.* 2002, Duke-Williams 2009).

This was a promising result for the statisticians as it appeared to solve an earlier problematisation of the mobility of a hard-to-reach student population. However, the statistician in charge of the project noticed that there was a significant decrease in the number of data points at a particular date in the one-year sample of Twitter posts. He later identified that this decrease occurred the day iOS 8 (an operating system used by Apple devices) had been released. The new operating system changed the default settings for location sharing, and as many devices stopped reporting their locations, they disappeared from the dataset. Following the disappearance of data points, he described the dataset as volatile, that is, unreliable and prone to sudden changes, and therefore unsuitable as a source of data for official statistics. The problematisation of the method as volatile meant that migrating subjects inferred using Twitter posts could be the result of a change in default settings for a popular mobile operating system. The reconfiguration of subjectivating forces, while solving one problem was thus generative of new ones. Using GPS coordinates to overcome the challenges of determining location based on a free-text field led to the problematisation of the volatility of the technology beyond their control such that the cause of a sudden change might be impossible to detect or understand.

The Twitter pilot began as a method of the live tracking of movement by inferring categories from clusters of data points made possible by a highly technical analysis. For us, the pilot demonstrated how subjectivation is configured differently by different methods, but also that its force is the product of the interactions and dynamics between human and technological actors, including categories, software, algorithms, and digital devices. While the identification method chosen by statisticians enacted students as mobile bodies, it arose from the complex interplay between location categories such as home and work, software settings, release schedules, and study design as well as the actions and inactions of subjects.

The Jedi campaign involved subjects actively and intentionally subverting anticipated categories, whereas migrating students were inferred from a series of technical and digital actions that generated variable and unstable geographic metadata. In both cases, however, student subjects were problematised through their use of a free-text field, a method that generated unanticipated categories or interpretations. As we saw in the case of the Jedi category, responses were resignified to support different arguments: one in the service of increasing the response rate and the other to support alternative methods and data sources for censuses. Regarding the latter, and like that which we will discuss below, the reconfiguration of subjectivation forces, while solving one set of problems also generated new ones. In the Twitter pilot, statisticians used GPS coordinates to overcome the challenges of the location free-text field but at the same time became vulnerable to technical forces of operating systems involving actions of others beyond their control or knowledge. In the next section, we explore this issue of control through another method experiment with digital technologies that involved statisticians designing their own digital platforms to conduct digital or online censuses.

Digital census solutions: calibrating subjects

Another project of the ONS Census Transformation Programme involves experimenting with an online census for the 2021 round of enumeration.[14] The method builds on developments in other countries and especially the Australian Statistical Bureau's (ABS) introduction of a 'digital census' in 2016, which was conceived as a 'transformation' rather than simply 'translation' of a paper questionnaire into digital format. It involved a move from digital publishing to digital transacting and interacting with subjects at all stages of enumeration, and a 'responsive' approach that made process adjustments in near 'real-time' as a result of field intelligence and response rates (Australian Bureau of Statistics 2015). A central management centre was set-up to achieve this by digitally monitoring a range of information, including online response rates, paper form requests and returns, and social media. For example when the response rate of an area lagged that of others, then a variation to the enumeration approach was designed, reviewed and actioned such as a field visit from an enumerator.

The relation to subjects was described as not just changed by merely putting a questionnaire online but transformed into an interaction that was 'easy, responsive, fun' through a design that provided more information through pop-up windows that guided correct responses; drag and drop techniques that facilitated the ease of completing questions; assistance prompts that guided experience such as supplementary questions; and images and summary compilations that visualised responses so that they could be verified by subjects. The digital census also included incentives intended to modify behaviour such as being able to donate data to an archive for future generations. Digital technologies were further deployed by a 'field force' of workers to better capture and monitor subjects. By digitally monitoring progress through handheld devices, constant updates on operational progress and instructions were fed back to workers to optimise their activity and highlight problem areas in response rates.[15] Social media platforms such as Twitter were also used by workers to communicate experiences and problems to each other so that problematic subjects and areas could be better targeted.

Through these and other digital technologies, subjectivation was transformed into an interactive and live process of calibrating and maximising the responses of subjects by prompting and guiding them and making the process fun and easy rather than burdensome. Subjects who did not submit or obey in ways anticipated, were then targeted either by digital techniques such as prompts or by enumerators deployed through offline modes in the field. Significantly, the digital techniques were deployed in real-time, in contrast to paper questionnaires, which, as we noted above, typically go through long processes of testing, piloting and field worker feedback. With digital censuses then, relations between digital technologies, central management, and the field workers that make up the force of the method anticipate, guide and subjectify respondents through a dynamic process.

The ONS experiment with an online census is also proposed to go beyond a simple translation of a paper questionnaire and incorporates many of the same elements such as contextual assistance for subjects to complete questions, detailed drop-down boxes to reduce coding, comprehensive validation within and between questions, and the design of questions fit smaller screens so that subjects can respond using handheld devices (ONS 2015). These elements all make use of paradata, a type of metadata that is not descriptive as in traditional metadata but process data on a subject'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. It includes data on which devices are being used; which buttons (help, back, forward) are being clicked and when; changes subjects make to answers, and so on (Statistics Austria 2015). For each, inferences can be made about myriad issues such as individual subjects and groups who do not submit to the census in ways anticipated and desired as a consequence of one of these design elements. Paradata thus involves the tracking of the relation between the census and the subject through metrics about data collection[16] and are part of a 'data driven approach,' which informs strategies for increasing the numbers of subjects who submit to its requirements. As a by-product of digital technologies, it is put in the service of better calibrating responses and the forces of subjectivation.

Additionally, through the use of smart technologies such as autocomplete, digital censuses operate like other digital platforms. Indeed, one justification for digital censuses is that subjects regularly engage with public and commercial platforms and thus are familiar with and have the skills necessary for their submission. At the same time, digital censuses adopt many of the elements that make up these other platforms – especially those of Google, Facebook and Amazon – which channel choices and direct queries (König and Rasch 2014). While query functions appear neutral, autocomplete techniques anticipate and predict what subjects may want to know and direct their queries through suggestions. Like smart type-in, logical controls on entries and assistance prompts, autocomplete is intended to make searching faster and easier and produce optimal results. These are some of the logics and practices innovated and designed by technology companies that digital censuses adopt and adapt to guide responses in desired directions.

However, these solutions to the problematisation of subjects have generated new problematic ones. The ONS, for example identified four groups of subjects in relation to their anticipated access to and/or willingness to use the internet to digitally engage with government (Figure 2). Problematic subjects were differentiated according to several criteria. For each group, their related sociodemographic characteristics were identified (age, location, etc.) as well as reasons for being unwilling to digitally engage (lack of trust, internet security, etc.). In this conception, a digital divide was not simply between who does or does not have access to the internet, but occurs along various combinations of identification such as where someone lives and their age. These characteristics were used to calculate the numbers of likely hard-to-count subjects and their relative concentration in different geographic areas. Response rates and patterns could then be tracked in these areas and direct follow-up field activities organised to increase the number of subjects who submit to the census when targets were not being met.

Subjectivation thus involves myriad deployments of digital technologies, from the online and digital census form, paradata and smart techniques to the differentiation of subjects and their variable targeting. Relations to subjects thus involve anticipating how they might act and then calibrating how they do act through ongoing process of digital management and directing of their submission. That is, an online or digital census does not simply involve deploying digital technologies, but managing their operation and the performance of subjects in relation to them as live processes of subjectivation. Such management as we noted, also involves offline modes as was demonstrated in the ONS's test of its online census in 2017, which was designed to evaluate options for maximising responses, self-completion, and the quality of responses. One element evaluated was the introduction of an 'Assisted Digital Service' to reach the 'more than 10% of UK adults who have never used the internet' and recognition that '21% of the population lack basic online skills' (Bexley 2017). The service involved setting up computer terminals in a local library with librarians to assist subjects in

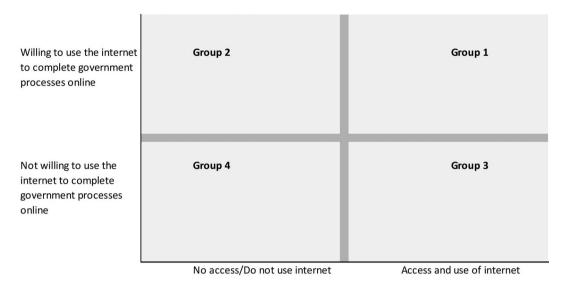Matrix with the four groups of respondents to consider in an online census

	No access/Do not use internet	Access and use of internet
Willing to use the internet to complete government processes online	**Group 2**	**Group 1**
Not willing to use the internet to complete government processes online	**Group 4**	**Group 3**

Figure 2. Categorisation of Respondents (ONS 2015).

completing an online questionnaire. For our argument, what the experiment demonstrated is how a digital census is generative of new problematic subjects and calls forth management solutions in the form of new actors (librarians) and sites (library computer terminals), which participate in subjectivation. But management is not only necessary to direct subjects and their submission, but also the instability and vulnerabilities of digital technologies. While this can take many forms such as a change in an operating system as we previously noted, a dramatic example was the disruption to the Australian digital census website, which suffered a mass outage and was shut down for 43 hours during the 2016 enumeration (MacGibbon 2016). Attributed to a Distributed Denial of Service Attack (DDoS), the failure lead to a major inquiry into cybersecurity. Loss of public trust and confidence were widely noted as a major consequence but for us the incident points to the vulnerabilities of digital technologies not only to operational failures but to other forms of subversion and how the introduction of new technological and human actors reconfigure those possibilities.

Conclusions

We have conceived of government methods as configured by and inventive of problematisations and how experiments with digital technologies introduce 'live' relations to subjects. While work on sociological and digital methods highlights how these features enable new modes of participation, we have attended to how they reconfigure what we have called the forces of subjectivation. Understanding subjectivation in respect to both sociological and governmental methods is ever more important as digital technologies are introducing new method agencements of human and technological actors.

Our analysis suggests that these forces are not linear and never settled but instead recursive and dynamically composed. For instance, while digital technologies are offered as solutions to problematic and subversive subjects, they also reconfigure forces and give rise to new problematisations. Automation, digital tracking and recursive designs may calibrate processes of subjectivation, but they are also inventive of new possibilities for subjects to act, be excluded or problematised. They also introduce new actors such as the assumptions, different objectives and biases of platforms and the decisions of operating system owners. However, rather than simply a question of reducing the potential of subjects to act, we have attended to how methods differently configure their

subjectivation, which can be anticipated and guided but not settled in advance. However, methods not only configure the capacities of subjects to obey, submit and subvert, they also configure their object, that is, the populations that are enacted. While populations have historically been understood as relatively stable objects that only require periodic measurement, experiments enact them as fluid and modulating (Ruppert 2012). In other words, new kinds of populations and modes of intervention are also invented. Furthermore, while typically based on self-elicited social categories, these experiments infer identification categories and populations from the data traces of subjects generated by their actions in relation to digital platforms. In these ways, not only do methods produce their subjects and their agential capacities but also the very object of population is transformed.

These conclusions are critical as digital technologies become ever more part of social life and at the same time part of method agencements for knowing and governing populations. How the dynamics of subjectivation will play out are by no means given or settled, but through our examples, we can suggest what they may mean for relations between subjects and the making of government population statistics. The critical import for our argument is that digital technologies often work in the 'background': from the technical configurations of digital censuses to the scraping of tweets to infer categories, what then are the possibilities of subversion, intervention or accountability? Subversion does not only mean to attack or undermine authority, but to make demands and claims about its operation. Given the long history of how NSIs have sought to secure the consent of subjects for both the collection and use of data about them, we suggest that providing possibilities for such interventions are important to secure the future legitimacy of government statistics. That is, the possibilities and potentials of citizens to act in their subjectivation and identification are as important, if not more, than the promises of digital technologies for more timely, efficient, cheaper and reliable statistics.

Notes

1. While there is some debate about the similarities and differences between forms of data generated through the internet and government administrative databases, for some statisticians, they both constitute what is now more commonly now known as big data. We follow this convention and when relevant specify a specific source.
2. Beyond these there are also many other communities of practice through which people are identified. Following Bowker and Star (1999), we recognise that government practices are connected to and require other communities of practice. This is significant because identification is not in isolation but in relation to numerous other practices.
3. We draw on ethnographic data collected by a team of researchers as part of a collaborative research project called ARITHMUS, funded by the European Research Council (Peopling Europe: How data make a people; www.arithmus.eu). It involved six researchers: Evelyn Ruppert (PI), Baki Cakici, Francisca Grommé, Stephan Scheel, Ville Takala, and Funda Ustek-Spilda.
4. This formulation is elaborated in relation to digital acts in Isin and Ruppert (2015).
5. Economic Commission for Europe, Conference of European Statisticians, Group of Experts on Population and Housing Censuses, Fifteenth Meeting, Geneva, 30 September–3 October 2013. Field notes.
6. Respondent burden also applies to businesses but our focus is on the individual subject. While cost reduction is often cited as a key driver, also relevant is the belief that innovating data, methods and technologies is a necessity otherwise NSIs will fall behind developments across government and in the private sector. Interview with Eurostat, 15.03.2016.
7. The conception of burden draws on presentations and discussions at international meetings that we have followed from 2013 to 2017.
8. Economic Commission for Europe, Conference of European Statisticians. Group of Experts on Population and Housing Censuses. Seventeenth Meeting. Geneva, 30 September to 2 October 2015. Field notes. An example provided at this meeting concerned stark differences in measurements of rates of disability depending on whether the question is self-completed on a paper questionnaire or asked in a face-to-face interview; differences were explained as a matter of trust.
9. Economic Commission for Europe, Conference of European Statisticians, Group of Experts on Population and Housing Censuses, Fifteenth Meeting, Geneva, 30 September–3 October 2013. Field notes.
10. Critics argued that religion is a private and sensitive matter that is not suitable for the census, while some faith groups campaigned for the inclusion of the religion question to increase the visibility of religious communities (Aspinall 2000, Southworth 2005, Sherif 2011).

11. The email, later revealed to be a hoax, claimed that Jedi would become an officially recognised religion if 10,000 people declared it as their religion in the census (BBC 2001).
12. In the 2011 census, the entries under the Other Religion category ranged from more recognisable entries such as Pagan (56,620 people), or Jain (20,288 people) to those not immediately recognised as religions, such as those who answered, 'Belief in God' (2969 people) or 'Own Belief System' (1949 people). Those who answered Jedi Knight (176,632 people) were counted under the No Religion category, together with those who answered Atheist (29,267), and Heavy Metal (6242 people), among others.
13. Using a clustering algorithm called DBScan, the tweets were assigned to places, under the assumption that tweets with coordinates close to one another were likely to have originated from the same location. Next, a set of rules involving the time of day and frequency of posts were used to infer whether these locations could be considered the home or the workplace of the poster (See ONS 2012).
14. Alongside the development of the online census, another strand of work will involve research into the use of administrative data for enhancing 2021 outputs and the development of methods to produce more detailed inter-censal annual statistics. Based on this work and the results of the 2021 census, ONS will then assess the potential use of administrative data and new surveys to replace that which is provided by a primarily online census (ONS N.D.)
15. Australian Bureau of Statistics (2015, p. 6).
16. Paradata is identified as a standard of statistical modernisation 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. Field notes.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Baki Cakici is Assistant Professor, Technologies in Practice, IT University Copenhagen. His research draws on theories from the field of science and technology studies to address issues of surveillance, classification, big data, algorithms, politics of numbers, and the history of computing. He was a postdoctoral researcher on ARITHMUS at Goldsmiths, University of London.

Evelyn Ruppert is Professor of Sociology at Goldsmiths, University of London. She studies how digital technologies and the data they generate can powerfully shape and have consequences for how people are known and governed and how they understand themselves as political subjects. She is PI of ARITHMUS, funded by an ERC Consolidator Grant (2014–2019).

ORCID

Baki Cakici http://orcid.org/0000-0002-1039-8177
Evelyn Ruppert http://orcid.org/0000-0001-7781-1491

References

Adams, V., Murphy, M., and Clarke, A. E., 2009. Anticipation: technoscience, life, affect, temporality. *Subjectivity*, 28 (1), 246–265.

Aspinall, P., 2000. Should a question on 'religion' be asked in the 2001 British Census? A public policy case in favour. *Social Policy & Administration*, 34 (5), 584–600. doi:10.1111/1467-9515.00212.

Australian Bureau of Statistics, 2015. *Big crocs, big snakes and small censuses—the story of Australia's digital-first census*. Paper presented to the economic commission for Europe, Conference of European Statisticians. Group of experts on population and housing censuses, seventeenth meeting. Geneva, Switzerland: UNECE.

Back, L. and Puwar, N., 2012. A manifesto for live methods: Provocations and capacities. *In*: L. Back and N. Puwar, eds. *Live methods*. London: Wiley-Blackwell/Sociological Review Monographs, 6–17.

Balibar, E., 1991. Citizen subject. *In*: E. Cadava, P. Connor and J.-L. Nancy, eds. *Who comes after the subject?* London: Routledge, 33–57.

BBC, 2001. *Jedi e-mail revealed as hoax*, 11 April. Available from: http://news.bbc.co.uk/1/hi/entertainment/1271380.stm.

Bexley, S., 2017. It's all about inclusion: how ONS plans to support the digital have-nots. Accessed 23 Mar 2018 from National Statistical: News and insights from the Office for National Statistics website. Available from: https://blog.ons.gov.uk/2017/04/04/its-all-about-inclusion-how-ons-plans-to-support-the-digital-have-nots/.

Bowker, G. C., and Star, S. L., 1999. *Sorting things out: classification and its consequences*. Cambridge, Massachusetts: The MIT Press.

Bruns, A. and Burgess, J., 2015. Easy data, hard data: the politics and pragmatics of Twitter research after the computational turn. *In*: G. Langlois, J. Redden and G. Elmer, eds. *Compromised data: from social media to big data*. New York: Bloomsbury Academic, 93–111.

Butler, J., 1997. *The psychic life of power: theories in subjection*. Stanford, CA: Stanford University Press.

Cremonesi, L., *et al.*, eds., 2016. *Foucault and the making of subjects; rethinking autonomy between subjection and subjectivation*. London: Rowman & Littlefield International.

Demunter, C., and Reis, F., 2014. Using mobile positioning data for official statistics: daydream nation or promised land? Presented at the Eurostat New Technique and Technologies for Statistics conference, Brussels.

Duke-Williams, O., 2009. The Geographies of student migration in the UK. *Environment and Planning A: Economy and Space*, 41 (8), 1826–1848. doi:10.1068/a4198.

Eurostat, 2011. *European statistics code of practice*. Available from Eurostat website: http://ec.europa.eu/eurostat/documents/3859598/5921861/KS-32-11-955-EN.PDF/5fa1ebc6-90bb-43fa-888f-dde032471e15.

Foucault, M., 1982. The subject and power. *Critical Inquiry*, 8 (4), 777–795.

Foucault, M., 1997. The birth of biopolitics. *In*: P. Rabinow, ed. *Ethics: subjectivity and truth*: *Vol. essential works 1*. London: Penguin, 73–79.

House of Commons, 2012. *Appointment of the Chair of the UK Statistics Authority: sixteenth report of session 2010–12 —Volume II: oral and written evidence* (No. HC 910-II). London, UK: House of Commons – Public Administration Select Committee.

Isin, E., 2012. *Citizens without frontiers*. London: Bloomsbury.

Isin, E., and Ruppert, E., 2015. *Being digital citizens*. London: Rowman & Littlefield International.

Kertzer, D. I., and Arel, D., 2002. *Census and identity: the politics of race, ethnicity, and language in national censuses*. Cambridge: Cambridge University Press.

König, R., and Rasch, M., 2014. *Society of the query reader: reflections on web search*. Amsterdam: Institute of Network Cultures.

Langlois, G., Redden, J., and Elmer, G., eds., 2015. Introduction. *In*: *Compromised data: from social media to big data*. New York: Bloomsbury Academic, 1–14.

Law, J., and Ruppert, E., 2013. The social life of methods: devices. *Journal of Cultural Economy*, 6 (3), 229–240. doi:10.1080/17530350.2013.812042.

Liu, W., and Ruths, D., 2013. What's in a name? Using first names as features for gender inference in Twitter. *AAAI Spring Symposium: Analyzing Microtext*, 13, 10–16.

Lury, C., and Wakeford, N., 2012. *Inventive methods: the happening of the social*. London: Routledge.

MacGibbon, A., 2016. *Review of the events surrounding the 2016 eCensus*: Australian Government, Department of the Prime Minister and Cabinet.

Mackenzie, A., 2013. Programming subjects in the regime of anticipation: software studies and subjectivity. *Subjectivity*, 6 (4), 391–405. doi:10.1057/sub.2013.12.

Marres, N., 2017. *Digital sociology: the reinvention of social research*. Cambridge, UK: Polity Press.

Marres, N., and Gerlitz, C., 2016. Interface methods: renegotiating relations between digital social research, STS and sociology. *The Sociological Review*, 64 (1), 21–46. doi:10.1111/1467-954X.12314.

Mislove, A., *et al.*, 2011. Understanding the demographics of Twitter users. *Fifth International AAAI Conference on Weblogs and Social Media*.

Mitchell, R., *et al.*, 2002. Bringing the missing million home: correcting the 1991 small area statistics for undercount. *Environment and Planning A: Economy and Space*, 34 (6), 1021–1035. doi:10.1068/a34161.

Mocanu, D., *et al.*, 2013. The Twitter of babel: mapping world languages through microblogging platforms. *PLOS ONE*, 8 (4). doi:10.1371/journal.pone.0061981.

Neyland, D., and Milyaeva, S., 2016. The entangling of problems, solutions and markets: on building a market for privacy. *Science as Culture*, 25 (3), 305–326.

Nguyen, D.-P., *et al.*, 2013. 'How old do you think I am?' A study of language and age in Twitter. *Seventh International AAAI Conference on Weblogs and Social Media*.

ONS, N.D. Census transformation programme. Available from: https://www.ons.gov.uk/census/censustransformationprogramme [Accessed 2 Jan 2018].

ONS, 2012. 2011 Census: Key Statistics for local authorities in England and Wales. Available from: http://bit.ly/2ncI70x [Accessed 28 Mar 2018].

ONS, 2013. FOI Request: Religion classification—Jedi. Office for National Statistics. Available from: http://bit.ly/2nvN1t5 [Accessed 23 Mar 2018].

ONS, 2015. *Research for 2021 Census England and Wales: possible innovations under consideration*. Presented at the Economic Commission for Europe, Conference of European Statisticians. Group of experts on population and housing censuses. Seventeenth meeting. Geneva, 30 September to 2 October.

ONS, 2017. Research Outputs: Estimating the size of the population in England and Wales, 2017 release. Available from: https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/sizeofthepopulation/researchoutputsestimatingthesizeofthepopulationinenglandandwales2017release [Accessed 23 Mar 2018].

Osborne, T., Rose, N., and Savage, M., 2008. Editors' introduction—reinscribing British sociology: some critical reflections. *Sociological Review*, 56 (4), 519–534.

Rose, N., 1996. *Inventing our selves: psychology, power and personhood*. Cambridge: Cambridge University Press.

Rose, N., 1999. *Powers of freedom: reframing political thought*. Cambridge: Cambridge University Press.

Rossi, A., 2016. *The labour of subjectivity: Foucault on biopolitics, economy, critique*. Lanham: Rowman & Littlefield International.

Ruppert, E., 2011. Population objects: interpassive subjects. *Sociology*, 45 (2), 218–233.

Ruppert, E., 2012. The governmental topologies of database devices. *Theory, Culture & Society*, 29 (4–5), 1–21.

Sherif, J., 2011. A census chronicle – reflections on the campaign for a religion question in the 2001 census for England and Wales. *Journal of Beliefs & Values*, 32 (1), 1–18. doi:10.1080/13617672.2011.549306.

Sloan, L., *et al.*, 2015. Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. *PLOS ONE*, 10 (3). doi:10.1371/journal.pone.0115545.

Southworth, J. R., 2005. 'Religion' in the 2001 Census for England and Wales. *Population, Space and Place*, 11 (2), 75–88. doi:10.1002/psp.361.

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.

Torpey, J., 2000. *The invention of the passport: surveillance, citizenship and the state*. Cambridge: Cambridge University Press.

UNECE, 2014. *Measuring population and housing: practices of the UNECE countries in the 2010 round of censuses* (No. ECE/CES/34; p. 220). Geneva, Switzerland: United Nations Economic Commission for Europe.

United Nations, 2014. *Fundamental principles of official statistics* (No. A /RES/68/261). Available from United Nations General Assembly website: http://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf.