Data Scientists: A new faction of the transnational field of statistics

Francisca Grommé, Evelyn Ruppert and Baki Cakici, Goldsmiths University of London

Version date: 30 May 2017

For *Ethnography for a Data Saturated World* (Knox and Nafus eds., Manchester University Press, Forthcoming 2017).

Introduction

National statistical institutes (NSIs) have historically sought to be a single point of truth about states. Their publications, as recent press releases point out, communicate objective statistics, such as "In October 2016, the French business climate is stable" (INSEE 2016), and "Finland's population in September for the first time closer to six than five million" (Saari 2016). Such statements may come across as 'dry' and 'factual'. Yet, they are the result of the painstaking work of translating questionnaires, surveys and registers into statistics that seek trust and legitimacy from governments, corporations and publics. Such trust is in part generated by the application of international quality standards for accuracy, reliability, and timeliness that national statisticians appeal to in carrying out their role.

The availability of new data sources – or Big Data - such as that from mobile phone locations and sensors is introducing new possibilities for the generation of official statistics. For instance, rather than the population register, counting Finland's population could be based on the number of residents who own a mobile phone. However, existing methods for official statistics are not suited to the analysis and interpretation of these data. Data science and data scientists are increasingly identified as the discipline and profession necessary for realising the potential of these new data sources, which require skills and knowledge of analytic techniques not typically used in official statistics such as machine learning, algorithms and predictive modelling. In this regard, Big Data are not only introducing new forms of data to the repertoire of official statistics but also new valuations of professional skills and expertise, or, as we conceptualise in this chapter, of cultural and symbolic capital (Bourdieu 1984).

In the making of official statistics, what exactly comes to count as data science and the profession of data scientist is not self-evident. While there is much talk about a new science and profession,[1] a science of data dates back to Peter Naur's "Concise Survey of Computer

[1] A quick look on Google Trends shows that searches for the term "data science" started increasing in frequency around 2012, and is still climbing.

Methods" (Naur 1974), which is often cited as the source of the term "data science". Naur defines data science as "the science of dealing with data once they have been established, while the relation of data to what they represent is delegated to other fields and sciences" (Naur 1974, p.30). Various genealogies from computer science, statistics, economics, and corporate sources point to other well-known names in statistics alongside Naur, including Gottfried Wilhelm Leibniz, Thomas Bayes, Florence Nightingale, and John Tukey, among others (K. Cukier 2010; Loukides 2010; Patil and Davenport 2012; Donoho 2015). Contemporary definitions of data science and data scientists are much more closely associated with Big Data, another recent term, and the definition of data scientists as experts who work with Big Data is cited in recent literature examining data science practices (Kitchin 2014; Ruppert, Law, and Savage 2013; Burrows and Savage 2014; Gehl 2015; Halavais 2015; Pasquale 2015).[2]

Our concern in this chapter is how data scientists are being defined in relation to national statisticians and how both professions are objects of struggle with ambiguous outcomes. While the concentration of the generation, accumulation and ownership of Big Data in the hands of private sector corporations has been well documented and argued, our interest is in professional struggles over the legitimate authority to generate official knowledge of the state. We refer to this struggle as the politics of method in which the objects of study and the professions that study them are being defined at the same time (Savage 2010). This entwinement of objects and professions means that this struggle is consequential for how national statistical institutes (NSIs) define and bring into being official statistics as matters of government (Desrosières 1998; Ruppert 2012). However, how national statisticians position themselves in relation to data scientists is not only the outcome of discussion and debate, but also of specific material-semiotic practices such as experiments, demonstrations and job descriptions. Furthermore, this struggle is not delineated by national interests and practices (or to any other level or scale), but part of transnational negotiations, contestations and tensions that cut across numerous NSIs and international statistical organisations.

These are the main arguments of our analysis of fieldwork that we draw on in this chapter, which involved a collaborative ethnography of several NSIs and international statistical organisations in Europe.[3] By observing meetings, conducting interviews, shadowing

[2] Some, often academic, statisticians have publicly spoken out against a division between data science and statistics, arguing that statistics are at the core of data science and that the volume of data does not change that fact (cf. Meulman 2016).

[3] The ethnographic fieldwork, which began in 2013, was part of an ERC funded project, Peopling Europe: How data make a people (ARITHMUS; 2014-19; CoG 615588) and involved a team of researchers: Evelyn Ruppert (PI), Baki Cakici, Francisca Grommé, Stephan Scheel, Ville Takala, and Funda Ustek-Spilda. This chapter has

statisticians, observing experiments and performing participatory exercises, we attended to how practices traverse, travel between, and connect sites and scales.[4] This is what we refer to as a 'transversal' ethnographic method (see Scheel et al. 2016), which involves working across national and disciplinary boundaries, spatial scales, individual(ised) projects, and standardised or predefined research techniques. In practice, it required the work of several researchers to follow, engage and establish relations with different field sites (and each other) and connect or contrast observed practices across sites. The method allowed us to analyse the implementation and use of new data sources and analytics in official statistics as the outcome of innumerable distributed practices, negotiations, struggles, tensions and constraints that traverse national statistical institutes and international statistical organisations. By following the details of these rather mundane and situated bureaucratic practices we identified how data science and the profession of data scientist are being constituted in relation to national statisticians. As we note later in the chapter, this did not mean the practices we observed were the same or equally distributed or that there was consensus and settlement on definitions across actors and sites. Rather, following situated practices enabled us to trace repetitions of the activities, stakes, rationalities, claims and arguments shaping the formation of these professions as well as their objects.

We first establish a conceptual understanding of how valuations of professional skills and expertise in relation to Big Data constitute cultural and symbolic capital and are changing the relations between factions of professions within the transnational field of statistics. We then take this up to discuss a NSI recruitment practice where job interviews enact a version of data science and the profession of data scientist anchored around declarations of accumulated skill and potential. Next, we show how being a data scientist in relation to official statistics is not only a matter of possessing specific skills, but also of acquiring certain sensibilities. We discuss how a habitus for data science is shaped through the material-semiotic practice of the data camp. Finally, we outline how the object of data science, Big Data, is consolidated through repetitions of definitions and appeals to 'learning by doing' at international conferences and meetings to distinguish the data scientist from the iStatistician and defend the role of NSIs in the legitimate production of official statistics.

---

benefited from the insights of all team members, and is the result of ongoing collective work, conversations and analysis.

[4] The field sites include national statistical institutes and international statistical organisations: the Office for National Statistics of England and Wales; Statistics Netherlands; Statistics Estonia; Statistics Finland; Turkstat; Eurostat; and the Statistical Division of the United Nations Economic Commission for Europe.

Data scientists: Field, Capital, Habitus

Pierre Bourdieu's (1984) conceptualisation of fields, capital and power provides us with a way to understand the formation of the figure of the data scientist in relation to the making of a professional field.[5] Bourdieu understands a field as a dynamic configuration of relational positions occupied by actors who compete with one another over recognition of the forms of capital that shape their relative positions and in turn power and authority. Within any given or emerging field actors seek to maintain or improve their positions in relation to each other through the valuation of different forms of capital, including cultural, economic, social, and symbolic capital. It is then through the accumulation of these various forms of capital that their relative positions are established within the field (Bourdieu and Wacquant 2006: 127-8). While Bourdieu's studies were mostly confined to the nation and in particular to France, others have taken up this conceptualisation to understand inter- and transnational fields. Most notably, are studies in the fields of international law (Madsen 2014; Madsen 2011; Dezalay and Garth 1996) and international political sociology (Bigo 2011). In Didier Bigo's understanding of a transnational field, the 'national' is not simply replaced by the 'transnational' or the 'global'. Rather, he advances that the transnational exists in the form of transnational networks and practices of professionals who 'play simultaneously in domestic and transnational fields' (Bigo 2011: 251). In this view, a transnational field is constituted by networks and practices between and amongst professionals who act at various non-hierarchically ordered scales of the transnational, national and local. Bourdieu's conceptualisation insists on attending to struggle and change, fragile moments, and the emergence of new kinds of practices (Bigo 2011, 240–41).

It is through such simultaneously national and transnational networks and practices that statisticians have operated and worked since the late nineteenth century but especially during the post WWII period to make up a transnational field of statistics. Through working with and in relation to professional organisations such as the International Statistical Institute (ISI) and the International Association for Official Statistics (IAOS),[6] and international governing organisations such as the United Nations Statistical Division (UNSD) and the European Statistical Service (ESS), they have come to constitute one faction of actors who have forged a transnational field of statistical practice. Like other fields, they form one faction of a broader

[5] The following draws on a working paper that we wrote with our colleagues. See: (Scheel et al. 2016).
[6] The ISI was founded in 1885 but international meetings began in 1853. It has had consultative status with the Economic and Social Council of the United Nations since 1949. The IAOS has existed since 1985 and it is a specialized section of the ISI consisting of producers and users of official statistics.

field that includes statisticians who work in different capacities within government but also in the academy, commercial and non-governmental sectors. And like other fields it is dynamic and has undergone specific transformations as a result of changing methods, technologies and governing strategies, problematisations and rationalities but also as a result of struggles within and between factions.

Rather than mapping the dynamics of this transnational field and the dynamics of who constitute its stakeholders, factions and their relative positions, we focus on changing valuations of methods, technologies, expertise, skills, education and experience in the making of data scientists, especially in relation to national statisticians and in turn the recognition of different forms of cultural capital within the field. It is through structured social practices and structuring fields that various agents and their interests generate forms of expertise, interpretations, concepts, and methods that collectively function as fields of power and knowledge. This is one of the lessons of Bourdieu's (1984) studies on the ways in which fields of knowledge constitute fields of power. As Mike Savage (2010) has documented, these struggles constitute a *politics of method* through which statisticians and other stakeholders (demographers, data scientists, domain specialists etc.) both within and outside NSIs struggle over the technologies, truth claims, budgets and methods involved in the production of official statistics in order to improve their relative position. Rather than possessing and having fixed advantages, resources and skills are 'mobilized to achieve advantage and classify social distinctions' within this particular context and field (Halford and Savage 2010: 944). It is through such mobilisations that new skills and expertise and the rise of new professions and positions such as data scientists, data wranglers and data infomediaries get valued and recognised. Another site of these struggles is that of university curricula and professional training programmes in data science, for example.

However, struggles also occur through myriad material-semiotic practices that demonstrate and challenge competing methods and truth claims. The stakes are thus not only relative power and capital but also what Bourdieu refers to as the exercise of symbolic violence over the production, consecration and institutionalisation of forms of knowledge:

> Symbolic power is the power to make things with words. It is only if it is true, that is, adequate to things, that description makes things. In this sense, symbolic power is a power of consecration or revelation, the power to consecrate or to reveal things that are already there (Bourdieu 1988:28).

While Bourdieu does not express this as performative, he asserts that a description 'makes things'; from populations to the economy and the making of professions, the outcome of struggles thus also involves the constitution of the object itself.

While many objects of struggle constitute the field, a prominent and current one in relation to official statistics is the framing of the threat (or not) of a new faction (and its definition) called the 'data scientist'. The faction emerged in relation to the technological expertise required to analyse Big Data, a term that became pervasive around 2011. Its proponents promised a new science of societies that challenged existing forms of data and knowledge such as that generated by traditional methods and practices of national statisticians. While many commentators called it a hype, the prevalence and claims about data science and data scientists became an object of struggle within the field:

> On the other hand, the Big Data industry is rising: the huge volume of digital information derived from all types of human activities is being increasingly exploited to produce statistical figures. These figures often make use of data from private institutions or companies. Leaving aside the current public debate on whether companies which collect the data should own the data and could use them for another purpose without consent, these new statistical figures may be seen as competitors of traditional official statistics.[7]

Reflecting on this struggle some two years later, the Director General of Eurostat argued that 'we are at the edge of a new era for statistics' as 'data is raining down on us' and, as he further put it, others are claiming that the data revolution could make national statisticians obsolete.[8] With a chief data scientist now located in the White House what then is their relation to the chief statistician, he asked?

> Ignoring these new developments, official statistics would lose relevance in future and risks to be marginalised similarly to what happened to the geographical offices with Google or TomTom heavily investing into satellite images, aerial photographs and topographic maps.[9]

It is in response to this question that national statisticians over the past several years have struggled to define the profession of data scientist as well as their position in relation to

---

[7] Eurostat. 2014. 'Big data – an opportunity or a threat to official statistics?' Presentation to the Economic Commission for Europe Conference of European Statisticians. Sixty-second plenary session. Paris, 9-11 April 2014.

[8] Fieldwork Notes. From the opening address of Walter Radermacher, the Director General of Eurostat at the 'New Techniques and Technologies for Statistics (NTTS)' 2015 conference in Brussels, BE, an international biennial scientific gathering organised by Eurostat.

[9] Eurostat 2015; internal document.

it. But in doing so they have been both reconstituting their profession as well as their object, official statistics. At the same time, as Bourdieu also advances, it is because of their similar socialisation, career trajectories, interests and practices that they have a shared habitus and in turn can develop a common sense of the stakes. There is a correspondence or homology between social positions of agents and their dispositions, and their ways of perceiving and evaluating. Dispositions are an embodied state of capital, which Bourdieu names habitus, a system of lasting dispositions that integrate past experiences. Acquired dispositions are part of one's habitus, are internalized, embodied social structures, and a 'primary form of classification that functions below the level of consciousness and language and beyond the scrutiny or control of the will' (Bourdieu 1984: 466). Rather than there being a mechanical or direct relation (e.g. reflection) between social positions and preferences, the concept of habitus captures how the dispositions of similarly positioned agents are defined relationally and in distinction from all others. The dispositions of others therefore serve as negative reference points. When social differences as embodied dispositions get translated into choices, then they serve to strengthen social distinctions (Bourdieu 1984).

In various contexts from small meetings in national offices and analyses in official documents to those of international task forces and conferences, national statisticians discuss, debate and struggle over the practices of data scientists, their valuation, and what they mean for the authority of their data, expertise and the knowledge that they generate. While often non-coherent and involving contradictory claims, at the same time, there are recitations, repetitions and reiterations of truth claims about what is the problem, what are the solutions and so on. There are also patterns and momentary settlements about the constitution of both the profession of the data scientist and its relation to that of the national statistician. This struggle is the focus of our analysis below. In doing so we follow Judith Butler (1990) who argues that it is through such citations, repetitions, and resignifications of claims that truth is not just described but is performed. In brief, this is what Butler draws attention to in her take up of John Austin's (1962) theorizing of speech acts.[10] While some speech acts are descriptive of a state of affairs (what Austin named constative), others are performative (what Austin named illocutionary and perlocutionary). The latter have a force that creates a potential effect in a state of affairs that they seek to describe. What determines whether this force will have an effect is whether there

[10] We only briefly summarise the key points for us of Austin and Butler on speech acts; for further elaboration see (Isin and Ruppert 2015).

is an uptake, that is, changes in a state of affairs such as the adoption of new conventions and practices.

For us, the force of recitations, repetitions and reiterations of claims about the data scientist and national statistician is to be found in their uptake in specific material-semiotic practices such as experiments, demonstrations, and job descriptions. That is, while discursively performed, definitions of what is at stake, problematisations and solutions are also constituted through doing things. Struggles are not just performed through words but through what national statisticians do. Drawing on Isin and Ruppert's (2015) conception of digital acts, we examine how statisticians define and differentiate the data scientist from their craft through not only what they say but what they do in practices such as data boot camps, hackathons, innovation labs, method experiments, job descriptions and sandboxes. Furthermore, it is through such doings that the habitus of the data scientist is constituted and experienced.

We thus attend to struggles within the field as involving how national statisticians talk about, define, problematize and become data scientists but also what they do, their practices, and how through these they also define, distinguish and dispositionally come to occupy their position within the transnational field. To capture the recognition of different types of capital we do not only draw on literal references to data science and data scientists in our analysis. When referring to 'the data scientist' or for that matter 'national statistician' moreover, we are adopting the terminology of the field. Our intention is not to suggest that there is one universal characterization of professions but that there are specific formulations within different practices.

Recruiting data scientists

Our first practice concerns the recruitment of data scientists, where the same process is used to interview and assess several candidates. The recruitment process materially frames data science through job interviews, where applicants appear in front of a committee and convince the committee that their previous experience and knowledge are sufficient for the role of the data scientist, while the committee itself judges whether the applicants' responses fulfil the requirements. The process is discursively framed through application documents, including the job description, guidance for candidates, sample multiple choice tests, and other supporting texts. The recruitment process constructs a kind of data scientist, clearly defined and framed prior to the encounter. However, the definition does not end with the documents supplied by the employer. It is further refined through the submitted CVs, the performance of the candidates, and the discussions of the committee. It is not merely an exercise of fitting people

into checkboxes, but rather the definition of the task and job are changed and refined throughout the process.

In this section, we draw on material gathered while shadowing a statistician in 2015 as she interviewed applicants for data scientist posts distributed across several departments within the government. The interview committee included two other civil servants, one from the human resources department of the NSI, and one from another government agency. Each interview lasted from 45 minutes to one hour. The candidates were also required to take a multiple-choice test in an adjacent room following their interview, which included questions on basic statistics knowledge such as term definitions, probability calculations, etc. Their eventual placement would be in various institutes within government, but their assignments were to be decided after the interview process.

The job description document presented an ideal type for the data scientist: a collection of skills in programming/computing, data, and statistics. The interviewers were asked to formulate questions that assessed the candidate's competency in different skills. They were provided with a "marking matrix", a document listing the categories and the grades they should use to assess the performance of the interviewees. The marking matrix referred back to the ideal type, distributed over two categories of questions, "job specific" and "competency", each with four subcategories. Job-specific categories referred to the technical skills of data scientists, including computing (mostly related to programming languages), scripting (related to statistical tools such as R, SAS, SPSS), software (referring to "Big Data technology" such as NoSQL, Hadoop, Spark, etc.), and statistical skills (referring to traditional statistics knowledge, which test should be used for which problem, how to determine if a sample is representative, etc.). The competency category referred to broader skills, applicable for all civil service positions, defining a common core of skills and attitudes that civil servants are expected to possess. These included collaboration, personal improvement, meeting deadlines, leadership, and communication (including the ability to explain technical issues to non-technical audiences).

To be accepted as data scientists, the candidates needed to demonstrate statistical expertise. Several questions at the interview were designed toward this end, such as "how do you know if your result is statistically significant", or "how did you know if your sample represented the population?" When one of the candidates provided inadequate answers to these questions, the interviewers added a note to his application during the assessment round, asking him to "Please look at the statistical techniques required [for the position]". When asked about tools, most candidates brought up Matlab, SAS, SPSS and R as familiar software. In particular,

R was discussed as an open-source tool, being "less clunky than SPSS" in the words of one candidate.

The interviewers queried the applicants' familiarity with Big Data through questions such as: "What did you learn from your experiences working with Big Data projects?" One candidate replied with "use fewer programming languages", displaying a familiarity with data science practice by referring to the shared perception of the proliferation of tools and languages. Through his answer, the candidate implied that some technologies were used for the sake of having used a new technology in the project, and that these types of activities did not belong in proper data science.

Following each interview the interviewers were required to individually assign different scores to the eight subcategories using a scale from one to seven. The evaluation also involved a multiple-choice assessment for some categories, where the interviewers were expected to tick under "positive", "needs development", or to leave it blank. After the interviewers filled in their forms individually, they were also required to reach consensus on the final assessment of the candidate, although this did not prove very difficult for them as their final scores in most categories were either the same, or differed by only one point. During one such discussion, one of the interviewers stated that given sufficient background in other related tasks (for example a quantitative PhD, or prior experience in statistical programming), the candidates would be able to pick up some of the skills even if they did not seem to possess them at the time of interview.

Data scientist as accumulated skill and potential

Who are the data scientists as enacted by the job interview? They are able to program, acquire new technical skills quickly, have basic statistical knowledge, be familiar with the discourse of Big Data, be reflexive about not only the division between the highly technical and the traditional statistical, but also their own position within various government departments. They are not merely programmers or developers as they possess statistical expertise, but they are also more than just methodologists as they do not rely on other developers to conduct their study or produce their results. The data scientists combine statistical knowledge with new forms of data analysis. At the same time, the data scientists of the job interview are not hackers. They do not solve problems through small, localised fixes. Instead, they follow specific methodologies informed by traditional statistical practice. In short, the job interview enacts the data scientist as a set of skills and attitudes. Candidates are expected to possess capital in the form of particular accumulated technical skills such as statistical analysis and programming

that can be converted to advantage in the ongoing struggle to define the field of data science (Halford and Savage 2010).

The data scientists sought in these interviews are not bound to a specific government task or practice. They can be placed in different government departments, but still contribute their own set of skills independent of domain. In other words, the capital of the data scientist is highly convertible, it allows them to work in different domains with the same set of skills. However, these posts are all part of government, and the recruited data scientists are thus expected to perform as civil servants, in ways listed under the competencies category of the marking matrix.

The candidates need to possess certain cultural capital such as statistical expertise and related technical skills to succeed in the recruitment process, but as the interviewers also acknowledge, they are evaluated in relation to their potential to become data scientists. That is, being a data scientist involves a process that builds on capital that a candidate already possesses but through which they must demonstrate the capacity to learn and acquire yet unknown skills. To become a data scientist, in other words, is an ongoing process of accumulating capital by engaging with different fields and acquiring additional skills such as new programming languages, or familiarity with new data analysis tools as technologies change and evolve. Rather than settled, the data scientist is understood as a profession constantly in-the making.

The recruitment of new data scientists valorises new skills in the statistical practices of government. However, skill alone is not sufficient but must be bundled with other forms of capital such as statistical knowledge, as well as a particular habitus as we discuss in the following section. However, in this specific job bundle, technical skills count for more when granting legitimacy to the performance of the data scientist candidate. Candidates argue for why different skills should be considered part of the bundle, attempt to configure what cultural capital advantages data scientists, and thereby define who should count as one. Some skills, for example familiarity with database management, once relegated to IT-specialists, now play a much more prominent role defining the skills of government statisticians.

The question and answer format of the job interview enacts a kind of data scientist, anchored around verbal declarations of accumulated skill coupled with the evaluation of their potential. However, as we have already suggested, being a data scientist also involves cultural capital understood as the acquisition of particular sensibilities. In the next section, we discuss the shaping of a data science habitus through our investigations of data camps.

Data camp: From skills to sensibilities

Big Data sensibilities

As NSIs start experimenting with Big Data statistics, they need employees not only with data science skills but also with "Big Data sensibilities", as stated by a senior national statistician. In this section, we discuss the shaping of what we call a data science habitus, embodied cultural capital that includes skills, tastes, habits, normative inclinations and other knowledges that are not normally made explicit. Indeed, in the absence of a singular definition of data science, the shaping of a habitus functions to form who are data scientists and how they differ from or resemble national statisticians. We discuss one practice through which a habitus is shaped to suggest that it is not only acquired through discursive practices, but also through practical exercises such as experiments, sandboxes and boot camps, that test, develop and demonstrate the possibilities of new data sources and analytic techniques.

The specific material-semiotic practice we discuss is a data camp in which national statisticians, students, and PhDs in computer science and related disciplines participated. The aims of the camp were to develop uses of Big Data for official statistics; to increase statisticians' knowledge and familiarity with techniques; to profile the NSI as a future employer for data scientists; and to strengthen the ties between the NSI and the university. The camp was modelled after a hackathon, and included skills training, lectures, presentations and, of course, collaborative work. We suggest that these practices 'make explicit' (Muniesa and Linhardt 2011) what Big Data sensibilities might entail.

Data camp

During the data camp, twenty participants and seven mentors from the NSI and the university stayed on a university campus for a week. The mixed NSI-university teams worked until late at night on topics such as profiling Twitter populations and using road sensor data as an indicator of economic growth, not even stopping work during the 'data dinners'. The participants engaged in three main activities. The first and central part of the data camp was collaborative work. To get statisticians up to speed with the programming languages and software necessary for processing and analysing large data sets (in this case mainly Spark and Hadoop), skill training sessions were provided during the first days. At the end of each day, the group evaluated their findings and identified problems they were struggling with. The second main activity was attending lectures by academic or professional experts, for instance on 'Big Data processing principles' or 'Setting up research questions'. The third main activity was participation in the final presentation day. The teams presented their results for the NSI

and university higher management. The event took place in the VIP room of the university, and included a ceremonial signing of a memorandum of understanding about future collaboration between the two institutions.

A closer look at these activities indicates that for statisticians to work with Big Data, they need to acquire more than skill. Regarding collaborative work, participants stated the relevance of algorithms by referring to the commands and codes that help them execute a wide variety of automated work: converting data sets, classifying data for more insight and analysis, codes that extract and select relevant data, mining text, calculating values, and finally implementing analytic models. 'Algorithms' therefore covered a wide variety of automated work. In their evaluations and reports, however, participants emphasised the relevance of algorithms to process data and to get insight into data sets. As one participant stated in his project report: "The initial work focussed on accessing and getting to know the data. We tried Spark-R and Pig-Latin for this. Since using Pig-Latin was successful, we did not try any other language. Pig-Latin was used to study the data set in more detail."

Such statements of relevance, we suggest, amount to more than acquiring necessary skills. The participants articulated that it required an 'appreciation' of algorithms. For instance, one of the outcomes of an evening evaluation session was that "algorithms love statistics" (see Figure 1). Yet, the participants also experienced that algorithms did not necessarily make data processing a quick and simple task. Data from sources such as Twitter differed from administrative sources, and required extra attention due to the variations in their formats and their changeability over time. One NSI participant commented on the relevance of the location of commas, and the elaborate work required to prepare data sets for analysis. This work, the group commented, required patience. As automated correction and processing work also happens at NSIs, using algorithms and cultivating patience was not entirely new for statisticians. What was new, however, was a conception of a relationship between algorithms and statistics. Statisticians expressed an imagined closeness between algorithms and statistics (and therefore themselves) that helped them not only understand but also capitalise on data.

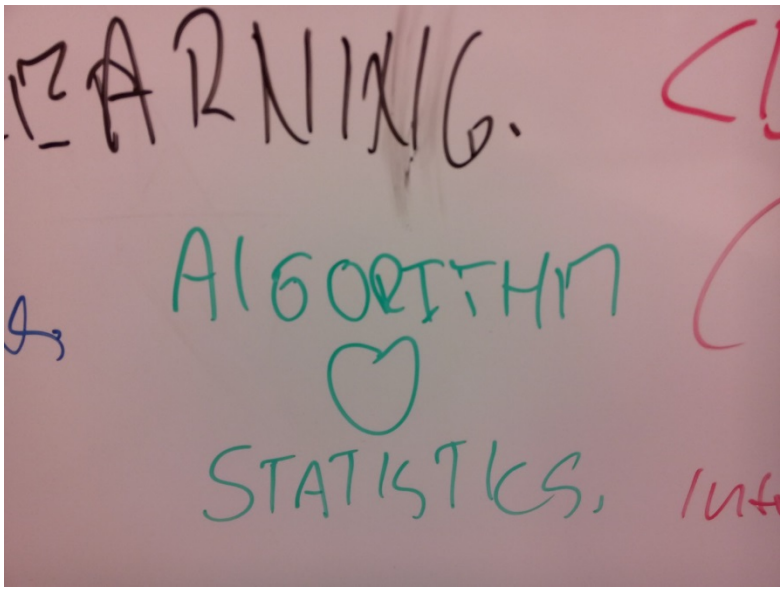

Figure 1: The love of algorithms. Photo of the data camp's whiteboard after group evaluation.

Two lectures on visualisation were especially instructive about data science sensibilities. The first lecture covered the practices of a NGO working according to the principle of 'objects of concern'.[11] They stated that this might mean increasing the visibility of a local phenomenon (like deforestation) on a map, in order to draw attention to it. 'You have to take a position', the organisation's CEO stated, 'not exaggerating is making a choice as well'. Similar practices were adopted by the teams for different purposes. For instance, one team created a threshold for visualising a phenomenon because they thought it would help their analysis.

The second lecture by an NSI statistician contrasted with this practice. When the presenter was asked whether his visualisations had an explicit political viewpoint, he responded that he left the politics to the audience, 'so you [analysts and statisticians] don't have to make choices'. The sensibilities negotiated here concerned relations between what they considered to be realities and their representations. As summed up by one of the mentors: "You lie a bit with statistics (…) you have to torture them until they confess". The lecture shows that national statisticians are not unfamiliar with such practices, even though they usually argue for the objectivity of statistics in public settings (as argued by Desrosières 2001).

The lectures were informative about another aspect of a data science habitus in-the-making. The NGO visualisations were aesthetically more pleasing than the NSI visualisations,

[11] After 'matters of concern' Latour (2004).

which were presented as mostly an aid to data processing. The NGOs visuals were detailed, interactive, applied subtle colour schemes, and were easy to grasp because they were based on geographic maps. The NSI visuals, although innovative, were clunky, less concerned with continuous and cohesive colour schemes and less easy to read for audiences other than national statisticians. The data camp mentors encouraged attractive visualisations by stating their preferences, as one NSI mentor stated: 'It would be great if we had something like the [NGO] visualisations on our website'. The teams were also coached actively to produce such visuals.

Appreciating the aesthetics of visualisation is relevant for two reasons. First, attractive visualisations draw in audiences. Second, and what we highlight here, they also serve as analytic tools. Visualisation is not only an end-product, but a method for the analyst to understand large volumes of data that are not easily analysed using traditional techniques such as graphs. Aesthetics contribute to this aim, in the form of contrasts, colours and animations that facilitate analysis. So, while maps, graphs and diagrams have always been part of statistical analysis, the difference here is the appreciation of, and the acquisition of a taste for, the possibilities of advanced aesthetics.

A final sensibility was introduced by the NSI mentors in the context of preparing for the closing presentations: doing result-oriented projects for a specified group of users. This included requirements typical for NSIs whose role is to supply statistics to clearly defined user groups such as departments and government agencies. Working within government, they are accustomed to contributing to public discussions and policy making as a professional responsibility. As the mentors also reminded, making a convincing case requires indicating, or at least hinting at, the value (or 'business case') of projects. Alongside these sensibilities, being inquisitive and taking risks were appreciated and encouraged by the participants, the mentors, and higher management. As the Director-General of the NSI stated on the final presentation day: "Especially experimentation I liked very much". In sum, the data camp showed how the statisticians' norms of contributing to government policy and debate co-existed, and sometimes clashed, with an appreciation of experimentation and a sensibility for doing trial-and-error work. Statements such as 'there is a lot in the data' often helped resolve such frictions because they indicated that a trail-and-error process might lead to results at a later stage.

A data science habitus for official statistics

Various elements of a new type of cultural capital were recognised through the data camp: tweaking representations of reality; a feel for the business case and users; the aesthetics of visualisations; experimentation; patience; and an affinity for and appreciation of algorithms. In

addition, the data camp allowed statisticians to build personal networks of support and generate demonstrations of their capabilities that might allow them to operate according to these sensibilities in their statistical institutes.

The foregoing is not exhaustive or comprehensive nor an account of the formation of a universal data science habitus. Rather, it is an account of practices involved in the shaping of a set of 'data science sensibilities' specifically in relation to a faction of the transnational field of statistics. Which sensibilities are valued and become part of a habitus that individuals start to embody happens through training, and in this case, by doing things (cf. Dobbin 2008; Franssen 2015). These sensibilities are not separate from but a consequence of changing relations between statisticians, forms of data, and their methods of analysis. Working with algorithms and visualisations, as we have noted, requires sensibilities for aesthetics, patience and a closeness to data that are otherwise not easily interpretable. These are dispositions of 'how to be with data' that form a shared habitus.[12]

In sum, the shaping of this habitus required practical exercises that put something to the test in order to be actualised. This is what Fabian Muniesa and Dominique Linhardt refer to as 'making explicit' (2011). Such actualisation does not unfold without problems, hesitations or tensions, as exemplified by the tension between trial-and-error work and producing short-term results. Making things explicit is exactly that: making sensibilities visible and putting them up for consideration, debate or negotiation. The significance of practical exercises is also recognised in appeals at international conferences and meetings where repetitions of definitions of Big Data lead to appeals to 'learn by doing' and in this way differentiate the data scientist from the 'iStatistician'.

Towards the iStatistician: Define, Demonstrate, Differentiate and Defend

> We need more investment, more infrastructure, more innovation, being 'smart', raw data into intelligence, skills needed, competitiveness, and so on. The repetition was rather numbing.[13]

[12] We are not suggesting that statisticians have a more or less embodied relationship with data compared to other professionals but that specific dispositions and their embodiment might be different. Studies of analytic practices in other fields point out that working with data also involves developing a sense for data (see, for instance, Myers 2014). When we refer to 'sensibilities' here, we primarily mean a sense of what dispositions are valued by this faction of the field and as part of how statisticians relate to data.

[13] Fieldwork Memo: European Data Forum 2015. Exploiting Data Integration, Luxembourg, 16-17 November 2015.

The quote above comes from a fieldwork memo that noted the mundane repetition of arguments of statisticians across various international meetings. Rather than deepening over time, such arguments usually remain undeveloped along with the uncritical adoption of the oft-repeated claim that Big Data is 'raw'.[14] These repetitions highlight that what is understood as data science and in turn a data scientist happens not only through 'doing things' but also discursive practices of imagining and defining the object of concern. Like other professional fields, this takes place in complex networks of interactions and exchanges from small meetings and official documents to those of international task forces and conferences where national statisticians discuss, debate and struggle over definitions.

In this section, we draw on practices that involve presentations and discussions at international meetings we observed (Figure 2). These include, for example, task forces of Eurostat and various meetings of population and housing experts at the UNECE and Conference of European Statisticians. Presentations invariably included a series of repetitions that were not defended as much as put forward as something that everybody knows. We identify four repetitions: *defining* what Big Data is, how it is better than traditional sources, and what skills it demands; *demonstrating* how Big Data can or cannot deliver official statistics; *differentiating* the data scientist from the figure of the iStatistician; and *defending* the authority of NSIs to establish, and evaluate adherence to, the legitimate criteria for the production official statistics.

Figure 2: International conference

[14] Fieldwork memo: Q2016 conference: Eurostat Biannual Quality Conference, Madrid, ES, 1 – 2 June 2016.

Define

> Big Data is characterized as data sets of increasing volume, velocity and variety; the 3 V's.[15]

The emergence of new professions involves identifying and defining an object of concern that calls upon new skills and expertise not currently being met. For this reason, defining an object and profession are very much entwined. When national statisticians defined Big Data they usually adopted the '3 Vs', one of the most prominent repetitions in this and other fields. It is a definition born of industry, though taken up and in some cases expanded within the academy.[16] One definition offered at an international meeting built on this formulation to highlight how it gives rise to new IT-issues compared to more traditional data sources:

> Big data is data that is difficult to collect, store or process within the conventional systems of statistical organizations. Either, their volume, velocity, structure or variety requires the adoption of new statistical software processing techniques and/or IT infrastructure to enable cost-effective insights to be made.[17]

Over time, the definition was further expanded:

> Every day, 2.5 quintillion bytes of data are created. These data come from digital pictures, videos, post to social media sites, intelligent sensors, purchase transaction records, and cell phones' GPS signals, to name a few, and comply with the following attributes: volume, velocity, variety, veracity, variability and value, in other words: Big Data.[18]

Such definitions led to repetitions of how Big Data is 'better' than data from traditional sources such as censuses and surveys. Most commonly, these potential improvements were identified in relation to the European Statistics Code of Practice, which defines five aspects of statistical output quality: relevance; accuracy and reliability; timeliness and punctuality; coherence and comparability; accessibility and clarity.[19] That Big Data could possibly meet these principles better than existing forms of data was key to its evaluation and the identification of what is at stake.

---

[15] UNECE. 2013. What does "Big data" mean for official statistics? Economic Commission for Europe. Conference of European Statisticians. Sixty-first plenary session Geneva, 10-12 June 2013.
[16] See for e.g., (Kitchin 2013) and (K. N. Cukier and Mayer-Schonberger 2013).
[17] UNECE. 2014. 'How big is Big Data? Exploring the role of Big Data in Official Statistics.' Draft paper. UNECE Statistics Wikis. Avail. at http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=99484307.
[18] UNECE. 2016. Interim Report of the Task Force on the Value of Official Statistics. Conference of European Statisticians. Sixty-fourth plenary session. Paris, 27-29 April.
[19] European Statistical Services Committee. 2011.

Definitions and valuations of Big Data were then often translated into the skills required and in turn what constitutes a data scientist. At one international meeting, a national statistician described 'hardcore data scientists' as having broad knowledge plus specialist skills and who can work with Big Data systems and process knowledge.[20] At another meeting a statistician stated that what he observes is that other disciplines have been quicker to adopt these skills such as computer scientists and they are taking up the 'high' positions that economists once did; they sell themselves better; and that it is harder for statisticians because their interests are substantively different.[21] In these ways, how national statisticians defined, problematized and valued Big Data were very much entangled with how they then came to identify the forms of cultural capital that make up the profession of a data scientist.

Demonstrate

The role of repetitions secured a relative settlement on a definition, problematisation and valuation of Big Data, and in turn the skills that it demands. But, while necessary, such definitional settlement remained insufficient as it lacked specificity in relation to how Big Data would practically compare to long-standing and trusted official statistics. If the profession of data science and in turn the scientist are to be relevant to the field of statistics, then experiencing, testing and demonstrating or as commonly recited – 'learning by doing' -- was required. Laboratories, sandboxes and pilots were practical means identified at meetings for moving from definitions to demonstrations:

> The key critical success factor in the action plan is an agreement at ESS level to embark on a number of concrete pilot projects. A real understanding of the implications of big data for official statistics can only be gained through hands-on experience, 'learning by doing'. Different actors have already gained experience conducting pilots in their respective organisations at global, European and at national level. The purpose of the pilots is to gain experience in using big data in the context of official statistics.[22]

Pilots were noted as ways to gain experience in identifying and analysing the potential of Big Data for statistical data production.

> [The] "Sandbox" environment has been created to provide a technical platform to load Big Data sets and tools. It gives participating statistical organisations the opportunity to test the feasibility of remote access and

[20] Fieldwork notes. February 2015.
[21] Fieldwork notes. April 2014.
[22] Fieldwork notes. Oct 2014.

> processing; test whether existing statistical standards / models / methods etc. can be applied to Big Data; ...[and] "learning by doing".[23]

Repeated definitions of Big Data or the data scientist were thus insufficient to appreciate the implications for the profession of national statisticians. Beyond discursive claims, specific material-semiotic practices of doing things were recognised as necessary. As such, the latter were not separate from the former but together were part-and-parcel of the struggle to understand the implications of Big Data for national statisticians through their practicing and testing of specific skills recognised as the domain of data science.

Differentiate

It is through demonstrations that discourses about the promise of Big Data led to more nuanced critiques and started to focus on differentiating the data scientist from the figure of the national statistician. This was exemplified in the description of the figure of the iStatistician by Walter Radermacher, the Director General of Eurostat. He set out this role in the context of a 'data revolution' and in relation to a genealogy of statistics beginning with its birth during the period of state formation in the first part of nineteenth century noting the shared etymology of state and statistics.[24] This he called the 'descriptive' era of Statistics 1.0. In the twentieth century came Statistics 2.0 with the move to mathematics, inference, surveys and sampling. Statistics 3.0 came later in that century with the introduction of new technologies and IT infrastructures. The twentieth century was then characterized as the 'data revolution,' big data, machine-to-machine communication and modelling and the era of Statistics 4.0.

The iStatistician, he argued, is the professional of Statistics 4.0. The classical profile of the statistician in earlier eras was that of a data gatherer and information generator who was invested in a 'data collection machinery.' In Statistics 4.0 that profile must move to less data generation to a focus on information and knowledge generation along with in-depth knowledge of both statistics and information technologies. That the iStatistician was chosen is an interesting strategic move. By 2016 the appropriateness of the term Big Data was in question; though always debated as a term and the hype it generated, Big Data started to lose fashion in 2015 as other practices such as machine learning, internet of things and artificial intelligence became more prominent. This was illustrated in the move of the UNECE Big Data project from

---

[23] UNECE. 2014. 'Sandbox.' UNECE Statistics Wikis. Avail. at www1.unece.org/stat/platform/display/bigdata/Sandbox.

[24] Fieldwork Notes. From the opening address of Walter Radermacher, the Director General of Eurostat at the 'New Techniques and Technologies for Statistics (NTTS)' 2015 conference in Brussels, BE, an international biennial scientific gathering organised by Eurostat.

a focus on Big Data to that of data integration. The rationale was that Big Data is insufficient to serve the purposes of official statistics and only when taken together and integrated with multiple and especially official data sources can their value be realised. This conclusion followed several years of experiments and pilots that individually revealed the limitations of data from mobile phones, search engine queries or social media sources.

The definition of the iStatistician proposed by the Director General marked out a similar distinction. Rather than focusing on data, his concern was the analytic skills required to manipulate high volume datasets, deal with uncertainty, and work with predictive analytics, machine learning, and modelling. But in a move to further differentiate the statistician, he noted that skills are insufficient: only the iStatistician can generate trust and confidence in advanced analytics and produce high quality products while at the same time deal with political issues such as data privacy. Relationally then, the data scientist is distinguished as having specific analytic skills but lacking other competencies that are the province of the iStatistician.

Defend

At a meeting where a draft version of this chapter was presented, national statisticians offered that the skills – or cultural and embodied capital - to work with Big Data are not the dominion of one professional, whether named a data scientist or iStatistician.[25] Rather, a 'Da Vinci team' composed of professionals with different and complementary skills is necessary that combines knowledge of IT with soft skills such as communication, for example. They also noted that the creation of data scientist positions within NSIs is a matter of ongoing debate; some argued that what is required is the reskilling of national statisticians and not new positions. We suggest that their observations resonate with the competencies of the iStatistician and their emplacement in collectives, the offices of national statistics. Furthermore, they highlight that while Big Data and data scientists are bringing into question the skills and competencies of national statisticians, they are also being mobilised to reinforce and defend existing values that they command such as trustworthiness, public accountability, civil service and democratic legitimacy. These principles constitute another repetition often asserted at international meetings in relation to Big Data; that the investments of NSIs in myriad forms of data and their capacities to secure the principles of official statistics ensure the relative advantage of national statisticians in the future:

[25] The meeting took place on 6 February 2017 at Goldsmiths, University of London.

> It is unlikely that NSOs [NSIs] will lose the "official statistics" trademark but they could slowly lose their reputation and relevance unless they get on board. One big advantage that NSOs have is the existence of infrastructures to address the accuracy, consistency and interpretability of the statistics produced. By incorporating relevant Big data sources into their official statistics process NSOs are best positioned to measure their accuracy, ensure the consistency of the whole systems of official statistics and providing interpretation while constantly working on relevance and timeliness. The role and importance of official statistics will thus be protected (UNECE 2013).

National statisticians, in other words, assert their authority to establish, but also to evaluate adherence to, the legitimate criteria for the production official statistics. The effects of Big Data are thus both disruptive and continuous. Data scientists are not 'taking over' or replacing national statisticians but they are being relationally reconfigured through a process of differentiation. As part of this, while requiring new skills, Big Data is also (potentially) reinforcing established values and norms for the legitimate production of official statistics. That is, struggles over the naming and defining of Big Data and professions are part of larger stakes in the production, consecration and institutionalisation of official statistics.

Conclusion

The production of official statistics involves practices where the uptake of Big Data is leading to changing valuations of the relative authority of skills, expertise and knowledge within the transnational field of statistics. Big Data is an object of investment whose value is being produced by the competitive struggles of professionals who claim stakes in its meaning and functioning. Exactly how a new faction - the data scientist - and the remaking of another – the national statistician - relationally take shape and have consequences for how official statistics are generated, valued and trusted is yet to be known. We have elaborated this as a struggle that is happening through three kinds of practices that are very much connected to political struggles over the legitimate exercise of what Bourdieu named symbolic violence: the power to name and make things. This is one meaning we give to the politics of method: that the object of study and the professions that invest in them are being defined at the same time.

A final aspect of the politics of method concerns our own practice, about which we draw three observations. First, ethnography has enabled us to account for often undocumented practices involved in the remaking and emergence of professions such as the formation of a habitus. Second, a collaborative ethnography enabled us to trace how these practices are transversal and happen across numerous sites that are not of similar significance but fulfil specific roles in the formation of professions. Third, we emphasised the importance of

documenting and interpreting not only discursive struggles but also material-semiotic practices. As we have noted, many national statisticians are aware of the role of 'learning by doing' as evident in their insistence on experimentation and demonstration. In that regard, we also suggest that such practices allow them to be their own ethnographers: to observe each other and reflect on what their observations mean for their profession.

Acknowledgements

The research leading to this publication has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement no. 615588. Principal Investigator, Evelyn Ruppert, Goldsmiths, University of London.

References

Austin, J. L. 1962. *How to Do Things with Words*. Oxford: Oxford University Press.

Bigo, Didier. 2011. "Pierre Bourdieu and International Relations: Power of Practices, Practices of Power." *International Political Sociology* 5 (3): 225–58.

Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgement of Taste*. Cambridge: Harvard University Press.

———. 1988. "Social Space and Symbolic Power." *Sociological Theory* 7: 14–25.

Bourdieu, Pierre, and Loic Wacquant. 2006. *Reflexive Anthropologie*. Frankfurt am Main: Suhrkamp.

Burrows, Roger, and Mike Savage. 2014. "After the Crisis? Big Data and the Methodological Challenges of Empirical Sociology." *Big Data & Society* 1 (1).

Butler, Judith. 1990. *Gender Trouble: Feminism and the Subversion of Identity*. Routledge.

Cukier, Kenneth. 2010. "Data, Data Everywhere." *The Economist*. http://www.economist.com/node/15557443.

Cukier, Kenneth Neil, and Viktor Mayer-Schonberger. 2013. *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murray.

Desrosières, Alain. 1998. *The Politics of Large Numbers: A History of Statistical Reasoning*. Edited by R. D. Whitley. Cambridge, MA and London: Harvard University Press.

———. 2001. "How Real Are Statistics? Four Possible Attitudes." *Social Research* 68 (2): 339–55.

Dezalay, Yves, and Bryant G. Garth. 1996. *Dealing in Virtue: International Commercial Arbitration and the Construction of a Transnational Legal Order*. Chicago: University of Chicago Press.

Dobbin, Frank. 2008. "The Poverty of Organizational Theory: Comment on: 'Bourdieu and Organizational Analysis.'" *Theory and Society* 37 (1): 53–63.

Donoho, David. 2015. "50 Years of Data Science" presented at the Tukey Centennial workshop, Princeton, N.J., USA, September 18, 2015. http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf.

European Statistical Services Committee. 2011. "European Statistics Code of Practice for the National and Community Statistical Authorities." Luxembourg: Eurostat.

Eurostat. 2014. ‘Big data – an opportunity or a threat to official statistics?’ Presentation to the Economic Commission for Europe Conference of European Statisticians. Sixty-second plenary session. Paris, 9-11 April 2014.
Franssen, Thomas. 2015. “How Books Travel: Translation Flows and Practices of Dutch Acquiring Editors and New York Literary Scouts, 1980-2009.” PhD Thesis, Amsterdam: University of Amsterdam.
Gehl, Robert W. 2015. “Sharing, Knowledge Management and Big Data: A Partial Genealogy of the Data Scientist.” *European Journal of Cultural Studies* 18 (4–5): 413–28.
Halavais, Alexander. 2015. “Bigger Sociological Imaginations: Framing Big Social Data Theory and Methods.” *Information, Communication & Society* 18 (5): 583–94.
Halford, S., and M. Savage. 2010. “Reconceptualizing Digital Social Inequality.” *Information, Communication and Society* 13 (7): 937–55.
INSEE. 2016. “In October 2016, the French Business Climate Is Stable.” *INSEE: Economic Indicators*. http://www.insee.fr/en/themes/info-rapide.asp?id=105&date=20161025.
Isin, Engin, and Evelyn Ruppert. 2015. *Being Digital Citizens*. London: Rowman & Littlefield International.
Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE.
Latour, Bruno. 2004. “Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern.” *Critical Inquiry* 30 (2): 225–48.
Loukides, Mike. 2010. “What Is Data Science?” *O’Reilly Media*. https://www.oreilly.com/ideas/what-is-data-science.
Madsen, Mikael Rask. 2011. “Reflexivity and the Construction of the International Object: The Case of Human Rights.” *International Political Sociology* 5 (3): 259–75.
———. 2014. “The International Judiciary as Transnational Power Elite.” *International Political Sociology* 8 (3): 332–34.
Meulman, Jacqueline. 2016. “When It Comes to Data, Size Isn’t Everything.” *STAtOR* 17 (2): 37–38.
Muniesa, Fabian, and Dominique Linhardt. 2011. “Trials of Explicitness in the Implementation of Public Management Reform.” *Critical Perspectives on Accounting* 22 (6): 550–66.
Myers, Natasha. 2014. “Rendering Machinic Life.” In *Representation in Scientific Practice Revisited*, edited by Catelijne Coopmans, Janet Vertesi, Michael E. Lynch, and Woolgar, Steve, 153–77. Cambridge, MA: The MIT Press.
Naur, Peter. 1974. *Concise Survey of Computer Methods*. Sweden: Petrocelli Books.
Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.
Patil, Thomas H., and D. J. Davenport. 2012. “Data Scientist: The Sexiest Job of the 21st Century.” *Harvard Business Review*. https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century.
Ruppert, Evelyn. 2012. “Seeing Population: Census and Surveillance by Numbers.” In *Routledge International Handbook of Surveillance Studies*, edited by K. Ball, K. Haggerty, and D. Lyon, 209–16. London: Routledge.
Ruppert, Evelyn, John Law, and Mike Savage. 2013. “Reassembling Social Science Methods: The Challenge of Digital Devices.” *Theory, Culture & Society, Special Issue on “The Social Life of Methods”* 30 (4): 22–46.
Saari, Matti. 2016. “Statistics Finland - Preliminary Population Statistics.” *Statistics Finland*. http://www.stat.fi/til/vamuu/2016/09/vamuu_2016_09_2016-10-25_tie_001_en.html.

Savage, Mike. 2010. *Identities and Social Change in Britain since 1940: The Politics of Method*. Oxford: Oxford University Press.
Scheel, Stephan, Baki Cakici, Francisca Grommé, Evelyn Ruppert, Ville Takala, and Funda Ustek-Spilda. 2016. "Transcending Methodological Nationalism through Transversal Methods? On the Stakes and Challenges of Collaboration." ARITHMUS Working Paper No. 1. Goldsmiths College, University of London.
UNECE. 2016. Interim Report of the Task Force on the Value of Official Statistics. Conference of European Statisticians. Sixty-fourth plenary session. Paris, 27-29 April.
UNECE. 2014. 'How big is Big Data? Exploring the role of Big Data in Official Statistics.' Draft paper. UNECE Statistics Wikis. http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=99484307.
UNECE. 2014. 'Sandbox.' UNECE Statistics Wikis. www1.unece.org/stat/platform/display/bigdata/Sandbox.
UNECE. 2013. What does "Big data" mean for official statistics? Economic Commission for Europe. Conference of European Statisticians. Sixty-first plenary session Geneva, 10-12 June 2013.