Data Practices

Making Up a European People

Evelyn Ruppert
Stephan Scheel

Data Practices

Data Practices

Making Up a European People

Evelyn Ruppert
Stephan Scheel (eds)

Goldsmiths
Press

First published in 2021 by Goldsmiths Press
Goldsmiths, University of London, New Cross
London SE14 6NW

Printed and bound by Versa Press
Distribution by the MIT Press
Cambridge, Massachusetts, and London, England

A CIP record for this book is available from the British Library

ISBN 978-1-912685-85-1 (hbk)
ISBN 978-1-912685-86-8 (ebk)

www.gold.ac.uk/goldsmiths-press

Goldsmiths
UNIVERSITY OF LONDON

Contents

Acknowledgements

Data Practices: Making Up a European People is one outcome of a project, Peopling Europe: How data make a people (ARITHMUS). It was a five-year (2014–20) ERC-funded project led by one of the co-editors and co-authors, Evelyn Ruppert (PI); the second co-editor and co-author, Stephan Scheel, was a postdoctoral researcher along with three other co-authors, Baki Cakici, Francisca Grommé, and Funda Ustek-Spilda, and a doctoral researcher and co-author, Ville Takala. The project was funded by the European Research Council under the European Union's Seventh Framework Programme ((FP/2007–2013)/ERC Consolidator Grant Agreement no. 615588, Goldsmiths, University of London). In addition to supporting the research that led to this book, funding from that grant also enabled the book to be published open access. Of note is that the book is the product of collaborative ethnographic fieldwork, analysis, and writing and can be considered as both a research monograph and an edited collection. Rather than consisting of a series of different positions on data practices, it is a single intervention about the role of data practices in the making up of the population and people of Europe. That said, because individual chapters have been co-written by different authors and in some cases build upon already published articles, we have maintained the format of an edited collection.

We are grateful for the generosity of time and support of numerous national and international statisticians who supported the project in many ways such as providing access to meetings and interviews at their institutions and participating on an ARITHMUS Advisory Group: Barteld Braaksma and

Eric Schulte Nordholt (Statistics Netherlands); Jane Naylor (UK Office for National Statistics); Şebnem Beşe Canpolat and Muharrem Gürleyen Gök (Turkish Statistical Institute); Tuulikki Sillajõe and Kaja Sõstra (Statistics Estonia); Pasi Piela (Statistics Finland); Michail Skaliotis, Albrecht Wirthmann, and David Thoroughgood (Eurostat); and Paolo Valente and Fiona Willis-Núñez (United Nations Economic Commission for Europe). We would also like to thank the countless statisticians, policymakers, methodologists, and data scientists who participated in interviews and accepted our presence at their innovation labs, offices, working group and task force meetings, and conferences. Without their support the ethnographic research on which this book is based would not have been possible. While they and other statisticians may question or disagree with our analyses and interpretations, we hope that the book contributes a fresh perspective on longstanding debates on the political stakes of data practices in the making of official population statistics.

We also appreciate the feedback and suggestions of six anonymous reviewers at both the proposal and draft manuscript stages. Their reviews and comments were insightful and contributed to many improvements to the manuscript. We especially acknowledge the close reading of one reviewer whose astute suggestions helped us to elaborate and strengthen the conclusion.

The book was conceived during a May 2018 meeting at the Netherlands Institute for Advanced Study (NIAS) during Evelyn Ruppert's tenure as a Fellow and Van Doorn Honorary Chair at the School of Social and Behavioural Sciences, Erasmus University Rotterdam. We are thankful for the use of facilities and space in the NIAS building located in the heart of the old city centre of Amsterdam and the support of NIAS staff in making the meeting possible. The meeting served as

an opportunity to reflect on our collaborative research and to discuss an initial framing and plan for this book which proved to be fruitful.

Our final thanks are to Susan Kelly, Ellen Parnavelas, and Sarah Kember of Goldsmiths Press for their assistance and support at different stages during the writing and editing of *Data Practices: Making Up a European People*.

Biographies

Co-Editors and Co-Authors

Evelyn Ruppert is Professor of Sociology and Academic Lead for Research at Goldsmiths, University of London, UK. Evelyn was PI Peopling Europe: How data make a people (ARITHMUS; 2014–20). She is Founding Editor and Editor-in-Chief of the SAGE open access journal, *Big Data & Society*.

Stephan Scheel is Assistant Professor for Transnational Cooperation and Migration Research at the Institute of Sociology of the University of Duisburg, DE. Stephan was a Postdoctoral Researcher with ARITHMUS (2014–2018).

Co-Authors

Baki Cakici is Associate Professor in the Technologies in Practice Research Group at the IT University of Copenhagen, DK. Baki was a Postdoctoral Researcher with ARITHMUS (2014–2017).

Francisca Grommé is Assistant Professor at the Erasmus School of Social and Behavioural Sciences, Organizational Dynamics in the Digital Society, NL. Francisca was a Postdoctoral Researcher with ARITHMUS (2014–2018).

Ville Takala is Research Associate at the UCL Institute for Innovation and Public Purpose (IIPP), UK. Ville was a Doctoral Researcher with ARITHMUS (2014–2019).

Funda Ustek-Spilda is a Postdoctoral Researcher and Project Manager on the Fairwork project at the Oxford Internet Institute, University of Oxford, UK. Funda was a Postdoctoral Researcher with ARITHMUS (2014–2018).

1

Introduction: The Politics of Making Up a European People

Evelyn Ruppert and Stephan Scheel

Technological advances have powered the Digital and Data Revolutions. These raise legitimate questions about how effectively NSOs [National Statistical Organisations] are using these new possibilities to expand the benefits they provide to our societies … There are opportunities presented by these developments, which if they are wise, official statisticians will take in order to build on previous successes. But there are also threats. Failure to recognise these or to react to them with complacency could have the most serious consequences. At worst, official statistics could find itself partly or largely replaced by other information and data providers.

(United Nations Economic Commission for Europe (UNECE), 2018: 1)

This warning to statisticians to innovate or perish was written by the National Statistician of the United Kingdom in his preface to a report of the UNECE Task Force on the 'Value of Official Statistics'. What he expresses is a key concern of statisticians: that official population statistics such as those generated by traditional methods like questionnaire-based censuses and surveys are going through a transition and are at a crossroads as methods and data sources are being innovated and diversified. Digital technologies such as the internet, handheld

devices, and what is usually referred to as big data offer new possibilities to innovate methods and generate statistical knowledge about populations. The pressure to realise the potentials and expectations associated with these technologies and data is driven by discourses on reducing costs and what is referred to as respondent burden—the time and effort people have to invest to complete a survey or census questionnaire—as well as calls to produce more up-to-date and detailed statistics. These pressures and calls have prompted numerous experiments with sources of big data, such as those generated by social media, mobile phones, search engines and other internet platforms of private corporations. Statisticians are engaging in experiments that repurpose this data to produce migration, price, employment, and other statistics. But most significantly, perhaps, the usual questionnaire-based methods of censuses and surveys are being increasingly replaced by methods that draw on government administrative data (e.g., employment, national insurance) and central population and housing registers of addresses and basic personal information. In brief, statisticians are seeking to repurpose existing data rather than produce their own data for knowing populations and other phenomena relevant for governments such as on the economy to secure the future relevance of official statistics more generally.[1]

Arguably, these changes in statistical methods are the most fundamental since the beginning of modern population statistics over two centuries ago. Since then, official statistics have gone through many changes or phases as depicted by Walter Radermacher, who was the Director General of Eurostat, the statistical agency of the European Union, from 2008 until 2016. Reflecting on these changes Radermacher regards 'the handling of big data' and the related 'digitisation of all spheres of life' as the critical 'fourth phase' in the development of

official statistics (Radermacher, 2020: 3). The phases characterised by Radermacher roughly coincide with the last three centuries and key methodological developments: the growth of statistics as a science in the 18th century; the introduction of major scientific innovations such as inferential statistics in the 19th century; and the rise of new methods and practices in relation to computerisation in the late 20th century. Historically, all of these phases were tightly interwoven with the consolidation of the modern nation-state as reflected in the etymology of the word 'statistics', which means essentially 'state-istics', the science of the state (Schmidt, 2005: 15). What was distinctive about the new science of statistics from the 18th century onwards is the ambition to describe all state phenomena through numbers rather than words (15). Porter (1986) refers to this as the 'rise of statistical thinking', which was indispensable to create a legible population, and which Scott describes as a 'central problem of statecraft' (1998: 2). As Foucault (2003: 243–245) notes, it was the institutionalisation of official statistics, along with the emergence of other sciences such as biology and medicine, that heralded the 'remarkable entrance' (Foucault, 2009: 67) of the population as a central object of government in the 18th century. Statistical thinking was central to knowing populations by making it possible to study large numbers and discover regularities, causal relationships, and correlations through which populations could be shaped and governed according to political objectives and priorities (Desrosières, 1998; Hacking, 1990; Porter, 1986).

Importantly, the fourth phase of official statistics that Radermacher refers to involves the innovation and diversification of data and statistical methods that also coincide (and potentially sit in tension) with the European Union's (EU) efforts to harmonise population statistics across member states. It is this moment of change in statistical regimes and what it

means for making up a European population and people that was the focus of a project on which this book is based: *Peopling Europe: How data make a people* (ARITHMUS).

We elaborate on the project later but here note that its approach was to study this unique moment of methodological change by ethnographically following and documenting the situated practices involved in assembling a multiplicity of national populations into a governable European population. It began with the proposition that doing so faces practical and political problems and, whether intentional or otherwise, contribute to the *making up* of a European people. In brief, the ARITHMUS project started from the proposition that statistical methods do not just describe, measure, or count a population that already exists. Statistical methods are performative; they help to enact – that is, make up – a European population as a knowable object of government. That was the central argument of the project and now the basis of this book. It is an understanding that brought focus to the performativity of knowledge practices in the fieldwork conducted by ARITHMUS researchers and their analyses: how methodological changes in population statistics are not simply technical but also political matters. For if statistical methods enact rather than reflect populations, then different versions of populations are possible and changes in methods have implications for political questions such as: What is 'Europe' and who are 'Europeans'?

What is 'Europe' and who are 'Europeans'?

Europe is not singular but multiple. To say this is not to be playful but to highlight that Europe is not given but variously brought into being by a complex of imaginaries, laws, and governing practices. This multiplicity is most evident in the

number of institutional bodies that bear the name 'Europe', and which consist of different combinations of states such as the Council of Europe, UNECE, and the European Border and Coast Guard Agency (Frontex). Politically, the European Union is most significant but even that consists of political arrangements that differentially combine EU member and non-member states such as the Eurozone whose members share a single currency (19 EU states) or the Schengen passport free zone (22 EU and four associated non-EU states). Moreover, those arrangements belie complexities such as formal agreements with 'microstates' like the Vatican, which use the Euro as their official currency, for example, or associated countries which can contribute to and participate in EU programmes under the same legal conditions as member states. One consequence of these and other political arrangements is that, since its inception, the EU has operated at different degrees of integration and forms of cooperation between its member and other states.[2]

So, to refer to Europe as singular is certainly problematic. But so too is it problematic to consider these various arrangements as independent from social and political struggles that crisscross and shape them. At the time of writing, this was not least evident in the withdrawal of the United Kingdom from the European Union on 31 January 2020 (Brexit) and its relation to numerous political fissures such as the so-called migration crisis or the crisis of the Euro, the EU's single currency. These struggles are connected to a more fundamental challenge, that of the Union's legitimacy. The EU for long has said to be suffering from a 'democratic deficit'. This diagnosis is often accompanied by claims that a European demos does not exist, and that the EU is unable to address itself to a constituted polity (Balibar, 2003; Bellamy, 2008; Habermas, 2006; Konstadinidis, 1999; Shore, 2000).

Critical scholarship offers that answers to the question of legitimacy are not to be found in grand political statements by major figures or theories of how Europe can be made into a new political entity (Balibar, 2003: ix). Rather than attempting to address an imaginary demos or conjuring it up in proclamations about a European identity, critical scholars have shown that it is through specific practices such as laws, regulations, customs, histories, and institutions that Europe is enacted and continuously remade (Barry and Walters, 2003; Bellamy et al., 2006; Isin and Saward, 2013; Shore, 2000). What this body of scholarship advances is that making Europe relies on standardised practices and institutions for forging narratives and identities similar to those involved in nation-building and nationalism (Anderson, 2006; Best, 2009; Kertzer and Arel, 2002; Savage, 2010). It is through political technologies and related practices of statecraft such as taxation, market regulations, maps, museums, censuses, surveys, and much more that 'imagined communities' have been forged (Anderson, 2006).

Our first starting point in response to the question of what is Europe begins with this understanding: that like the formation of nations, the EU is brought into being via myriad practices. This is an argument put forward in several books in a series on *Making Europe* that maps the various institutional arrangements and networks that have come to constitute Europe during the 'Long Twentieth Century' (1850–2000). For example, in *Building Europe on Expertise*, Kohlrausch and Trischler (2014) examine how the integration of Europe historically involved the transnational circulation of knowledge amongst various networks of expert communities. They document how scientific and technological experts, and specifically a trained technical elite, were central to the construction and reconstruction of Europe (Kohlrausch and Trischler, 2014).

In a similar fashion, Kaiser and Schot (2014) argue in their book *Writing the Rules for Europe* that European integration is driven by a 'hidden integration' which operates through international technocratic committees that often work behind 'closed doors'. Critically, the book series does not conflate Europe with the EU and instead traces different 'zones' of collaboration where boundaries were often 'fuzzy' and extend beyond what is typically understood as Europe. We share the view that the EU is just one such zone, albeit one that has become a 'hegemonic force of Europeanization' (Kohlrausch and Trischler, 2014: x).

This book builds on and contributes to this body of scholarship by studying one such force involving expertise and rules that were central to the making of nations and now the EU: the census and more generally official population statistics. In the 18th century the census was, along with the map and museum, a political technology that 'profoundly shaped the way in which the colonial state imagined its dominion – the nature of the human beings it ruled, the geography of its domain, and the legitimacy of its ancestry' (Anderson, 2006: 163). Just as censuses connect numbers and nationhood (Patriarca, 1996) and make people 'singular' and 'legible' (Scott, 1998) so too, we argue, contemporary censuses contribute, along with population statistics more generally, to the making of the population and people of Europe. Our point of departure is that population statistics are not simply a set of knowledge practices that measure and numerically describe a European population that already exists. Rather, methods and their classifications, categories, definitions, and visualisations like graphs, maps, and tables help to *enact* a European population as an intelligible object for various strategies and technologies of government. At the same time, for the European project the matter of concern is not only knowing 'How many are we?' but

also to know 'Who are we?' If the EU is more than the sum of its national parts then, as Durkheim famously argued in relation to national statistics, what are the regularities that provide 'evidence of the autonomous existence' of a European society? (Hacking, 1990, 1999, 2002).

If EU institutions are to represent interests beyond those of individual member states, then that 'presupposes a transnational European public whose "general will" arises from common interests that can be represented and championed by [these] supranational bodies' (Shore, 2000: 19). Population statistics are thus critical because, like surveys such as the Eurobarometer, they do not just measure phenomena on a European scale, but rather help to constitute something like a European public in the first place (Law, 2009). Likewise, censuses and population statistics more broadly do not just count the people of Europe. They help to constitute what is the population and who are the people of Europe and are amongst many political technologies through which the EU seeks to secure its legitimacy.

It is with this understanding of the relation between political technologies like statistics and projects of nation-building that we approach the making of the political union of Europe. Rather than political debates and institutional struggles, we consider how apparently minor, technical struggles over methods shape the making of a European population and people. There are related critical approaches to such an understanding of the making of population statistics. The book *Demystifying Social Statistics* (Irvine, Miles, and Evans, 1979) argues, for instance, against the 'widely-held view of statistical data as a form of knowledge untainted by social values or ideology' where 'the role of the statistician is simply to clinically collect and preserve *the* facts' (italics in original, 1). In relation to censuses, others have directed attention to the categories

that are used in what are apparently only technical enumeration exercises in order to show how censuses help to constitute populations in particular, historically contingent ways. Nobles (2000), for instance, examines how censuses have been inextricably linked with discourses 'where ideas about race are worked through, categories constructed and then applied to public policy' (84). An edited collection by Kertzer and Arel (2002) demonstrates, in turn, that 'the census does much more than simply reflect social reality; rather, it plays a key role in the construction of that reality' and especially the division of 'national populations into separate identity categories: racial, ethnic, linguistic, or religious' (2).

While ARITHMUS took inspiration from these works, it departs from their conception of censuses and population statistics as social products or as something constructed. Rather than simply deliberate projects of willful human actors following predefined plans, we build on a related argument advanced by Scott that emphasises the performative power of practices of statecraft. Following Scott, maps 'were … not just maps. Rather, they were maps that, when allied with state power, would enable much of the reality they depicted to be remade. Thus a state cadastral map created to designate taxable property-holders does not merely describe a system of land tenure; it creates such a system through its ability to give its categories the force of law' (Scott, 1998: 3). As Scott suggests, technologies such as official statistics or maps do not simply represent populations and territories but literally enact them as objects of power. That is, they do not enumerate populations just to satisfy curiosities, but constitute populations as intelligible, negotiable, and actionable objects of government. In other words, like surveys (Law, 2009) and public opinion polls (Bourdieu, 1979; Osborne and Rose, 1999), censuses enact, that is, both represent and bring into being that which they

ostensibly only reflect (Law, 2009; Savage, 2010; Shore, 2000). For this reason, we put quotation marks at the beginning of this chapter around the words Europe and Europeans to specify that they are not given but enacted.

At the same time, as we have indicated above, official population statistics and censuses help to enact a distinct form of peoplehood (Lie, 2004) where commonality is more important than differences (Porter, 1986). For states, the dominant commonality is who are the people within their territory and thus under their control (Scott, 1998). Historically, this determination has been based on the conception of the people as an immobile, sedentary, and enclosed body politic within a territory (Isin, 2018). This sedentary bias in conceptions of the national demos, in turn, has constituted various mobile peoples as residual parts such as nomads, migrants, and refugees, itinerants, gypsies, and wanderers who move through (or find themselves in) multiple and intersecting positions across a spectrum (120). Thus, to define who are the people of Europe is to constitute who are the dominant (the people) and the dominated (peoples), which deeply carries nationalist and colonial legacies. For Scott (1998), such domination has consisted of various efforts of statecraft to make subjects sedentary and make them legible as part of a (national) society. Historically, efforts to render people legible required organising a population in ways 'that simplified the classic state functions of taxation, conscription, and prevention of rebellion' (2). Population statistics and censuses thus are part of enacting a duality: Europe both as a population – an object of government and biopolitical interventions that seek to optimise its health, wealth, and economic productivity (Foucault, 2009) – and as a distinct people and 'imagined community' (Anderson, 2006) of shared territory, history, and values.

Data Practices

While we understand and study censuses and official population statistics as technologies of government in the tradition of the bodies of scholarship introduced above, we make two important empirical and conceptual departures that we elaborate in Chapter 2. First, we focus on how population methods and statistics require specific *data practices* to enact populations. From defining, standardising, categorising, cleaning, and editing to inferring, estimating, and harmonising data, population statistics involve numerous data practices that are part of what Law (2004) has defined as 'method assemblages'. Method assemblages consist of technologies, materials, rules, things, concepts, and people, 'a large hinterland of inscription devices and practices' (31) and a wide range of literary and material arrangements (29). For population statistics this includes the standards, routines, materials, and infrastructures of censuses, administrative registers, and social surveys, which shape the data practices that make them up. While many data practices are enrolled in the enactment of populations, some may be more significant, yet all can have effects. Thus, to study the data practices of official population statistics requires attending to the relations between different elements of method assemblages and their variable effects. So, while this book is concerned with making up a European population, its focus is on examples of how this is accomplished by specific data practices across national statistical institutes (NSIs), supranational organisations like Eurostat and UNECE, and private organisations, such as those that reuse sources of big data for statistical purposes. Rather than focusing on the final object, that is, the population that is enacted, circulated, taken up and so on, the book examines the data practices that come to matter for how a European population is made up. A second

and related departure is the focus on how the relations that constitute method assemblages are transnational and include rules, conventions, and standards that exceed and traverse national contexts of individual states within and beyond the EU. Like the expertise and rules that have made up Europe mentioned previously, data practices are situated in, circulate, and help to shape a transnational field of practices, namely the field of statistics, as we explain in more detail in Chapter 2.

Data practices that make up official population statistics and their transnational relations were the focus of the fieldwork and empirical analyses of the ARITHMUS project. Together, six researchers carried out a multi-sited and multi-method collaborative ethnography of the data practices of EU national and international statisticians. From 2015–18, we followed their practices at five NSIs (Estonia, Finland, the Netherlands, Turkey, and United Kingdom). Four EU NSIs were selected not to study them as best examples but to investigate each as an instance of specific issues that concern many such as innovation labs and experiments (e.g., Netherlands), digital censuses (e.g., UK and Estonia), questionnaire-based censuses (e.g., UK), register-based censuses (e.g., Finland), and experiments with big data (e.g., Estonia, Netherlands). Additionally, the NSI of a non-EU country but that had applied for membership was included because they participate in EU statistical programmes as part of their candidacy (Turkey).[3] We also followed two international organisations (Eurostat, the statistical agency for the European Union, and UNECE, the statistical division of the United Nations Economic Commission for Europe) but our research also led to us following others (e.g., Expert Group on Refugee and Internally Displaced Population Statistics (EGRIS) and International Organisation for Migration (IOM)). Of note and importance to our understanding of a transnational field is that these

institutions work closely with each other and often cooperate on the development and adoption of standards, definitions, and methods. This is reflected in our decision to study these institutions through a collaborative ethnography, which we have documented elsewhere (Scheel et al., 2016, 2020). The main reasons follow from our conception of data practices and their transnational relations.

We first recognised that to analyse data practices requires following the everyday activities, conversations, meetings, negotiations, technical work, and so on of statisticians. This includes not only tracing discourses, but also material and technological work such as data cleaning, modelling, and visualisation. Hence, following data practices also means to follow relations to technologies and to trace technological forms as they are produced, exchanged, and travel to different sites both within and beyond European statistical institutes. This is related to a second reason. Practices do not happen in isolation but are part of forces and dynamics that cut across national and international statistical organisations (Scheel et al., 2020). This calls for moving beyond nationally bounded case studies, a research practice that has been problematised as 'methodological nationalism' (Wimmer and Glick Schiller, 2002). We thus conceive of data practices as part of a transnational field of statistics where scales of the local, the national, and the international overlap and intersect in practices that enact neither a 'national' nor 'European' population and which perpetuate nationalist and colonial legacies. To follow and trace such relations and dispersed practices required defining a corresponding collaborative method which we came to call 'transversal' (Scheel et al., 2016). Building on the initial project formulation, it is in these two senses that we interpret how data practices involve 'peopling' and not simply reflecting who are the people of Europe.

The following chapters analyse in detail some of this ethnographic fieldwork with a focus on how enacting the population and people of Europe requires standardising, harmonising, and assembling data that have been produced by NSIs via multifarious practices that make up method assemblages of which censuses are a part. It is a partial and selective account of this fieldwork and we do not attempt to equally cover all of the sites that we followed and noted above. Critically, this fieldwork took place during initial debates about and experiments with new digital technologies and sources of big data and what they mean for statistical methods that we noted previously.

Our collaboration involved not only tracing and documenting these changes and sharing fieldwork material and notes. It extended to the analysis and writing of a working paper and an article that elaborated our method (Scheel et al., 2016, 2020), which have guided the conception of data practices and analyses in the chapters of this book. Consequently, this book can be considered as a hybrid: it is both a research monograph and an edited collection. Rather than consisting of a series of different positions on data practices, it is a single intervention about the role of data practices in the making up of the population and people of Europe.

Contribution

This book contributes to scholarship on official population statistics and the making up of Europe as well as related theoretical debates on biopolitics, the authority of numbers, the politics of method, the performativity of categories, and the multiple entanglements between the production of knowledge and practices of governing, statecraft, and nation-building. It is distinctive in connecting contemporary political and philosophical debates about Europe and European identity to

the practical problem of knowing Europe as a population and a people. It does this through a focus on the role of censuses and more generally population statistics, which are amongst myriad standardised practices and institutions such as laws, regulations, and maps that have historically been part of the formation of nation-states. Critically, it does this by considering how new digital technologies and big data are potentially changing statistical regimes and what this might mean for the political legitimacy of the EU.

The contribution of the book also resides in its timing, which coincides with the 2020–21 round of censuses. In general, censuses are conducted every ten years and governed by national laws and guided by international agreements, protocols, and guidelines or regulations in the case of the EU. During 2020–21, most NSIs around the world will have conducted national censuses using various methods from questionnaire-based to register-based censuses and many will have conducted online censuses. In the context of the EU, the 2020–21 census round will involve the further implementation of an intense programme of harmonising population data across NSIs with the objective to provide a singular account of the European population (Eurostat, 2019). These efforts confront policy demands to 'do more with less' which feed imperatives to innovate, and in turn, contribute to methodological diversification across NSIs. Hence, the lead up to the 2020–21 round of censuses was a unique moment to study practical and political struggles over the making of official statistics and what this means for enacting the population and people of Europe. It enabled following, observing, and detailing the usually out-of-public-sight data practices of statisticians: their debates, struggles, tensions, discourses, techniques, material devices, logics, rationalities, values, assumptions and so on.

Through a focus on the data practices of statisticians, this book also locates the 'politics of numbers' not primarily in political debates and institutions of government where statistical figures are invoked as evidence to back claims and promote political agendas where politics happen *after* numbers have been produced and circulated. Rather, the chapters of this book show that politics also happen *in* and *through* data practices that produce and circulate these numbers and contribute to enacting the realities to which they refer. The connection between the politics of numbers and data practices was powerfully revealed in the controversy over the addition of a question on citizenship status to the US 2020 census questionnaire. In brief, the Trump administration tried to include the following question on the citizenship status of any household member: 'Is this person a citizen of the United States?' The Trump administration argued that the question was needed to provide the Justice department with more accurate data for implementing the *Voting Rights Act*, ostensibly to protect ethnic minority voters (BBC, 2019). Statisticians of the US census bureau had, however, conducted a study in 2018 which concluded that 'inclusion of a citizenship question will likely suppress response rates in households with immigrants and minority groups' as the latter may fear that data will be shared with authorities enforcing deportation (ibid.). Thus, adding the citizenship question could result in up to four million people – mostly African-Americans and Latinos, that is, the very people whose voting rights the citizenship question is meant to protect – could go uncounted (Urban Institute, 2018). Moreover, critics maintained that the objective behind adding the question was to suppress response rates among mostly democratic-voting minorities in order to allow the Republican Trump administration to redraw electoral boundaries in their favour. This would not only affect future election-results. The

undercount of ethnic minorities would also deprive their communities of public funds for schools, roads and other public services (BBC, 2019).[4]

The political uses to which population statistics may be put also has a long history in Europe. Most famously is the use of censuses and other official statistics by Nazi-Germany to organise the mass murder of Jews, Sinti and Roma, and other 'undesirables' (Hannah, 2012). Such uses were recounted in the boycott of German censuses in the 1980s which was also fuelled by concerns about the sharing of census data facilitated by digital technologies. The performative effects and political implications of censuses have also been palpable in contexts where tensions between ethnic groups are rife, such as in the Western Balkans (Hoh, 2018). In Bosnia-Hercegovina, the delayed publication of census results in 2016 prompted a political crisis because Bosnian-Serb politicians rejected the results which reported that the number of ethnic Serbs living in the country had declined even more than the number of other ethnic minorities (Agence France-Presse, 2016). Yet, that different census methods enact contested versions of a population as the chapters of this book argue is not unique to these contexts but rather a condition of all population statistics that involve 'micro-politics of method' (Scheel, 2020).

Moreover, this book contributes to debates on the so-called migration crisis, which erupted in 2015 and during the time when ARITHMUS researchers were doing fieldwork. Consequently, migration empirically became a major matter of concern. Heated political debates affected, and in turn, were taken up in the transnational field of statistics through calls for more timely, detailed, and reliable data and statistics on migration. At the same time, these calls revealed how migration is politically one of the most difficult statistical categories to define and measure. Yet, amongst other things,

how it is defined and measured are consequential for determining who belongs to and should be counted as part of a population. This especially holds for the European project which enables a form of citizenship strongly intertwined with freedom of movement: the right of EU citizens to settle and work in other member states as granted under the Maastricht Treaty. In consideration of the empirical and political import of migration we therefore analytically came to focus on different categories of 'migrants' or mobile subjects in the making up of a European population and people. This focus echoes de Genova's (2016: 76) observation that what is disputed in today's heated debates on the 'migration question' are, first and foremost, competing notions of 'Europe and Europeanness.'

The book is moreover relevant to another significant political event that took place during the fieldwork of ARITHMUS: the vote of a slim majority of the UK electorate to leave the EU. However, contrary to assumptions that a clear break from the EU is possible as suggested by the UK government's political programme of 'Brexit,' this book shows how statistical practices, like so many other practices, are part of a transnational field and connect Europe in ways beyond formal political arrangements, treaties, laws, and unions. While Shore (2000) examined the complexities and difficulties of European integration, this book highlights the complexities and difficulties of its disassembly when a member state seeks to exit from the ties that bind it. While the Office for National Statistics (ONS), the NSI of the UK, will no longer be subject to EU regulations on the production of official population statistics, it will remain a member of the UNECE. As such, the ONS will be subject to UNECE guidelines and conventions and the imperative that the statistics it produces are internationally comparable and recognised. Furthermore, those guidelines, which concern statistical methods, categories, definitions, standards and so on,

closely correspond to EU regulations as they are a product of cooperation between Eurostat and the statistical division of UNECE. But perhaps more significantly, ONS statisticians will continue to engage with and perform within the transnational field of statistics that shapes conventions, innovations, practices, and methods of national statisticians. Thus, the UK will remain entangled in statistical laws, rules, and conventions and related professional fields. So, while integration has never been fully achieved as argued by Shore, so too will it likely be for any form of disassembly.

Last, at the time of the final writing and editing of this book, the outbreak of COVID-19 – a novel, highly contagious coronavirus – was declared by the World Health Organization in March 2020 as a global pandemic with no signs of its imminent abatement. The pandemic revealed the fragility and inadequacy of government services, most significantly those of health and social care which, in many countries, were weakened by years of underfunding and lack of state investment. It also revealed the inadequacy of government statistical services to produce timely and relevant data: 'For official statistics it pointed among others to the weak coordination between domains of involved statistics, to the lack of timeliness of official data, to incomplete and erroneous statistics and to the opportunity this crisis gave to fake and purposely misleading statistics popping up' (Everaers, 2020: 243).

The search for data and statistical alternatives resulted in heightened attention to digital technologies and big data sources, especially those of private corporations to track population movements that are more detailed and timelier to know the effects of the pandemic and inform policy responses. For example, the ONS started producing 'early experimental data on the impact of the coronavirus (COVID-19) on the UK economy and society' through the development of faster indicators

based on rapid response surveys, novel data sources such as online job adverts, and price changes and experimental methods (ONS, 2020). For the editor of the leading international journal of statistics, the 'new normal' of the pandemic will lead to changes in the 'whole set of social, economic and business statistics' (Everaers, 2020: 244). The new normal may well include the very conduct of censuses: the outbreak of the pandemic coincided with the 2020–21 censuses which led to adjustments or delays in some countries due to their reliance on field- and questionnaire-based methods. The import for this book is that the new normal will not simply herald innovations in methods but also have consequences more broadly for how populations are known and governed (Isin and Ruppert 2020). In this sense the pandemic returns this chapter to its opening reflections on the implications of digital and technological changes for the future of official population statistics and making up a European population and people. It is this issue that begins Chapter 2, which develops a conception of data practices through which the transnational field of statistics is being transformed.

Outline of the Chapters

Chapter 2, 'Data Practices', develops an understanding of data practices as empirical objects and a conceptual register for analysing the activities of practitioners by drawing on theories that are generally referred to as part of the 'practice turn' in contemporary social sciences. It adopts a conception of data practices as 'embodied, materially mediated arrays of human activity centrally organized around shared practical understanding' and 'occur within and are aspects or components of the field of practices' (Schatzki, 2001: 10–11). Building on this, the chapter adopts five theoretical commitments and related analytical sensitivities for analysing data practices. It

then introduces the book's focus on data practices involved in classifying and encoding people into categories which then contribute to making up a European population and people.

Chapter 3, 'Usual Residents: Defining and Deriving', explores how the international harmonised definition of 'usually resident' conceives of the European population as sedentary and relatively fixed to national locations. The chapter attends to two problems that the harmonised definition confronts. First, it is often at odds with multiple modes of mobility in the EU – what the chapter refers to as the complexity of mobility – which challenge the relevance of the category. Second, the harmonised definition confronts pre-existing national definitions, rules, technologies, priorities, and histories – what the chapter refers to as the complexity of methods. How statisticians address these problems is explored through two data practices they deploy to classify and encode subjects as 'usually resident' and which are necessary to sustain the definition: *defining special cases* and *deriving usual residents*.

Chapter 4, 'Refugees and Homeless People: Coordinating and Narrating', examines two groups identified as special cases in the definition of the usually resident population, which are considered 'hard-to-count'. It argues that the category of refugee conflates people who occupy different legal statuses and the category of the homeless includes people living under myriad conditions. This contributes to national differences in methods for enumerating these populations and stands in the way of achieving internationally comparable data. The chapter argues that this is resolved by producing 'good enough' data through two data practices that enact refugees and homeless people as 'excess populations' that overflow the 'usual': data practices that *coordinate* international numbers of refugees across the world; and data practices that *narrate* national numbers of homeless people across the EU.

Chapter 5, 'Migrants: Omitting and Recalibrating,' shows that the enactment of mobile populations in Europe is intertwined with the production of non-knowledge. It attends to two data practices – *omitting* and *recalibrating* – to illustrate how the enactment of migration as a coherent, precisely measurable reality hinges on the production of non-knowledge about the known limits of quantifying migration. The chapter does this through a study of the 'Global Migration Flows Interactive App' (GMFIA) which was hosted by the International Organization of Migration (IOM) until it was deactivated in 2019. The GMFIA is interpreted as an example of how actors in the field of migration management mobilise seemingly precise figures from the field of statistics about stocks and flows of migrants – often by assembling them into interactive visualisations – in order to perform themselves as knowledgeable, competent actors capable of 'managing' migration according to predefined policy objectives.

Chapter 6, 'Foreigners: Inferring and Assigning' attends to the performativity of statistical categories to highlight their role in the enactment of the people of Europe. It does this by analysing two new statistical identity categories introduced in Estonia and the Netherlands in the context of register-based population statistics. One important implication of register-based methods is that people do not allocate themselves to identity categories through practices of self-identification. Hence, the chapter attends to two data practices – *inferring* and *assigning* – that are used by statisticians to allocate individual subjects to the statistical identity categories of the 'third generation of the foreign-origin population' (Estonia) and the 'Caribbean Netherlands.' The analysis shows that statistical identity categories enact more than the groups to which they refer. They also enact national identities and notions of national belonging of majoritarian groups in the host countries in ways that perpetuate and carry colonial legacies.

The next two chapters step back from how data practices enact kinds of people that constitute Europe to consider two subject positions that data practices also produce and require: the data subject and the statistician subject. Chapter 7, 'Data Subjects: Calibrating and Sieving', explores how data practices that make up a population and people involve 'forces of subjectivation' that differently configure the capacities of data subjects to intervene, challenge, and influence how they are classified and encoded. It takes up this conception to then explore how subjectivation plays out differently in two experiments with digital technologies that are offered as solutions to traditional paper questionnaire-based censuses where data subjects are problematised for not revealing themselves truthfully. The first involves experiments with Twitter data to know student migration in the UK through the data practice of *sieving* tweets; and the second the design and development of digital censuses in the UK and elsewhere through the data practice of *calibrating* responses. The chapter shows how data subjects do not pre-exist but come into being through data practices that configure the relations, interactions, and dynamics between human and technological actors.

Chapter 8, 'Statistician Subjects: Differentiating and Defending', considers how the statistician subject is also shaped through the valuing and performing of data practices, which in turn come to influence what constitutes the profession of national statistician. The chapter argues that this happens through 'professionalising practices' such as job interviews, innovation events, and professional conferences. It argues that it is through these practices that the skills, capacities, mindsets, and ethical positions of the profession of national statistician are being repositioned in relation to a new faction in the transnational field of statistics, that of data

scientists. It examines how this is happening relationally: by valuing and adopting some of the skills and dispositions of data scientists – described as *entrepreneurial* – and by defending and differentiating those of national statisticians – described as *public service*.

Finally, Chapter 9, 'The Politics of Data Practices', provides a brief overview of the foregoing chapters to then focus on key political questions that cut across them to emphasise how politics happen in and through data practices but also how data practices are irreducibly political. In sum, the issues concern: (1) the sedentary bias of population statistics; (2) the double edge of enumeration; (3) the production of non-knowledge and the performativity of what is absent; (4) the politics of knowledge and the performativity of what is present in categories; and (5) the politics of method in and of data practices. The chapter then concludes by considering what these issues mean for official population statistics, academic research, and citizen data rights in relation to making up a European population and people.

References

Agence France-Presse (2016) Bosnia-Herzegovina Has Lost a Fifth of its Pre-war Population, Census Shows. Available at: https://bityl.co/519Z (accessed 23 January 2020).

Anderson B (2006) *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Revised and extended. London: Verso.

Balibar É (2003) *We, the People of Europe?: Reflections on Transnational Citizenship*. Princeton, NJ: Princeton University Press.

Barry A and Walters W (2003) From EURATOM to 'Complex Systems': Technology and European Government. *Alternatives: Global, Local, Political* 28: 305–329.

BBC (2019) US Supreme Court Blocks Census Citizenship Question for Now. London. Available at: https://bityl.co/51Ba (accessed 23 January 2020).

Bellamy R (2008) Evaluating Union Citizenship: Belonging, Rights and Participation Within the EU. *Citizenship Studies* 12(6): 597–611.

Bellamy R, Castiglione D, and Shaw J (eds) (2006) *Making European Citizens: Civic Inclusion in a Transnational Context*. Basingstoke: Palgrave Macmillan.

Best H (2009) History Matters: Dimensions and Determinants of National Identities Among European Populations and Elites. *Europe-Asia Studies* 61(6): 921–941.

Bourdieu P (1979) Public Opinion Does Not Exist. In: Mattelart A and Siegelaub S (eds) *Communication and Class Struggle*. New York: International General, pp. 124–129.

de Genova NP (2016) The European Question. Migration, Race, and Post-coloniality in Europe. *Social Text* 34(3): 75–102.

Desrosières A (1998) *The Politics of Large Numbers: A History of Statistical Reasoning* (ed. RD Whitley). Cambridge, MA and London: Harvard University Press.

Eurostat (2016) *A Vision for Census Statistics After 2021*. DSS/2016/Oct/3.2, Meeting of the European Directors of Social Statistics. Luxembourg: European Commission.

Eurostat (2019) *EU Legislation on the 2021 Population and Housing Censuses – Explanatory Notes*. Available at: https://bityl.co/51ER (accessed 13 January 2020).

Everaers P (2020) Editorial. *Statistical Journal of the IAOS* 36(2): 243–251.

Foucault M (2003) *'Society Must Be Defended': Lectures at the Collège de France 1975–76*. New York: Picador.

Foucault M (2009) *Security, Territory, Population: Lectures at the Collège de France 1977–1978* (eds M Senellart, F Ewald, A Fontana, et al.; trans. G Burchell). London: Palgrave Macmillan.

Gawthorpe A (2020) Trump is Ending the Census Count Early: It's a Feat of Profound Political Cynicism. *The Guardian*. Available at: https://bityl.co/40ql (accessed 18 October 2020).

Habermas J (2006) Is the Development of a European Identity Necessary, and Is it Possible? In: *The Divided West*. Cambridge: Polity, pp. 67–82.

Hacking I (1990) *The Taming of Chance*. New York: Cambridge University Press.

Hacking I (1999) Making up People. In: Biagioli M (ed.) *The Science Studies Reader*. London: Routledge, pp. 161–171.

Hacking I (2002) *Historical Ontology*. Cambridge, MA: Harvard University Press.

Hannah M (2012) Interview: The Boycott of the German Census in the 1980s. *Exploring Geopolitics Blog*. Available at: https://bityl.co/51C8 (accessed 30 December 2020).

Hoh A-L (2018) *Counting for EU Enlargement? Census-taking in Croatia, Bosnia and Macedonia*. PhD Thesis. Maastricht: University of Maastricht.

Irvine J, Miles I and Evans J (eds) (1979) *Demystifying Social Statistics*. London: Pluto Press.

Isin E (2018) Mobile Peoples: Transversal Configurations. *Social Inclusion* 6(1): 115–123.

Isin E and Ruppert E (2020) The Birth of Sensory Power: How a Pandemic Made it Visible. *Big Data & Society* 7(2): 1–15.

Isin E and Saward M (2013) *Enacting European Citizenship*. Cambridge: Cambridge University Press.

Kaiser W and Schot J (2014) *Writing the Rules for Europe: Experts, Cartels, International Organisations*. Making Europe. Series editors Schot J and Scranton P. London: Palgrave Macmillan.

Kertzer DI and Arel D (2002) *Census and Identity: The Politics of Race, Ethnicity, and Language in National Censuses*. New Perspectives on Anthropological and Social Demography. Cambridge: Cambridge University Press.

Kohlrausch M and Trischler H (2014) *Building Europe on Expertise: Innovators, Organizers, Networkers*. Making Europe. London: Palgrave Macmillan.

Konstadinidis SV (1999) *A People's Europe: Turning a Concept into Content*. EC/International Law Forum S. Aldershot: Ashgate.

Law J (2004) *After Method: Mess in Social Science Research*. London and New York: Routledge.

Law J (2009) Seeing Like a Survey. *Cultural Sociology* 3(2): 239–256.

Lie J (2004) *Modern Peoplehood*. Cambridge, MA: Harvard University Press.

Nobles M (2000) *Shades of Citizenship: Race and the Census in Modern Politics*. Stanford, CA: Stanford University Press.

ONS (2020) Coronavirus and the Latest Indicators for the UK Economy and Society. Available at: https://bityl.co/3wcN (accessed 14 October 2020).

Osborne T and Rose N (1999) Do the Social Sciences Create Phenomena?: The Example of Public Opinion Research. *The British Journal of Sociology* 50(3): 367–396.

Patriarca S (1996) *Numbers and Nationhood: Writing Statistics in Nineteenth-Century Italy*. Cambridge: Cambridge University Press.

Porter TM (1986) *The Rise of Statistical Thinking, 1820–1900*. Princeton, NJ: Princeton University Press.

Radermacher WJ (2020) *Official Statistics 4.0: Verified Facts for People in the 21st Century*. Cham, Switzerland: Springer International Publishing.

Savage M (2010) *Identities and Social Change in Britain since 1940: The Politics of Method*. Oxford: Oxford University Press.

Schatzki TR (2001) Introduction: Practice Theory. In: Schatzki TR, Knorr-Cetina KD, and Savigny E von (eds) *The Practice Turn in Contemporary Theory*. London: Routledge, pp. 10–23.

Scheel S (2020) Biopolitical Bordering: Enacting Populations as Intelligible Objects of Government. *European Journal of Social Theory* 23(4): 571–590.

Scheel S, Grommé F, Ruppert E, et al. (2016) *Transcending Methodological Nationalism through Transversal Methods? On the Stakes and Challenges of Collaboration*. ARITHMUS Working Paper 1. Available at: www.arithmus.eu (accessed 21 May 2020).

Scheel S, Grommé F, Ruppert E, et al. (2020) Doing a Transversal Method: Developing an Ethics of Care in a Collaborative Research Project. *Global Networks* 20(3): 522–543.

Schmidt D (2005) *Statistik Und Staatlichkeit*. Wiesbaden: Springer VS.

Scott JC (1998) *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven and London: Yale University Press.

Shore C (2000) *Building Europe: The Cultural Politics of European Integration*. London and New York: Routledge.

UNECE (2018) *Recommendations for Promoting, Measuring and Communicating the Value of Official Statistics*. New York and Geneva: United Nations.

Urban Institute (2018) In: *Five Reasons the Addition of a Citizenship Question to the 2020 Census is Detrimental to its Goal*. Available at: https://bityl.co/3wIz (accessed 14 October 2020).

Wimmer A and Glick Schiller N (2002) Methodological Nationalism and Beyond: Nation-State Building, Migration and the Social Sciences. *Global Networks* 2(4): 301–334.

2

Data Practices

Evelyn Ruppert and Stephan Scheel

What Are Data Practices?

Chapter 1 outlined the fundamental challenges of digital and technological changes to produce official population statistics and what they mean for making up a European population and people. It argued that these changes are stimulating – and perhaps even driving – experiments with new sources of data, the diversification of methods and relatedly the role of NSIs and their relations to other data producers. That is the context for this chapter, which we consider in relation to two key principles. First, while much can be learned from describing these broad contours of change in generalising statements, the book approaches change as an object and outcome of political struggles over the power, authority, and legitimacy to name and know populations. Those struggles happen not only through debates and political programmes, but through specific social, cultural, and material data practices. Second, we consider how the data practices through which change is happening in the transnational field of statistics are one small part of what is happening more broadly in relation to the proliferation of big data in contemporary societies as a consequence of the 'datafication' of everything (Mayer-Schönberger and Cukier, 2013; van Dijck, 2014).

According to advocates, big data produced by the digital interactions and transactions of people with various government, commercial and social platforms, devices and apps make it possible to measure, monitor, track, and analyse myriad aspects of social lives in near real-time (Ruppert, 2011). Besides

providing more timely information, the big data they produce are promoted as a means of capturing and reflecting the conduct of entire populations with unprecedented efficiency at a reduced cost (Kitchin, 2015). Such attributions to big data give rise to a new naïve empiricism and set of interrelated assumptions that 'data can speak for themselves' and can capture an entire domain 'free of human bias or framing' (Kitchin, 2014a: 5). In this way the 'big data revolution' revives a new version of what Labbé (2000) refers to as 'statistical realism', that is, the belief that statistical data are collected about already existing realities and reflect, measure, and represent those realities more or less accurately. Even critics of big data and related processes of datafication often imply such an understanding. For example, critical studies of the datafication of border and migration management suggest that migration flows are increasingly 'datafied' by 'an ever expanding network of surveillance systems and databases aimed at visualising, registering, mapping, monitoring, and profiling mobile (sub)populations' (Broeders and Dijstelbloem, 2016: 243).

In contrast, a growing scholarship critiques such assumptions or makes more explicit that data is not 'raw' or a mere reflection (Gitelman, 2013) because 'data do not exist independently of the ideas, instruments, practices, contexts and knowledges used to generate, process and analyse them' (Kitchin, 2014b: 2). This is why data are not neutral representations of external realities but carry particular institutional agendas, political and economic interests, cultural norms, preferences, and tacit assumptions. In other words, data are never 'raw', but always 'cooked' according to particular 'recipes' (Gitelman, 2013).

It is with the latter understanding that many researchers have considered how specific data practices such as selecting, formatting, editing, storing, cleaning, and analysis are involved

in producing data. For instance, Leonelli (2016) details the data practices of curation in the biological sciences through which billions of data are centrally brought together for scientific research and how this involves '*packaging procedures* for data, which include data selection, formatting, standardization, and classification, as well as the development of methods for retrieval, analysis, visualization, and quality control' (italics in original, 16). Garnett (2016) ethnographically traces data practices that make air pollution data and how they actively shape 'what constitutes air, and how air is experienced and engaged with' and the 'ways in which environmental data gain scientific and political affordance' (2). Gabrys, Pritchard and Barratt (2016) attend to the data practices of citizens who deploy a range of air pollution monitoring technologies and techniques to not just generate amateur accounts, but provoke the political possibilities of data. An edited collection by Knox and Nafus (2018) includes a range of ethnographic studies of data practices that deploy a variety of concepts to understand how digitally collected data are one of many ways of knowing social lives. There are numerous other accounts documenting the manifold practices through which data are produced, repurposed and processed such as categorising, sensing and cleaning (e.g. Ardittis and Laczko, 2017; Edwards, 2010; Leahey, 2008).

However, while empirically rich and informative, most of these works do not make explicit a theory or conception of data practices.[1] Rather, they approach practices in a way that is more generally critiqued by Gad and Jensen (2014): as an empirical object that captures a variety of activities of practitioners or as a conceptual register where the scope and meaning of practices 'is rarely explicated' (699). Consequently, 'practice approaches are slippery: they can slide easily between empirical and conceptual registers, without at any point losing their

aura of common sense' (ibid.). Practices are, in other words, treated as simply what practitioners do and where theoretical and analytic choices for interpreting those practices are seldom made explicit.

We take up this distinction to develop an understanding of data practices as both empirical objects and conceptual registers to analyse the activities of practitioners. It is an understanding that we initially developed to frame articles in two journal special issues we edited.[2] The special issues include articles by ARITHMUS and other academic researchers that examine data practices involved in governing education, health, citizenship, residence, social policy, and migration through which populations of Europe are enacted. Our initial understanding, and that which we elaborate here, draws on theories that are part of what is more generally referred to as the 'practice turn' in contemporary social sciences (Bueger and Gardinger, 2015; Gad and Jensen, 2014; Hui, Schatzki and Shove, 2017; Reckwitz, 2002). For these and other authors, the practice turn is marked by a shift from interpreting social phenomena as 'structures,' 'systems,' 'life worlds,' 'events,' and mere 'actions' of individual agents to that of socially, culturally, and materially embedded 'practices.' Considering the work of Foucault and Taylor to that of Wittgenstein and recognising that there are many theories and no unified approach, Schatzki's (2001) significant contribution brings together work of leading researchers in STS and the social sciences. That work includes theories that understand how practices involve following rules that encompass patterns of behaviour and normative verbal 'accounting' (Bloor, 2001); the role of tacit knowledge in the study of scientific practices (Collins, 2001) and convergences in what people learn and share in tacit rules (Turner, 2001); the mutual constitution of material and human agency (Pickering, 2001); the 'relational dynamics' that link subjects and objects (Knorr Cetina, 2001);

and the ordering of the field of practices through discourses, activity patterns, and social relations (Swidler, 2001). Drawing on this work, Schatzki (2001) offers a core conception that seeks to capture this variety and scope of contemporary practice theories: 'practices are embodied, materially mediated arrays of human activity centrally organized around shared practical understanding' and 'occur within and are aspects or components of the field of practices' (10–11). Importantly, Schatzki emphasises that the 'linchpin' is to 'treat the field of practices as the place to study the nature and transformation of their subject matter' (11). Likewise, Reckwitz (2002: 249) defines a practice as 'a routinized type of behaviour which consists of several elements, interconnected to one another: forms of bodily activities, forms of mental activities, "things" and their use, a background knowledge in the form of understanding, know-how, states of emotion and motivational knowledge'. Reckwitz emphasises that performing a practice depends on the interplay of all these elements and can, consequently, not be reduced to any of them (on this point: Bueger and Gardinger, 2015: 451).

These understandings highlight three important points for the core conception of practices informing this book. First, practices cannot be reduced to routinised techniques or technical operations. Rather, practices are *activities* such as articulating discourses, making drawings, and creating designs that are performed by humans and in relation to various materials and other actors that are part of a field of practices. To analyse practices thus calls for different methods of following activities such as close ethnographies of everyday work activities involving relations between humans and materials; identifying and tracking relations between actors; tracing discourses as they circulate in documents, reports, and meetings; and identifying the field of practice of which activities are a part and how those activities are shaped by and shape that field.

Second, practices de-centre the notion of 'human agency as a highly reflexive and formally rational enterprise' (Reckwitz, 2002: 258). Instead, practice theories attend to the beliefs and values of actors, as well as available material resources and the external environment involved in a particular doing. In this way practice theories seek a 'unified account of knowing and doing' (Bueger and Gardinger, 2015: 453). Rather than considering practices as ex-post outcomes of rational-choice calculations or coherent norm-oriented planning, they consider the role of tacit knowledge of actors and how that becomes ingrained in practices. Consequently, the study of data practices cannot be reduced to the investigation of particular actions and operations, but also has to consider the discourses, knowledge regimes, legal norms, materialities, and technical affordances that shape and inform these practices. Finally, practice theories embrace – similar to socio-material approaches as they have been developed in STS – a performative understanding of phenomena in which realities only exist as long as they 'are enacted, enacted again and enacted yet again' (Law, 2008: 635). Following this understanding, it is through the reiteration of a range of practices which establish and reconfigure relations between actors, objects, bodies of knowledge, discourses, legal norms, material artefacts, and so forth that realities are accomplished.

We adopt the core conception of practices cited previously and draw on these three points to develop our understanding of *data practices* as empirical objects and a conceptual register for analysing the activities of statisticians. More specifically, we adopt five theoretical commitments and related analytical sensitivities.[3] In brief, we conceive of data practices as (1) *sociotechnical* in that they involve relations between humans, materials, infrastructures, and technologies; (2) *situated* in and produced by sets of relations; (3) *performed* by

actors as stakes in struggles over authority and power within specific professional fields of practice; (4) *contingent* in that they do not have a 'prior and determinate form' (Law, 2004: 38) but involve practical adjustments to address complex and changing conditions; and (5) enrolled in the *enactment* of their object. Below, we elaborate on each of these theoretical commitments, which inform and guide the empirical analyses in the chapters that follow.

To begin with, data practices are *sociotechnical* in that they involve various relations between humans (practitioners, policymakers, regulators, subjects, etc.) and technologies (materials, infrastructures, devices, rules, standards, protocols, etc.). Rather than contained or bounded, those arrangements and relations occur within a 'hinterland of pre-existing social and material realities' that constitute 'method assemblages' (Law 2004, 13). That is, data practices are configured by and are performed in relation to things such as existing rules and infrastructures, which they both depend upon but also affect. Rather than separate, 'human and material agency are reciprocally and emergently intertwined' such that they 'are mangled in practice, meaning emergently transformed and delineated in the dialectic of resistance and accommodation' (Pickering, 1995: 21–23). In other words, they are enabled and constrained by their relations to other elements of the method assemblages which they also affect and form a part of.

While data practices are part of method assemblages – of technologies, materials, rules, things, concepts, and people – they are also 'bound to a specific site' (Mol 2002, 55) and located in and produced by sets of *situated* relations (Haraway, 1988; Law, 2004; Mol, 2002; Suchman, 2007). Sets of relations consist of partial connections with the various elements that make up method assemblages (Law, 2004). Sites and situations can include relevant and specific histories and legacies of past data

practices to the different and particular technologies and relations between actors that get assembled at different sites such as between national statistical offices and administrative departments. Which relations are relevant cannot be defined a priori, but only through empirically following and tracing practices and the connections they establish.

As already suggested, that data practices are situated and part of method assemblages does not mean that they are determined. On the one hand, data practices are part of knowledge regimes that contain recurring patterns, regularities, logics, strategies, self-evidence and rationalities 'where what is said and what is done, rules imposed and reasons given, the planned and the taken-for-granted meet and interconnect' (Foucault, 2000: 225). That is, numerous regimes such as statistical rules and conventions configure what and how data is produced by practices. However, data practices involve a 'more or less messy set of practical *contingencies*' (Law, 2004: 13) that are complex and variable. That is, while regimes such as official or scientific rules and discourses configure practices, what is done 'takes work and effort' and is an accomplishment that does not have a 'prior and determinate form of its own' (38).

Data practices are *performed* by actors and function as stakes in competitive struggles over authority, influence, and resources within specific fields of practice. As Schatzki (2001) notes, practice approaches develop an account that treats 'the field of practices as the place to study the nature and transformation of their subject matter' (11). In this regard, we take up Bourdieu's (1990) understanding of fields where each actor's position and authority are configured by their relative possession of different types of recognised capital (cultural, economic, social, and symbolic) including their embodied forms (perceptions, know-how, skills, experience, judgements, tacit knowledge). These constitute both objective (positions) and

subjective (dispositions) forms of knowledge that Bourdieu argues are part of the logic of practices (Bourdieu, 1990; Bourdieu and Wacquant, 1992).

Our final theoretical commitment is to the concept of enactment introduced in Chapter 1: data practices contribute to making up the very objects and subjects that they seek to represent. In other words, data practices are not only performed but also performative in the sense that they help to enact – that is make up – the very realities they ostensibly only describe. Hence, rather than 'constructing' an object or 'reflecting' an already existing reality, the concept of enactment specifies that realities are made up and reproduced by data practices.

In this context, a brief clarification regarding terminology is needed. To avoid confusion with the stance that data practices are performed by certain actors such as statisticians, data scientists, or enumerators in the field of statistics, we use the word *enact* whenever we write about, describe, or analyse the performative effects of data practices. While work on enactment has certainly been influenced by that on performativity, researchers who adopt enactment do so to avoid connotations carried by the notion of performativity. In this regard, Mol notes, for instance, that the term is too closely related to the word performance which, while carrying some useful meanings such as that of a script being performed by certain actors, it is also potentially misleading as it 'may be taken to suggest that there is a backstage, where the real reality is hiding' (Mol, 2002: 32). To avoid such Goffmanian associations of a frontstage and a backstage, Mol suggests using a word 'without too much academic history' that does not carry such terminological baggage, namely the word enact. The term is also particularly useful in that studying how data practices enact realities means, according to Mol (2002: 33)

to attend to the activities and 'techniques that make things visible, audible, tangible, knowable'.

Moreover, enactment avoids connotations that are carried by terms such as 'constitution' – which suggests a one-time creational act – or 'construction' which suggests not only stability and fixity (Ruppert, 2011: 223) but also that materials are assembled and put together according to a predefined plan by wilful human subjects (Mol, 2002: 32). Speaking of enactment allows, in contrast, to highlight that the activities involved in the enactment of realities, as well as their effects, are not fully controlled by the actors performing them. The reason is, as we have noted above, that data practices are enabled by and part of complex and always shifting assemblages comprising a multitude of human and non-humans.

What follows from this is that making up populations is a volatile and contingent accomplishment that hinges on mutable data practices whose operation and maintenance requires continuous work (Law, 2008; Mol, 2002; Ruppert, 2011). This understanding is the basis for what has come to be known as 'ontological politics'. In brief, practices involve normative values, political agendas, and tacit assumptions that bring one reality into being and not others (Mol, 2002). In relation to data practices, they are political insofar as they enact and sustain certain versions of the real while marginalising or even precluding other possible versions from emerging. To say so is also to acknowledge that the realities produced at the same time exceed the will to power whereby practices come to attain constitutive powers. That is, as Hacking (1999) argues in relation to the subversive effects of categories, practices can produce realities that are 'incidental' and 'unintentional' and which are effects of the relations and practices that make up method assemblages (Law, 2012: 156). What is enacted is thus neither controlled and determined by individual human

actors, nor are enactments reducible to predefined outcomes of plans. By attending to the effects of data practices thus means to engage in a 'politics of the real' by paying attention to how practices shape and reconfigure realities such as populations.

This conception of populations as enacted by data practices challenges the overly simplistic epistemological register of statistical realism (Labbé, 2000). The latter postulates that statistics only measure, account for, and describe realities that already exist. The concept of enactment concerns, in contrast, the (onto-)political qualities of statistics. This constitutes an 'empirical ontology' (Law and Lien, 2012) whereby data practices help to make up and reproduce the very objects to which they refer such as particular versions of the population and people of Europe.

These five theoretical commitments and related analytical sensitivities of data practices are not exhaustive, nor are they all explicitly addressed in all chapters of this book. Rather, each chapter first identifies specific data practices that have been ethnographically observed and documented, and then interprets them by drawing on the theoretical commitments and analytical sensitivities that are most relevant. Informed by the core conception of practices offered by Schatzki, and summarised above, each chapter attends to and empirically analyses two particular data practices, such as defining and deriving (Chapter 3), coordinating and narrating (Chapter 4), or inferring and assigning (Chapter 6). Methodologically, we follow Knox and Nafus's (2018) proposition that ethnographies of data practices can generate new ways of theorising and understanding digital data and relations of knowledge production. Critically, as the chapters elaborate, this involved studying data practices through multi-method and multi-sited ethnographies that entailed observing conferences and meetings, compiling and analysing documents and reports,

conducting interviews, and engaging in conversations with statisticians across myriad European sites (Scheel et al., 2020). While statisticians engage in numerous data practices, the chapters focus particularly on those involved in making up categories of people for, as we argue below, it is by allocating individuals to categories that statistical methods enact the population and people of Europe.

Classifying and Encoding Individuals into Categories: Making Up a People

Practices of classifying and encoding are central to the production of statistics. In a study of the relation between statistics and the making of the modern state, Desrosières (1998) argues that statistics involve establishing 'categories of equivalence' that transcend the singularities of individual situations and thereby 'make a priori separate things hold together' (236). Writing about the emergence of nationalism, Anderson (2006) shows, in turn, how the production of ethnic and racial categories by imperial powers shaped the emergence of imagined communities along national and colonial lines. However, censuses have historically been made up of numerous classifications and categories that have enacted populations according to residence, age, sex, nationality, birthplace, and citizenship. But classification involves more than defining categories; it also requires practices of encoding through which individuals are allocated to categories. That is, it is through data practices that categories first get defined and then populated as individuals are allocated to them so that they can be constituted as parts of a population.

Enumeration demands the identification of kinds of people to count and it is through categories that this has been done in censuses and population statistics (Hacking, 2015).

Hacking notes that many of the categories 'we now use to describe people are by-products of the needs of enumeration' (280). He argues that 'biopolitics as the transition from the counting of hearths to the counting of bodies' follows from this. Furthermore, 'the subversive effect of this transition was to create new categories into which people had to fall, and so to create and to render rigid new conceptualizations of the human being' (281). This subversive effect of categories, he argues, is the result of a circular process he calls 'dynamic nominalism': a kind of person comes into being when the kind itself is invented. Put differently, the category and the categorised are co-constituted and emerge through 'feedback-loops' between the two (Hacking, 1999). For Hacking (2015: 280) '[t]he fetishistic collection of overt statistical data about populations' is generative of this unintended effect.

The question that follows from Hacking is not whether categories are real but how they have been enacted through practices that involve battles over truth, definitions, controversies, and so on (cf. Grommé and Scheel, 2020). However, once settled, a category can be said to exist and can be investigated, acted upon and identified with (Ruppert, 2007). This is, however, a historically contingent outcome: some categories are enacted, put to use and sanctioned as 'official' through their usage in population statistics, for instance, while others are not. Hacking says there are many possible descriptions that are true of the world, but the struggles that establish the truth of one version close off other equally true versions. This contingency does not disqualify the truth status of versions of the world but does account for why some things become true rather than others, or why some categories become authoritative, and others do not. Once authoritative, categories can then be deployed administratively, shape social development, support particular political projects, have practical consequences

for the distribution of resources, and shape collective identities (Kertzer and Arel, 2002).

While building on Hacking, we depart in two important ways. Whereas Hacking (1999) focuses on categories that make up specific 'kinds of people' (such as heterosexual or autistic people), we are also concerned with the categories through which censuses come to *make up a people* (cf. Isin, 2018). Just as the various categories that make up different kinds of people are not given, so too are those that come to make up *a people*. That is, the invention of different kinds of people is bound up with the invention of *a* people. Unsurprisingly, the categories we focus on in this book such as migrants and other mobile subjects are the consequence of the still dominant conception that the population and the people of Europe are sedentary. The different categories we examine are kinds of people who are imaginable as a consequence of having first constituted sedentary people as a norm and thus mobile people as an exception (Isin, 2018: 121).

A second distinction from Hacking is that the performative powers of categories are not only located in the feedback loops between categories and the named, but also in the data practices that are used to enact these categories in the first place. Hence, we highlight how classifying and encoding people into categories involves various data practices that are developed, negotiated, and experienced by experts entrusted with their production: national and European statisticians who operate within a transnational field of statistics. That is, beyond arguing that data practices are important, we specify how categories, from 'usual resident', 'refugees', and 'homeless people' to 'migrants', are done through specific data practices and how these data practices come to matter.

The central role that categories play in enacting the population and people of Europe is also reflected in the structure

of this book. Five of the book's seven chapters focus on different categories of migrants or other mobile subjects to analyse who is enacted as part of the European population. These categories include foreigners, refugees and asylum seekers, and homeless people, but also the usually resident population. The latter is of central importance for population statistics but considered particularly difficult to establish precisely because of the different forms of mobility of increasingly large segments of society.

All of these five chapters investigate the making of categories by attending to two particular data practices such as defining, estimating, inferring, sieving, recalibrating, or narrating through which categories of equivalence are defined and literally populated with individuals. The data practices they analyse are part of different census methods such as traditional questionnaire-based or population register-based methods. However, some are part of other methods of producing population statistics such as those involving experiments with big data. The reason is not empirical randomness. Rather, data practices are not confined to any one method but are part of repertoires such as cleaning, estimating, and inferring that are variously taken up and adapted across methods.

Additionally, other methods are connected to, and in many cases rely upon, censuses, which have traditionally been considered 'the benchmark for population counting at national and local levels' (CES, 2006: 5). In other words, it has served as the 'gold standard' and 'ground truth' for other methods such as a sample surveys. The reason is that the census is meant to provide an 'inventory', that is, a comprehensive account of the total population (e.g., Puur and Tammaru, 2012) instead of just capturing – like a survey – a sample from which the whole can be estimated. Moreover, some statisticians prefer traditional questionnaire-based methods that involve face-to-face

enumeration over register-based ones which, they argue, replicate the information held in administrative registers and may thus provide an incomplete account of a population (Puur, Sakkeus, and Aben, 2013).

The point is that the data practices involved in making up the population and people of Europe circulate and are part of overlapping method assemblages (as defined at the beginning of this chapter). Census methods can be said to form distinct assemblages of standards, routines, people, technologies, materials, and infrastructures especially in relation to their national contexts. Assemblages carry national and colonial legacies including different political and cultural understandings of issues such as privacy, identity, national belonging, citizenship, and so on. At the same time, through their transnational relations, they overlap, interact, and are related in myriad ways to those of other national contexts including the data practices that they borrow, take up, and adapt. As such, method assemblages are not stable and fixed but always in flux in part due to the changing data practices that circulate and make them up both nationally and internationally. This brings us back to the opening reflections about the implications of digital and technological changes for the production of official population statistics in the opening of this chapter. Those changes are introducing new sources of data, technologies, and practices as well as new actors, such as data scientists and data producers such as platform owners to the assemblages that make up official population statistics. What this means for the data practices that contribute to the enactment of a European population and people, and the ontological politics of such enactments, are the focus of the chapters that follow.

References

Anderson B (2006) *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Revised and extended. London: Verso.

Ardittis S and Laczko F (2017) Measuring Irregular Migration: Innovative Data Practices (Introduction to Special Issue). *Migration Policy Practice* 7(2): 2–3.

Bloor D (2001) Wittgenstein and the Priority of Practice. In: Schatzki TR, Knorr-Cetina KD, and Savigny E von (eds) *The Practice Turn in Contemporary Theory*. London: Routledge, pp. 103–114.

Bourdieu P (1990) *The Logic of Practice*. Cambridge: Polity Press.

Bourdieu P and Wacquant L (1992) *An Invitation to Reflexive Sociology*. Cambridge: Polity Press.

Broeders D and Dijstelbloem H (2016) The Datafication of Mobility and Migration Management: The Mediating State and Its Consequences. In: Van der Ploeg I and Pridmore J (eds) *Digitizing Identities: Doing Identity in a Networked World*. London: Routledge, pp. 242–260.

Bueger C and Gardinger F (2015) The Play of International Practices. *International Studies Quarterly* 59(3): 449–460.

Cakici B, Ruppert E, and Scheel S (2020) Peopling Europe Through Data Practices: Introduction to the Special Issue. *Science, Technology & Human Values* 45(2): 199–211.

CES (2006) *Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing*. Geneva: UNECE and Eurostat.

Collins HM (2001) What Is Tacit Knowledge? In: Schatzki TR, Knorr-Cetina KD, and Savigny E von (eds) *The Practice Turn in Contemporary Theory*. London: Routledge, pp. 115–128.

Desrosières A (1998) *The Politics of Large Numbers: A History of Statistical Reasoning* (ed. RD Whitley). Cambridge, MA and London: Harvard University Press.

Edwards PN (2010) *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA and London: MIT Press.

Fotopoulou A (2019) Understanding Citizen Data Practices from a Feminist Perspective: Embodiment and the Ethics of Care. In: Stephansen H and Trere E (eds) *Citizen Media and Practice: Currents, Connections and Challenges*. Oxford and New York: Routledge, pp. 227–242.

Foucault M (2000) *Power: Essential Works of Foucault, 1954–1984* (ed. JD Faubion). London: Penguin.

Gabrys J, Pritchard H and Barratt B (2016) Just Good Enough Data: Figuring Data Citizenships Through Air Pollution Sensing and Data Stories. *Big Data & Society* 3(2): 1–14.

Gad C and Jensen CB (2014) The Promises of Practice. *The Sociological Review* 62(4): 698–718.

Garnett E (2016) Developing a Feeling for Error: Practices of Monitoring and Modelling Air Pollution Data. *Big Data & Society* 3(2): 1–12.

Gitelman L (2013) *'Raw Data' Is an Oxymoron*. Cambridge, MA: MIT Press.

Grommé F and Scheel S (2020) Doing Statistics, Enacting the Nation: The Performative Powers of Categories. *Nations and Nationalism* 26(3): 576–593.

Hacking I (1999) Making up People. In: Biagioli M (ed.) *The Science Studies Reader*. London: Routledge, pp. 161–171.

Hacking I (2015) Biopower and the Avalanche of Printed Numbers. In: Cisney VW and Morar N (eds) *Biopower: Foucault and Beyond*. Chicago: University of Chicago Press, pp. 65–80.

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

Hui A, Schatzki TR and Shove E (eds) (2017) *The Nexus of Practices. Connections, Constellations, Practitioners*. New York: Routledge.

Isin E (2018) Mobile Peoples: Transversal Configurations. *Social Inclusion* 6(1): 115–123.

Kertzer DI and Arel D (2002) *Census and Identity: The Politics of Race, Ethnicity, and Language in National Censuses*. New Perspectives on Anthropological and Social Demography. Cambridge: Cambridge University Press.

Kitchin R (2014a) Big Data, New Epistemologies and Paradigm Shifts. *Big Data & Society* 1: 1–12.

Kitchin R (2014b) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.

Kitchin R (2015) The Opportunities, Challenges and Risks of Big Data for Official Statistics. *Statistical Journal of the IAOS* 31(3): 471–481.

Knorr-Cetina K (2001) Objectual Practice. In: Schatzki TR, Knorr-Cetina KD, and Savigny E von (eds) *The Practice Turn in Contemporary Theory*. London: Routledge, pp. 172–183.

Knox H and Nafus D (eds) (2018) *Ethnography for a Data Saturated World*. Manchester: Manchester University Press.

Labbé M (2000) Censuses, Plebiscites and the Categorizations of Identities. In: Whose Self-Determination? Agency and Amnesia in the Disintegration of Yugoslavia (conference paper), Charlottesville, VA: Watson Institute and Brown University.

Law J (2004) *After Method: Mess in Social Science Research*. London and New York: Routledge.

Law J (2008) On Sociology and STS. *Sociological Review* 56(4): 623–649.

Law J (2012) Collateral Realities. In: Rubio F and Baert P (eds) *The Politics of Knowledge*. Abingdon and New York: Routledge, pp. 156–178.

Law J and Lien ME (2012) Slippery: Field Notes in Empirical Ontology. *Social Studies of Science* 43(3): 363–378.

Leahey E (2008) Overseeing Research Practice: The Case of Data Editing. *Science, Technology, & Human Values* 33(5): 605–630.

Leonelli S (2016) *Data-Centric Biology*. Chicago: University of Chicago Press.

Mayer-Schönberger V and Cukier K (2013) *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murray.

Mol A (2002) *The Body Multiple: Ontology in Medical Practice*. Durham, NC: Duke University Press.

Pickering A (1995) *The Mangle of Practice: Time, Agency, and Science* . Chicago and London: University of Chicago Press.

Pickering A (2001) Practice and Posthumanism: Social Theory and a History of Agency. In: Schatzki TR, Knorr-Cetina KD, and Savigny E von (eds) *The Practice Turn in Contemporary Theory*. London: Routledge, pp. 172–183.

Puur A and Tammaru T (2012) Census Under-Coverage and Data Correction Opinion and Proposals. Unpublished Paper presented at the Scientific Council Meeting of Statistics Estonia on 25 June 2012. Statistics Estonia.

Puur A, Sakkeus L and Aben S (2013) *Projekti lõpparuanne. Aruande versioon v 1.04 18.09.2013 [Methodology Study. final version]*. Tallinn: Institute of Population Studies. University of Tallinn.

Reckwitz A (2002) Toward a Theory of Social Practices: A Development in Culturalist Theorizing. *European Journal of Social Theory* 5(2): 243–263.

Ruppert E (2007) Producing Population. CRESC Working Paper Series, Paper No. 74. Available at: https://www.cresc.ac.uk/.

Ruppert E (2011) Population Objects: Interpassive Subjects. *Sociology* 45(2): 218–233.

Schatzki TR (2001) Introduction: Practice Theory. In: Schatzki TR, Knorr-Cetina KD, and Savigny E von (eds) *The Practice Turn in Contemporary Theory*. London: Routledge, pp. 10–23.

Scheel S, Ruppert E and Ustek-Spilda F (2019) Introduction: Enacting Migration Through Data Practices. *Environment and Planning D: Society and Space* 37(4): 579–588.

Scheel S, Grommé F, Ruppert E, et al. (2020) Doing a Transversal Method: Developing an Ethics of Care in a Collaborative Research Project. *Global Networks* 20(3): 522–543.

Suchman L (2007) *Human-Machine Reconfigurations: Plans and Situated Actions*. 2nd ed. Cambridge: Cambridge University Press.

Swidler A (2001) What Anchors Cultural Practices. In: Schatzki TR, Knorr-Cetina KD, and Savigny E von (eds) *The Practice Turn in Contemporary Theory*. London: Routledge, pp. 83–101.

Turner S (2001) Throwing Out the Tacit Rule Book: Learning and Practices. In: Schatzki TR, Knorr-Cetina KD, and Savigny E von (eds) *The Practice Turn in Contemporary Theory*. London: Routledge, pp. 129–139.

van Dijck J (2014) Datafication, Dataism and Dataveillance: Big Data Between Scientific Paradigm and Ideology. *Surveillance & Society* 12(2): 197–208.

3

Usual Residents: Defining and Deriving

Francisca Grommé, Evelyn Ruppert,
and Funda Ustek-Spilda

Who Is Usually Resident?

In the context of a population census, a country is free to enumerate (in the sense of collecting statistical data on) any person in its territory, as well as to define population counts which meet national needs (CES, 2013: 5).

The fiction of the census is that everyone is in it, and that everyone has one – and only one – extremely clear place. No Fractions. This mode of imagining by the colonial state had origins much older than the censuses of the 1870s (Anderson, 2006: 166).

Who should be counted as part of a national, and in turn the European, population may appear to be a simple matter. However, the determination of who should be counted has long constituted a fundamental challenge for the making of censuses and other state statistics. Yet, defining who should be counted and then encoding each person to a single location is the foundation of a rationality of knowing who are the subjects of governing within specific political jurisdictions. It is a foundation based on the dominant understanding that subjects are sedentary and settled in a state's territory and constitute its people (Isin, 2018). What has also been referred to as a 'sedentary bias' in development and migration studies, the underlying assumption of both state and non-state practices that are based on this understanding is that 'people want

to remain in their place' (Bakewell 2007, 10). This assumption, however, has resulted in the production of statistics that treat national boundaries as containers of populations as well as methodological nationalist assumptions of the social sciences (Sager, 2016). Many theories in geography, anthropology, and sociology assume people are sedentary and that stability, meaning, and place are normal, and distance, change, and placelessness are abnormal (Sheller and Urry, 2006: 208).[1] As noted in Chapter 1, understanding subjects as sedentary is part of making a society legible, which Scott (1998) argues is a central objective of statecraft. Historically, this bias has 'simplified the classic state functions of taxation, conscription, and prevention of rebellion' (2) and continues to serve these as well as other functions such as voter registration, parliamentary representation, and social rights to state services.

At the same time, since the inception of modern national censuses some 200 years ago, statisticians have recognised that a population is not singular, but that there are various target populations of interest (or population bases) such as workday and out-of-term populations. These bases are also referred to as 'theoretical' populations to distinguish the 'population to be enumerated' (the set of persons whom the country decides should be covered by the census, regardless of their subsequent exclusion from any specific population count) and the 'enumerated population base' (those persons who have actually been enumerated) (CES, 2015: 76). Consequently, populations can be understood as multiple as national statistical institutes (NSIs) differently define their population bases and conduct censuses. For example, some NSIs define a population base according to the principle of de jure (by law, that is, legally resident) while others apply the de facto principle (by presence, that is, found to be resident) to determine if

a person is to be counted on a designated census reference date. Some conduct censuses based on their population registers while others conduct questionnaire-based censuses. Whatever the population base, how it is defined involves decisions that make some people 'present "in-here", whilst making others absent "out-there" ' (Law, 2004: 14). What these differences also highlight is that not only are populations contained in national boundaries but also enacted by nationally defined methods. As the opening quote states, countries are 'free' to determine how and who they enumerate. For Eurostat this includes the data sources, methods, and technologies that 'best' suit a member state's context (Eurostat, 2011: 9).[2] The dominance of the national is thus not only to be found in a sedentary bias and the assumption that borders contain populations, but also in the methods through which populations are enacted.

This chapter addresses how the dominance of the national is especially problematic for the production and comparison of European and international population statistics. In response, and as part of a broader programme of statistical harmonisation, the UN, UNECE, and EU have progressively developed and adopted in tandem a definition of the 'usually resident population' to serve as the population base for international comparison.[3] For the EU, a harmonised population base is not only necessary for comparison, but for governing functions such as policymaking and the allocation of resources; for Qualified Majority Voting (QMV) in the European Council; and the allocation of MEP seats in the European Parliament (Eurostat, 2017).

The objective of a harmonised definition is to 'allocate each person to one, and only one, place of usual residence' (CES, 2015: 76). Allocate means to literally assign, connect,

or link a person to a particular location.[4] To meet this objective requires the two processes identified by Desrosières and discussed in Chapter 2: classifying and encoding, which both involve myriad data practices. For instance, classifying involves the practice of defining categories and encoding the practice of assigning and allocating individuals to those categories. Both are required to enact what is the usually resident population of a country and in turn the populations of the EU and UNECE.

Through task forces, working papers, and meetings from 2012–15, the following definition was adopted by the UNECE (together with Eurostat and closely mirrored in EC regulations) for the 2020 enumerations:[5]

> The '*place of usual residence*' is the geographic place where the enumerated person usually spends their daily rest, assessed over a defined period of time including the census reference time.
>
> The population base to be used for international comparisons purposes is the 'usually resident population'. The '*usually resident population*' of a country is composed of those persons who have their place of usual residence in the country at the census reference time and have lived, or intend to live, there for a continuous period of time of at least 12 months. A 'continuous period of time' means that absences (from the country of usual residence) whose durations are shorter than 12 months do not affect the country of usual residence. The same criteria apply for any relevant territorial division (being the place of usual residence) within the country.
>
> (CES, 2015: 78; italics in original)

This chapter focusses on two problems that developing and implementing this harmonised definition encounter. The first concerns how the category of usually resident is often at odds with modes of living that are experienced by numerous people because of choice, circumstance, law, or force.

While historically many modes of living have differed from government definitions, contemporary problematisations of the category are in part related to EU citizens exercising their right to freedom of movement (live and work and have rights to social and other benefits when they do so) granted by the Maastricht Treaty. The exercise of this legal right has been generative of new 'mobile people' (Isin, 2018) in the EU who challenge the sedentary bias of statecraft and the category of usually resident such as people who have residences in more than one country or people who live and work across national borders.[6] We refer to these multiple modes of transborder movements as the problem of the complexity of mobility.

The second problem concerns how the harmonised category of usually resident confronts pre-existing definitions, rules, technologies, priorities, and histories, all of the sociotechnical arrangements that make up the method assemblages of different countries. These assemblages also extend beyond NSIs to include a hinterland of relations (as defined in Chapter 2) for example, those of other government agencies such as administrative departments, which use different definitions that fit purposes such as taxation, health services, electoral rolls, and education (Potter and Champion, 2014). The implementation of a new definition requires changing or unlocking these relations, which is often very difficult to do.[7] Furthermore, such difficulty – including even the possibility of implementing a definition – varies considerably depending on a country's existing census method. For example, questionnaire-based methods can often be more readily changed to fit new definitions whereas register-based methods are more 'locked-in' to pre-existing administrative definitions as we illustrate later. Defining the category of usually resident may seek to anticipate and ameliorate these difficulties. However, the analysis that follows brings attention to how defining interacts with

the contingencies, capacities, circumstances, politics, and histories of national method assemblages through which people are encoded.[8] As we will elaborate, this includes different political and cultural understandings of issues such as privacy, identity, national belonging, citizenship and so on. We refer to this problem as the multiplicity of methods.

This chapter examines two data practices which aim to ameliorate these problems: defining special cases and deriving usual residents. They are part of two essential stages of statistical work of classifying and encoding identified by Desrosières (1998): the former involves defining categories of people as exceptions to the definition but who nonetheless should be included in the usually resident population (classifying) and the latter involves deriving which individuals should be allocated to the population (encoding). Both practices do not bring into question the category of usually resident and its sedentary bias. Instead, they involve elaborate data practices that serve to implement and sustain both. That is, while harmonisation and international comparability are touted as the main objectives, the usually resident category also serves to sustain the primacy of nationalist assumptions that people are sedentary and emplaced in one and only one national territory. Furthermore, as we elaborate below, the data practices also sustain the primacy of national methods in making up a European population. We return to the implications of this for the European project in the conclusion and reflect on how the category of usually resident reveals political tensions in the enactment of a European population and people.

Defining Special Cases

The definition of who is usually resident is based on two criteria: location – where a person 'usually' spends their daily rest – and time – where a person has 'usually' lived

for a continuous period of 12 months (without absences of more than 12 months). While it is intended to establish a clear delineation between who is or is not usually resident, it introduces what is acknowledged as 'uncertainty' (CES, 2015). The sources of uncertainty were discussed at numerous international contexts such as conferences and task force meetings where national and international statisticians considered proposals for the 2020 round of censuses.[9] These situated discussions reveal the uncertainties, tensions, and compromises made to meet the objective of allocating each person to one, and only one, national place of usual residence.

One uncertainty arising from the definition concerns the possibility of a precise meaning of 'daily rest' and, even if defined, whether it is possible to determine where the majority of daily rest is spent for people who have multiple places of rest.[10] Is rest the best criterion for establishing who is usually resident or should ownership or the location of a person's belongings be more appropriate, for example? A second uncertainty concerns the patterns of mobility of particular 'population groups' that do not easily fit the definition such as people who have residences in several countries or have no residence such as homeless people (CES, 2015: 79). How were both resolved? Differences and ambiguities in the interpretation of terms such as daily rest were deemed a matter of implementation and 'population groups' that do not fit the definition were delimited as 'particular' or 'special cases' (which we will refer to simply as special cases). While the definition of the category of usually resident occupies about a paragraph, rules were developed for the UNECE guidelines and EC regulations for the following special cases, which take up about two pages or 15 paragraphs respectively: persons who live in more than one residence; primary and secondary students away from home during school term; tertiary students away

from home while at college or university; persons living in institutions; persons doing military service; homeless or roofless persons; nomads, vagrants, and persons with no concept of usual residence; children who alternate between two places of residence; merchant seamen and fishermen; persons who may be illegal, irregular, or undocumented migrants, as well as asylum seekers and persons who have applied for, or been granted, refugee status or similar types of international protections; children born 12 months before the census reference time; persons whose stay in a country is exactly one year; military, naval, and diplomatic personnel and their families; and persons usually resident but absent at the time of the census.[11] The rules specify conditions which must be met for including each of these cases in the usually resident population.

These additional specifications did not resolve all uncertainties about the inclusion or exclusion of persons within each of the special cases. For example, for the special case of tertiary education students, some countries allocate students to the family home to reduce an overcount due to double-counting.[12] They argue that not doing so would have a significant impact on the age structure of a population, especially in small countries where many young people study abroad. Others use the term-time address because tertiary education is generally the time when a person starts to break away from their family nucleus, and because some university towns can double in population during term-time. In the interests of harmonisation, the UNECE guidelines thus added yet another stipulation:

> Students in tertiary education should be allocated to their term-time address, when studying within the country. When studying abroad they should not be included in the population of the country of their family home, since their place of usual residence should be the

term-time address in the country where they study, even if they are regularly returning to the family home. However, it is acknowledged that in some countries there may be considerations (such as higher coverage during field enumeration, or particularly high quota of emigrating student population) that would justify the allocation of these students at their family home (CES, 2015: 80–81).

In other words, exceptions were made to the exceptions so that under certain conditions NSIs can allocate students to the family home. A similar rule was adopted in the EC regulation, which specifies that while the term-time address shall be the usual residence for tertiary students regardless of whether they are pursuing their education elsewhere in the country or abroad, 'exceptionally', where the place of education is within the country, the place of usual residence may be considered to be the family home.[13]

The specification of rules for special cases acknowledges two issues that arise with a harmonised definition of usually resident. First, it recognises the multiplicity of national methods, which are usually based on variable definitions that often carry cultural meanings and long legacies that cannot easily fit the harmonised one. Second, it recognises the complexity of mobilities, which are not only diverse and difficult to define but also hard to enumerate. The harmonised definition mediates these issues through special cases and exceptions that make it sufficiently flexible to adapt to national differences and sufficiently robust to maintain commonality. It is in this sense that the harmonised definition can be said to operate as a 'boundary object' between national and international data practices:

We define boundary objects as those objects that both inhabit several communities of practice and satisfy the informational requirements of each of them. In working practice, they are objects that are

able both to travel across borders and maintain some sort of constant identity. They can be tailored to meet the needs of any one community (they are plastic in this sense, or customizable). At the same time, they have common identities across settings. This is achieved by allowing the objects to be weakly structured in common use, imposing stronger structures in the individual site tailored use (Bowker and Star, 1999: 16).

However, at the same time, by mediating robustness and flexibility, special cases help sustain the definition and its nationalist premises. A few of the identified special cases do not concern cross-border mobility per se (e.g., persons in institutions or people who are homeless within a national territory), but most do. While the mobilities of people defined as 'migrants' are commonly problematised, the special cases reveal how the mobilities of relatively privileged groups are also problematic for data practices that seek to define and implement the category of usually resident. However, there are other mobilities not identified as special cases but which also bring into question the sedentary bias of the category such as people engaged in weekly commuting, seasonal movements, 'living apart together', and transient labour migration (Potter and Champion, 2014).[14] This was noted at a meeting of the British Society for Population Studies, which was framed around the question of whether the concept of usually resident has reached its 'sell-by date' (ibid.). One presentation focused on the rise of transnational 'super commuters', which is leading to 'multilocal living'. The category refers to the use of two or more residences by the same occupants, their circular mobility between residences and alternating phases of presence and absence in each of the residences. One study estimated that 11 per cent of French residents and 28 per cent of the Swiss fall within this definition and that some people spend more

of their time in an average year at a residence other the one that they regard as their most important (Duchêne-Lacroix, 2014). Other studies were noted which highlight the complexities of mobile people such as weekenders, weekly commuters, FIFOs (FlyIn/FlyOut workers), and business travellers (Potter and Champion, 2014). These findings are echoed in studies of mobility made possible by new digital technologies such as mobile phones, which have found that the 'activity spaces' of people include multiple locations with their patterns that are diverse and complex.[15]

Special cases exclude such mobilities, which concern people who move between countries within (and/or outside) the designated 12-month reference period and also for varying lengths of time in the ten years between censuses. As the examples above indicate, some of these concern different forms of labour mobility exercised by people who the EC have named 'mobile citizens' (Eurostat, 2018).[16] In 2017, four types of EU mobile citizens were defined and counted in statistics produced by Eurostat.[17] The statistics are based on the EU Labour Force Survey (EU-LFS), which draws on data that NSIs are required to collect and report quarterly to Eurostat as well as data from other sources. 'Long-term movers' are people who lived in an EU country other than their country of citizenship for more than 12 months and made up approximately 4 per cent of the EU population (17 million EU citizens, an increase from 11.8 in 2016) (Fries-Tersch et al., 2018).[18] A second is 'cross-border workers', that is, citizens who reside in one country but are employed or self-employed in another and who, for this purpose, move across borders regularly. Based on the EU-LFS, there were approximately 1.4 million cross-border workers in 2017. The third is made up of approximately 2.8 million mobile citizens reported as 'posted' workers, people regularly employed in one member state but sent to another by the same

employer for a limited period of time. And finally, the fourth comprises 680,000 nationals who returned to their country of origin after an absence of more than 12 months ('return mobility').[19] Two of these categories – long-term movers and return mobility – adhere to the definition of usually resident in that only a person who moves to another country for a period of at least 12 months is included (which also defines a long-term migrant) (CES, 2015: 80). However, as explored in the next two sections, the cross-border worker and different forms of return mobility are not recognised and constitute yet additional exceptions to exceptions. Discussions about the cross-border worker exemplify the problem of the multiplicity of methods; while discussions of return mobility focus on the complexities of mobility involving repeated moves between two or more countries, as defined in yet another category, that of 'circular migrants'. That the identification of, and statistics on, both of these categories were compiled from a variety of different sources including the EU-LFS, further points to how a harmonised definition cannot account for different modes of living.

Cross-border Workers

A discussion at a meeting of an ESS task force considering regulations to implement the harmonised definition identified a number of problems. One concerned the collection and reporting of data on the country of work of people who were deemed usually resident in one member state but who work in another.[20] While countries that conduct questionnaire-based censuses collect data on place of work, countries which conduct register-based censuses do not collect this data and must rely on other sources. One statistician, while agreeing this is a problem, questioned the relevance of collecting this data

in the first instance since it is estimated that 99.5 per cent of people work in their country of residence; what, then, is the value of knowing this statistic? Additionally, reporting on the country of work for relatively small numbers of people working outside their country of residence would potentially lead to data confidentiality problems. Some statisticians argued that the census is not suited to measuring these movements and that data from the EU-LFS is more useful. Others offered that perhaps data from mobile phones, tax, and social insurance registers might be more relevant. The solution came in the form of agreeing that when data is available, then a general and singular category of 'Not in the territory of the Member State' be reported as the location of work rather than specifying the countries. Member states would then determine, based on their census method and other available data, what numbers to report, and if not measurable or relevant, enter zeros for the category. As such, rather than driven by the definition, data on cross-border movements would be driven by what was possible by national methods.

While acknowledging the practicalities of counting cross-border workers raised by NSIs, Eurostat statisticians at the meeting noted that the EC deems this a very important statistic as it is related to the right of free movement of EU citizens to live and work in member states other than their country of citizenship. For this reason, knowing the scale of the exercise of this right was expressed as necessary from a 'political point of view'. They acknowledged that numbers may be small in some cases, and the majority of movements likely apply more to some countries than others and to bordering regions. However, there is a 'political perception' that such movements can have a big effect on neighbouring countries with significant differences in salary levels. Yet, at the moment, there is insufficient statistical evidence about these movements. The difference between what

EC policymakers deemed important and what national statisticians deemed measurable reveals a tension between policies and statistics. On the one hand, policies enable mobilities (e.g., laws on free movement) which in turn affect and shape statistical categories and definitions that are needed to render mobilities objects of management and governing. A statistician at a different meeting described this as an 'interaction between political decisions and statistics' and argued that the two are 'developed in tandem.'[21] At the same time, while being shaped by policies, what data practices come to enact is not a simple reflection of policies but also, as argued above, involve an interaction between the complexities of mobility and multiplicity of methods.

Harmonised definitions, while operating as boundary objects to manage robustness and flexibility, also serve the governing logic of knowing who are the subjects of specific political jurisdictions. This relation is also exemplified in the case of circular migration, which is another exception to exceptions, but distinct from the category of return mobility.

Circular Migration

Like the other categories of mobile citizens, return mobility is based on the crossing of one international border in a move that begins and ends in the same country (from country A to B and then return from B to A), is bi-directional, is for a continuous period of more than 12 months, and, save for cross-border workers, one-time only (Figure 3.1). However, this does not capture people who sometimes move repeatedly between two or more countries. This pattern is referred to as 'circular migration' and was defined by the EC in 2007 as 'a repetition of legal migration by the same person between two or more countries' (European Migration Network, 2011: 14).

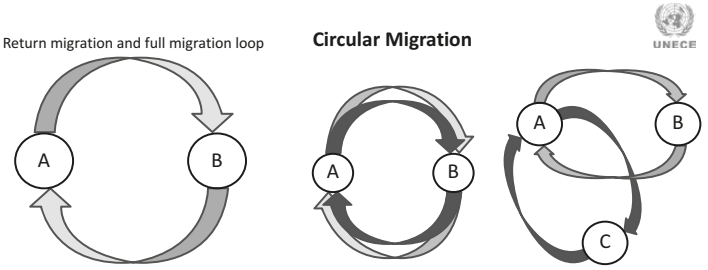

Figure 3.1 Return Migration and Two Forms of Circular Migration[a]
[a]Source: CES, 2016a, 11

The EC definition was introduced as part of a growing interest amongst 'policymakers and researchers alike' in the early 2000s who were heralding circular migration as a

> migration 'tool' which creates a 'triple win' situation by producing three beneficiaries: the host society whose labour shortages will be filled; the migrant who will have greater opportunities to increase his/her employability; and the country of origin which will benefit from remittances as well as newly-acquired skills of returning migrants (European Migration Network, 2011: 10).

Many national labour market initiatives have sought to achieve these 'wins' through policies, for example, on the recruitment of temporary migrants such as agricultural workers, care providers and workers in the hospitality sector. However, the growing political interest in circular migration and its promotion and management coincided with a diversity of national and international definitions (despite the EC recommended definition), statistics, legislation, and policies and a dearth of statistics (only a handful of EU countries measured it in the period leading up to 2011).

These were some of the conclusions of a 2011 report of the European Migration Network (EMN), which was established by the EC in 2008.[22] Its main recommendation called for harmonising key concepts and improving data collection on circular migration. A subsequent UNECE Task Force on Measuring Circular Migration was set up in 2013 to review this recommendation and prepare a proposal for an internationally harmonised definition. The illustrations in Figure 3.1 are from the 2016 final report of that task force.[23] The report noted that while the EMN definition was a first step towards a harmonised definition, many different versions persisted on the part of international bodies such as those of the EU, UNECE and national governments. It argued that one of the reasons is that most definitions, including that of the EMN, are 'conceptual' in that they broadly *describe* circular migration but do not constitute 'statistical' (or *operational*) definitions. That is, they only aim to describe what is to be measured but do not define the practicalities of such measurement (CES, 2016b: 16). The lack of agreement on both conceptual and statistical definitions was identified as the main reason for why data on this category continues to be incomparable across states. The task force thus recommended standard definitions for both and adopted the following EMN conceptual definition:

> [a] repetition of legal migration by the same person between two or more countries'; and, of a statistical definition: '[a] circular migrant is a person who has crossed the national borders of the reporting country at least 3 times over the past 10 years, each time with duration of stay (abroad or in the country) of at least 12 months (CES, 2016b: 16–17).

Of significance is that the statistical definition for circular migration also adopted the 12-month time criterion of the

usually resident definition. That is, only movements greater than 12 months would count.

The task force recommendations were presented at an international conference on migration statistics organised by Eurostat and UNECE in 2016.[24] Discussions at the conference revealed how the problems of the complexity of mobility and the multiplicity of methods came to matter for the statistical definition of circular migration for two reasons: first, it excluded short-term movements (less than the 12-month period specified in the definition of usually resident); and second, depending on their census method, NSIs are variously able to implement it, whatever the definition.

On the first, one statistician explained that, 'with this definition we also end up not capturing the group between short- and long-term circular migration' whereas another commented that trying to differentiate migration in too much detail would be too complex and result in confusion:

> One of my concerns is mixed migrations. Do we just ignore them? This seems impossible. Clear distinction between short and long term is hardly possible. I am not quite sure about [what] the starting point of any circular migration should be? Does it have to be in the starting country? Also is there an overlap between circular emigrant and circular immigrant? Do we count certain people multiple times? How do we distinguish them? By definition, does it not have some overlap between certain countries? [Is it] because certain countries will be able to capture circular migration much more clearly because they have more advanced statistical systems?

Others offered that there is a risk of confusing short-term circular migration with seasonal migration, and that short-term migration statistics are not well collected by many countries. Regarding the latter, another statistician highlighted that while

differentiating migration in this way makes a lot of sense in principle and conceptually, producing data for these categories was easier said than done. They reminded others at the conference that some countries, such as the UK, rely on survey data to produce data on migration (e.g., UK's International Passenger Survey)[25] and would not be able to provide this level of detail; they certainly would not be able to construct a 10-year migration history for a person. They also noted that even if people were asked, it would be a challenge for them to recall their movements over a 10-year period.

In contrast to the UK, countries that conduct register-based censuses were imagined to more easily differentiate and enumerate migration movements because they would be recorded in at least one administrative register. Sweden tested this based on a definition of a circular migrant as 'a person that in migration purpose had crossed the Swedish border at least twice during the period 1969–2009', both for short- and long-term periods (Statistics Sweden, 2016). The estimates used for the test were based on joining up multiple administrative registers (e.g., income, education, population) into a new register, called the 'Circular Migration Register', which could be used to produce annual statistics. In their presentation, the Swedish statistician who led the test explained that 'yo-yo migrants' (a term used to refer to circular migrants) during this period were estimated to be around 236,000 (with the total circular migration events estimated to be around 900,000).[26] He noted that the biggest circular migration group was from Nordic countries where there has always been an 'in-out-flow' and the second largest group comprised migrants from Asia, Iraq, and Iran. In noting that most migrants return home within five years, for circular migrants the time between the first and last migration for the majority was estimated to be 16 years, rather than the 10 years specified in the statistical definition.

While administrative registers, such as those used by Statistics Sweden, are more suited to tracking circular migration, they are not without their problems. Statisticians often note that registers are not reliable sources on migration as people often do not de-register when they move even though they might be legally required to do so.[27]

Based on these and other responses to the UNECE Task Force report, the CES amended and then approved the following statistical definition so that 'policy needs' on short-term migration could be accommodated:

> A circular migrant is a person who has crossed the national borders of the reporting country at least 3 times over a 10-year period, each time with duration of stay (abroad or in the country) of at least 12 months.
>
> To meet the policy needs for information on shorter durations of stay, the extended statistical definition allowing for short-term migrations is as follows: A circular migrant is a person who has crossed the national borders of the reporting country at least 3 times over a 10-year period, each time with duration of stay (abroad or in the country) of at least 90 days (CES, 2016a: 19).

These recommendations for harmonised definitions of the cross-border worker or circular migrant came too late to be considered in the guidelines or regulations for the 2020–21 round of censuses. Rather, the two categories will continue to be documented separately. For the EU, that will happen though the EU-LFS, which currently constitutes cross-border workers as one type of 'mobile citizen' whereas circular migrants will continue to occupy an ambiguous position between the categories of long-term mover or return mobility migrant.

As expressed in the 'political view' of the EC, some mobilities not recognised as special cases such as cross-border workers who are a consequence of EU citizens exercising their

legal right to move, reside, and work freely within the territory of member states secured by the Maastricht Treaty.[28] From circular migration to cross-border living and working, the right to freedom of movement and residence has enabled the enactment of mobile citizens who are not contained in national borders. This reveals two entangled aspects of the performativity of definitions. The first concerns how the definition normatively establishes non-mobile, emplaced people as the basis on which populations are enacted. This is the legacy of which this definition is a part – the concept of 'people' as an 'immobile, sedentary, and enclosed body politic bounded within a territory' (Isin, 2018: 116). In turn, the definition enacts mobile people as exceptional. A second concerns how being categorised as usually resident is not a phenomenon independent of political decisions but also an effect of them. Whether rights to move and reside or labour market strategies, political decisions effect the enactment of mobility and, in the case of the EU, the naming of the 'mobile citizen.' In turn, the naming of the category and a kind of people are emerging at the same time (Hacking, 2002). This is most apparent when the statistics on, and naming of, EU mobile citizens cited above are used to make rights claims about representation in the European Parliament, for example. One commentator who advocates for the recognition and rights of the category of mobile citizens cited the Eurostat figures to argue that they are 'the most "European" [but] are the least politically represented in Europe' (Alemanno, 2019). They argued that to exercise the right to vote in European parliamentary elections, citizens are expected to register in their country of residence, which is often administratively difficult for mobile Europeans. The making of such rights claims is an instance of what Hacking calls a subversive or unintended effect of categories. Another is how statistics on 'migrants' have reinforced debates on

migration as exceptional and a problem that must be managed (Sager, 2018). In these ways, the performative effects of defining categories such as usually resident, mobile citizen, or migrant bring attention to how government policies, naming, and political claims interact.

But allocating people to a single, nationally bounded territory does not simply involve data practices that define the category of usually resident. A range of data practices are required to then encode people into the category. In the following section, we analyse one such data practice, that of deriving who should be allocated to the usually resident population.

Deriving Usual Residents

> The Dutch demographic statistics are entirely based on the Dutch population registers. As such, describing demographic statistics in the Netherlands basically boils down to describing the definitions and practices used in the population registers (Statistics Netherlands, 2016: 8).

It is through the process of encoding that individuals are allocated to the category of usually resident and held together as part of a population. However, encoding involves data practices that are often 'hidden in routinized chains of production' which also involve decisions 'laden with further consequences' (Desrosières, 1998: 247). For example, in relation to law, Desrosières notes that judges do not simply apply the law but consider arguments and debates of public proceedings and interpret rules and jurisprudence that have accumulated from previous cases. Similarly, statisticians do not simply apply or implement a definition, but engage in numerous data practices that draw on their experience and involve judgements about encoding people in a category. These practices are also

part of method assemblages that make up national methods, which is well illustrated in the conduct of feasibility studies undertaken by NSIs from 2015–16 to evaluate the implementation of the harmonised definition of usual residence for the production of annual EU demographic statistics on population and vital events (births, deaths).[29] Given the multiplicity of methods NSIs use to produce not only decennial censuses but also annual demographic statistics, the feasibility studies sought to identify problems and differences in implementing the definition. They also coincided with the adoption by the ESS in 2017 of a vision and strategy on post-2021 EU population censuses, which called for more frequent annual population statistics to accompany decennial censuses based on a common international definition of the usually resident population (Eurostat, 2017: 3).

Reports on the studies revealed that while all EU member states declared problems implementing the definition, it was deemed most problematic for countries using population registers to produce both census and demographic statistics (EC, 2018: 8). For example, in some countries, the population base for the purposes of population statistics and censuses is equivalent to a person's registered place of legal residence, which is different from that of the harmonised definition. We describe some of those differences below but here note two issues. First, unlike questionnaire-based censuses whereby being usually resident can be determined by asking people directly, register-based censuses must determine this indirectly. Second, and relatedly, this highlights an important difference between statistics based on registers which serve government administrative purposes and statistics based on questionnaires that are specifically designed to serve statistical purposes.

How then are these differences resolved? Given that countries are increasingly adopting registers to conduct their

censuses, addressing these differences is deemed critical to ensure harmonised and comparable European population statistics now and in the future.[30] However, the objective of the feasibility studies was not to change national register-based methods and practices, which involve method assemblages consisting of a hinterland of administrative rules and practices that have historically developed to encode people for the purposes of governing them. Rather, the feasibility studies experimented with data practices that would enable estimating the numbers of usual residents indirectly by adjusting data from population registers. We discuss our observations and analyses of one such data practice that statisticians in a population register-based country, Statistics Netherlands (SN), engaged in as part of its feasibility study, that of deriving usual residents.

Since 2001, population register data produced by municipalities are used to constitute the population base of SN's demographic and census statistics. Population register data was adopted after wide resistance to the 1971 full enumeration census (a traditional questionnaire-based census), followed by a two-decade hiatus in the national census programme. Among the topics of public contention in 1971 were issues of privacy and concerns about the use of census data beyond statistical purposes (such as to verify administrative records). Following this resistance, response rates dropped considerably, while political concerns started to emerge about the high costs of full enumerations. A period of legal reform ensued, during which full enumeration censuses were called off. To satisfy the demands of the European Community's census programme, a 'compensation programme' (Van Maarseveen, 2002: 94) was then initiated in which demographic data for 1981 and 1991 were collected from the municipal population registers, supplemented by data about education and employment from the Labour Force Survey and data from other

surveys (e.g. about housing) to produce a centralised population register dataset (PR data set).[31] While the adoption of the PR dataset at SN's main population base initially led to strong critiques from the research community (Van Maarseveen, 2002), ongoing efforts to improve and test its population coverage and reliability have supported its continuation along with justifications based on cost-efficiency. These efforts have included technical upkeep, quality committees, and regular evaluations, which have facilitated its acceptance in the Netherlands as a method for producing national population statistics.

At the outset, the SN feasibility study identified a major issue that other register-based countries also reported: whereas the usually resident population is defined for statistical purposes, the population base for register-based countries is founded on rules and procedures that serve government administrative purposes. Municipal population registers are set up by Dutch municipalities according to the national Basic Population Register Act (PR Act), which outlines rules and admission criteria for registering individuals. Each registered person is assigned a unique personal identification number, known as a Citizen Service Number (in Dutch: BSN), which governs access to various services and rights such as national taxation and insurance, education, and health care. In fact, inclusion in the municipal population registers establishes who is treated as a full member of Dutch society, as exemplified in the statement that 'it is nearly impossible to live a regular life in the Netherlands for people who are not registered' (Statistics Netherlands, 2016: 40). At the same time, it establishes who are the subjects of governing and forms the basis of Dutch national policymaking.

Elaborate rules are set out in the PR Act that begin with the requirement that everyone born in the Netherlands is

registered and that immigrants must register five days after arrival and must meet the following criteria:

1. their stay is legal according to the Immigration Act (for people who do not have Dutch nationality);
2. the intended stay is at least two thirds of the forthcoming six months;
3. the person is properly identified. The latter means that a valid passport or other official document is shown for identification (Statistics Netherlands, 2016: 40).

Of note is that the category of immigrant includes EU mobile citizens (such as those discussed in the previous section) in addition to people immigrating from non-EU countries. What connects the two groups is the stipulation that a person intends to stay for more than four months. Amongst others, these rules are at odds with the definition of usually resident, which specifies a(n) (intentioned) 12-month residence. Additionally, registration rules require that people must prove their legal status (e.g., that they are an EU citizen, or have a visa) and have a legal address, two conditions that are not specified in the definition of usually resident. One consequence is that people who are illegally in the country (e.g., without a visa) are usually excluded from the municipal population registers, and thus also from SN's PR dataset. Furthermore, people may fail to register in the municipal population registers when they arrive or fail to deregister when they leave the Netherlands, which also affects who is included or excluded in the PR dataset. Finally, like special cases (and exceptions) of the harmonised definition of usually resident, so too are specific rules provided for how municipal authorities are to deal with cases such as tertiary students, asylum seekers, and homeless people and their inclusion or exclusion in the municipal population registers.[32] These more or less mirror issues were discussed in the previous

section. While all methods potentially exclude some special cases – questionnaire-based censuses, for example, only include people who have a residence and thus exclude many homeless people – those exclusions are variously acknowledged or acted upon and here we examine how they play out for a register-based census.

In the remainder of this section, we discuss how SN's feasibility study addressed what it described as 'gaps between national and usual residence population definitions' and how those gaps were filled by experimenting with two estimation methods: the catch-recatch method and the micro-register method. Rather than simply applying or implementing these methods, we describe how the experiments involved data practices that included decisions, assumptions, and judgements about who was or was not likely included in the municipal population registers. These data practices did not change the PR data set, rather, they resulted in estimations for Eurostat that would allow SN to maintain it and its current practices.

Catch-recatch: Deriving the Whole

The catch-recatch method (CRC) assumes that there are people residing in the country but are not registered, which is understood to be more common in countries with high levels of immigration. Statistics based on municipal population registers are thus understood to undercount the 'real' population and the CRC method is designed to address this by 'catching' nonregistered people. To do so, it involves comparing the PR dataset with one or more other administrative registers. One of the statisticians working on this topic drew on a piece of paper the following graphics to explain the (considerably simplified) basics of the CRC during a 2015 interview:

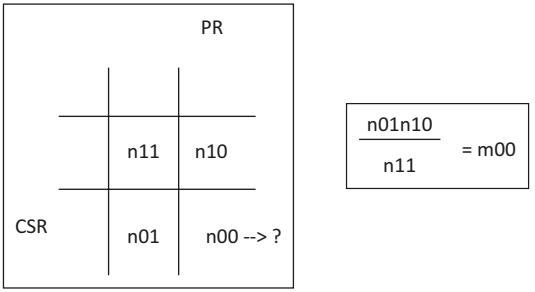

Figure 3.2 Graphic of CRC Method

The 'PR' is the PR dataset, and the 'CSR' the Crime Suspect Register that contains information on all persons reported to the Dutch police force as crime suspects, which may include people who meet the definition of usually resident but who have not registered. The rationale for using the police register was that, in theory, it can include anyone present in the country, legal or illegal. Moreover, it can contain people who would not be present in any other register, as one statistician explained at a meeting, while most other registers overlap with the PR dataset.[33] Another statistician at a different meeting elaborated that the roots of the method are to be found in biological population counts. In biology, the method involves comparing two consecutive samples from a single population (Figure 3.3).[34] Animals captured in the first sample are marked and then sent back into the field. A second sample is then captured and the number of newly captured as well as recaptured (i.e., marked) animals is counted. Based on those counts, an estimate is made of the total population.

The method has also been used to estimate human populations based on two or more samples. In the SN version depicted in Figure 3.2, two registers function as the samples

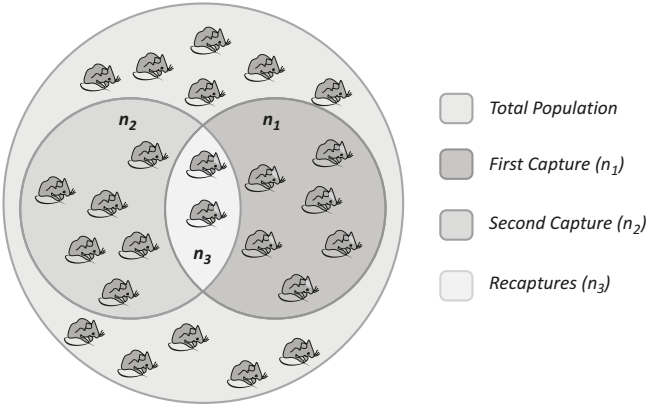

Figure 3.3 Biological Population Assumption of CRC Method[a]
[a]Fieldnotes. Drawing based on internal presentation at Statistics Netherland, 2015

(n1 and n2 circles in Figure 3.3); n00 is the area outside of the n1 and n2 circles, which is the population of people not included in either register.[35] However, the CRC is a mathematical estimation model based on many assumptions such as that there are no erroneous 'captures'; that is, whereas all animals have more or less equal chances of being caught in a trap, this cannot be assumed for persons 'caught' in registers. The model must therefore be adjusted to make it work for the purposes of estimating the under-coverage of a human population.

For SN, these adjustments involved the data practice of running the model multiple times for different groups deemed most likely to have not been registered. Since this is a time-intensive practice and there are limitations to computing capacity, the model could only be run a limited number of times. As a consequence, the decision was made to focus on unregistered immigrants (legal or illegal, EU citizens or

non-EU visa). They were deemed the most likely unregistered groups based on their geographical proximity, general size, and 'relevance' to the Netherlands (e.g., Polish nationals were identified because they are one of the largest groups of EU nationals residing in the Netherlands). Seven groups were so distinguished based on size and similarity of visa requirements: ' "EU15", "Polish", "Other EU", "Balkan and other former Soviet states", "Turkey, Morocco and others", "Iraq and others" and "Other western countries" ' (Statistics Netherlands, 2016: 55).[36] The aim was not to learn about the particulars of each of these groups, rather, distinguishing them was a means to an end: achieving a better estimate of under-coverage. As one statistician explained at a seminar, this method is 'about the maths, explaining the groups and their differences is not that interesting'. What works best for the method, in other words, is the criterion.

Other assumptions were adopted to address an expected under-coverage of specific groups such as young people (0–14) and elderly people (65+), which could not be addressed using the CRC method. Adjustments were made to the estimates for each of these groups, which SN acknowledged introduced uncertainty as they are unverifiable or based on educated guesses (e.g., that all children attend primary school). Taken together, the method involved data practices of running of the CRC model and making these and other adjustments, which resulted in an 'interval estimate of the under coverage' reported in the feasibility study: 113,000 to 136,000 people.

The CRC involved data practices that derived usual residents at the aggregate level, that is, by estimating the *number* of people missing from the PR dataset based on a mathematical estimation model rather than identifying and encoding each person who meets the definition of usually resident. As such, it derived the whole – that is, the number of usual residents – by

estimating how many people are likely not included in the registers, that is, the under-coverage. However, it did not identify people in the population registers who do not meet the definition. For this a second method was trialled to estimate over-coverage through data practices of deriving which *individuals* should not be encoded as usual residents.

Micro-register: Deriving the Parts

The micro-register data method (referred to here as the micro method) aimed to identify people included in the PR dataset but who, for instance, no longer live in the Netherlands. Because it is suited to determining who has moved out of the country, variations of it are adopted by countries with high levels of emigration. In addition, the micro method could also provide an estimate of under-coverage and in this way serve as a plausibility check for the CRC estimation of under-coverage (and vice versa).

As statisticians explained in an interview, the method 'operates at the level of individual records'. Or, as explained in the feasibility study report: 'The essence of this method is that records are explicitly added to or deleted from the population register data' (Statistics Netherlands, 2016: 17). In other words, it involves changing who is encoded as usually resident. This is achieved through data practices that involve combining data from other administrative registers such as those on taxation, employment, social security, education, and health care. Like the CRC method, it involves making a series of assumptions about which persons in these other administrative registers should or should not be encoded ('added or deleted') as usual residents. For example, two groups not required to be in the population registers but identified in the harmonised definition of usually resident as 'special cases' to be included are

diplomats and asylum seekers. A register maintained by the Ministry of Foreign Affairs was used to identify the former; and that maintained by the Immigration and Naturalisation Service (IND) for the latter. Other administrative registers also provide a way to identify people who may be present in the population register but should not be encoded as usually resident, for instance, because they emigrated but failed to deregister.

Statisticians combined data from the population and administrative registers into a new dataset to identify these and other special cases likely to be missing.[37] For each individual the registration start and end dates from each register were combined so that a person's total time present in the Netherlands could be determined. When doing so they detected overlaps and absences between registers and decided which data sources to prioritise in cases where there was a contradiction (for instance, someone registered as working in the Netherlands in the Employment Register but registered as having emigrated to Spain in the population register). Together, these operations were referred to as 'flattening the files': 'horizontally' aligning an individual's mutations in the register (e.g., an address change) so a timeline could be created. The result, as sketched by a researcher during an interview, looked like this (1, 2, 3 represent persons, see Figure 3.4):

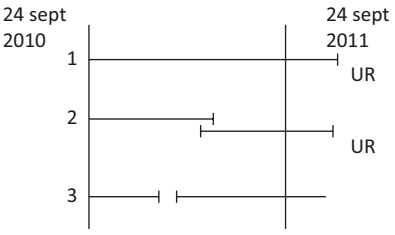

Figure 3.4 Determining a Timeline for Three People

Persons 1 and 2 would be derived as usual residents but the case of Person 3 raised questions during a project group meeting. If a person leaves for a month, should they be included? Some statisticians stated that the difficulty answering such questions is that Eurostat regulations and NSI interpretations often deviate. While the regulations state a usual resident is a person who has lived, or intends to live, at their place of residence for a continuous period of time of at least 12 months, NSIs vary in their interpretations of which periods of absence are permitted. Questions of interpretation extend more generally to the meaning of the definition of usual resident versus what seems 'right'. For instance, one statistician asked: 'what [do we] do with the group of people who live in Spain six months in the year, and in the Netherlands for another, where do they belong?' Another commented that 'you could also ask what people feel', or start a separate European population register to accommodate measurements of migration. In general, they pointed to how the complexities of mobility discussed in the previous section introduce problems of measurement even when a carefully crafted definition of usual resident along with special cases is provided.

When 'making decisions' about how to address these complexities, statisticians repeatedly mentioned that while crucial, this was a difficult and time-consuming part of the feasibility study. The project team undertaking the micro method thus decided to confine its work to six identified 'variants' and then estimated over-coverage according to each (see Table 3.1):

Table 3.1 Two of the Six Variants[a]

Variant 2. Both deregistration from the PR less than 365 days and in between registration in ER of less than 32 days were neglected. In case of incompatible data (example: an individual is registered in PR and receives monthly allowances abroad) the presence abroad was followed. People who reside in the country for more than 365 consecutive days are considered usual residents.
Variant 6. People who reside in the country for more than 365 days (not necessarily consecutive) during two consecutive calendar years are considered usual residents. In case of incompatible data (example: an individual is registered in PR and receives monthly allowances abroad) the presence abroad was followed.

[a]Source: Extracted examples from Statistics Netherlands 2016, 22–23

The possible variations demonstrate the uncertainties of interpreting and implementing the harmonised definition. They also show how defining who is usually resident does not end with agreement on a definition but also happens through data practices that seek to encode people. According to the project leader these variations provoked 'tough discussions'. For example, there was much debate about how to evaluate and accept the results of applying one of the variants. One statistician argued that if a variant led to a result of 12 million usual residents that would not be plausible given the Netherlands has 17 million people in the PR dataset. But a result of 17.2 or 17.3 million would be more plausible. On which grounds then should a decision be taken? On the one hand, plausibility was asserted as an important criterion for such decisions.

On the other hand, relying too much on plausibility would lead to a confirmatory bias towards the PR dataset based on municipal population registers. They offered that 'there really isn't any way to solve this' because there is no way of conclusively knowing who resides in the country. The conclusion was that variant 2 best fit the harmonised definition, which resulted in the identification of about 20,000 people 'unjustly included' and that 'some 104 thousand usual residents [were] not covered by the [municipal] population registers' (Statistics Netherlands, 2016: 24).

The results of the CRC and micro methods were then combined by first evaluating their relative plausibility on over-coverage including the acknowledgement that many of their assumptions were impossible to validate such as those on which the variants were based. Nevertheless, the report arrived a net under-coverage in the number of usual residents: 136,700. Thus, on 1 January 2013, the total usually resident population for the Netherlands was estimated at 16.9 million people. The report concludes that executing either or both methods annually would be too cost- and time-intensive and therefore advised that estimates be repeated once every five years and an extrapolation method be used for the intervening years.

Sustaining the National

The above can be read as a struggle over SN's epistemological position: the Netherlands population is defined by and equivalent to the number of people in its municipal population registers. In other words, its '"reality" is nothing more than the database to which they have access' (Desrosières, 2001: 346). This is a more or less pragmatic choice from the point of view of statisticians working with these methods. They are always

aware that it can be questioned and destabilised, as demonstrated from the issues raised while developing the micro method. By using various types of register-based data (from population and other administrative registers) statisticians developed methods and related data practices to derive usual residents and in turn satisfy international definitions.

However, their choices and what eventually may become stabilised conventions have consequences. To be registered is to be sedentary and settled in the country but more significantly it is a governmental necessity because 'it is nearly impossible to live a regular life in the Netherlands for people who are not registered' (Statistics Netherlands, 2016: 40). While affording considerable advantages and rights, the assumption is that only people who are fully incorporated into the state administrative system are a legitimate part of the population. In this way, SN's position is not only epistemological but also political as evident in the normative assumptions about the modes of living of groups that do not fit the logic of the municipal population registers such as homeless people, unregistered immigrants (which also include EU citizens), or 'illegal and undocumented persons.' Those normative assumptions include suggestions that many homeless people are also illegal (Statistics Netherlands, 2016: 22) and that homelessness is not a noteworthy problem in the Netherlands despite rising numbers (Nieuwenhuis, 2019).[38] So, while the data practice of deriving sought to account for these and other special cases to meet the harmonised definition of usually resident, it was deemed costly, time-consuming, and wrought with assumptions that could not be validated. In other words, unlocking the definitions, rules, technologies, priorities, and histories of the method assemblages that make up SN's PR dataset stand in the way of implementing a harmonised definition. However, acknowledging this reduced the issue to a technical effort and sustained the national method rather than bringing it

into question in the face of a Europe (and world) characterised by multiple modes of living. At the same time, while the data practices were treated as technical efforts, they also opened up normative issues such as assumptions about who belongs and does not belong in the population, issues that statisticians deem as 'political' and which are generally understood to be outside of their professional jurisdiction. In this way, data practices invented to implement and sustain international categories such as usual residents bring to the fore normative and political assumptions, which might otherwise remain obscure. However, the data practice of deriving deflected attention away from these questions by sustaining the national method and SN's long-standing position that despite its over- and under-coverage, the municipal population registers are its 'reality'.

Conclusion

By documenting the complexities of mobilities and methods, the objective of this chapter is not to advocate for a more harmonised and fitting definition of who is usually resident. Neither is it to criticise or advocate a method and its associated data practices. Rather, it is to identify the consequences of the persistence of the national order of things, which is not only to be found in a sedentary bias and the assumption that borders contain populations, but also in the methods and mundane data practices through which populations are enacted.

The data practices of defining and deriving discussed above, while noting the problems of the complexities of mobility and multiplicity of methods, engage in elaborate work to sustain these governing logics. From special cases that make exceptions, to rules that make exceptions to exceptions, much work is required to hold differences together and sustain the category of usually resident. Such data practices have many

political consequences. For one, they make mobilities exceptional and, in turn, objects to be (potentially) problematised as witnessed in political debates about migration numbers that have come to influence elections and referendums. However, those debates typically focus on questions of the accuracy of numbers and the best methods for reflecting an assumed independent reality rather than the underlying assumptions of the data practices that produce those statistics.

While being usually resident is not based on citizenship, the category of mobile citizens introduces a valuation of certain forms of mobility that are connected to the European project. However, as suggested in Chapter 1, the European project involves political technologies such as monetary, education, and cultural policies as well as population statistics through which Europe, understood as a people and a population, is constituted. But just as other governing technologies are locked-in to the national order of things, so too are censuses and population statistics wedded to their national frames. It is this inheritance that European – or international – practices and politics confront.

References

Ahas R, Silm S, Järv O, et al. (2010) Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones. *Journal of Urban Technology* 17(1): 3–27.

Ahas R, Tiru M, Saluveer E, et al. (2014) Mobile Telephones and Mobile Positioning Data as Source for Statistics: Estonian Experiences. Available at: https://bityl.co/555T (accessed 1 May 2020).

Alemanno A (2019) There Are 17 Million Mobile EU Citizens: We Deserve a Political Voice. *The Guardian*, 5 January. Available at: https://bit.ly/2DLXatj (accessed 5 February 2019).

Anderson B (2006) *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.

Bakewell O (2007) Keeping Them in Their Place: The Ambivalent Relationship between Development and Migration in Africa. *International Migration Institute* Paper 8. Working Papers.

Bowker GC and Star SL (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

CES (2013) *Population Definitions at the 2010 Censuses Round in the Countries of the UNECE Region.* Paper presented to the Meeting of the Group of Experts on Population and Housing Censuses. Geneva: UNECE.

CES (2015) *Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing*. ECE/CES/41. Geneva: UNECE.

CES (2016a) *Defining and Measuring Circular Migration: Final report of the Task Force on Measuring Circular Migration*. Geneva: UNECE.

CES (2016b) *Defining and Measuring Circular Migration: Final report of the Task Force on Measuring Circular Migration* (Draft 20 January 2016). Geneva: UNECE.

Desrosières A (1998) *The Politics of Large Numbers: A History of Statistical Reasoning* (ed. RD Whitley). Cambridge, MA and London: Harvard University Press.

Desrosières A (2001) How Real Are Statistics? Four Possible Attitudes. *Social Research* 68(2): 339–355.

Duchêne-Lacroix C (2014) *Multilocal Living Arrangements in Switzerland*. Presentation to the British Society for Population Studies Meeting on Usual Residence. London School of Economics and Political Science.

EC (2017) *Commission Implementing Regulation (EU) Laying Down Rules for the Application of Regulation (EC) No 763/2008 of the European Parliament and of the Council on Population and Housing Censuses as Regards the Technical Specifications of the Topics and of Their Breakdowns*. COM(2017) 1728 final. 22 March. Brussels.

EC (2018) *Report from the Commission to the European Parliament and the Council on the Implementation of Regulation (EU) No 1260/2013 on European Demographic Statistics* COM(2018) 843 final. 14 December. Brussels.

European Migration Network (2011) *Temporary and Circular Migration: Empirical Evidence, Current Policy Practice and Future Options in EU Member States*. Available at: http://bit.ly/2QPPiOQ (accessed 1 November 2019).

European Parliament (2018) Free Movement of Persons. Available at: https://bityl.co/555m (accessed 30 April 2019).

Eurostat (2011) *EU Legislation on the 2011 Population and Housing Censuses: Explanatory Notes*. Eurostat Methodologies and Working Papers. Luxembourg: EC.

Eurostat (2017) *Feasibility Studies on 'Usual Residence': Annex 1*. Working Group 'Population and Housing Censuses.' Luxembourg: EC.

Eurostat (2018) *EU Citizens Living in Another Member State: Statistical Overview*. Available at: https://bityl.co/555o (accessed 10 October 2019).

Eurostat (2020) EU Labour Force Survey Microdata 1983–2019, release 2020, version 1. *Eurostat: Your Key to European Statistics*.

Fries-Tersch E, Tugran T, Markowska A, et al. (2018) *Annual Report on Intra-EU Labour Mobility*. Brussels: European Commission Directorate General for Employment, Social Affairs and Inclusion.

Hacking I (2002) *Historical Ontology*. Cambridge, MA: Harvard University Press.

Isin E (2018) Mobile Peoples: Transversal Configurations. *Social Inclusion* 6(1): 115–123.

Järv O, Müürisepp K, Ahas R, et al. (2014) Ethnic Differences in Activity Spaces as a Characteristic of Segregation: A Study Based on Mobile Phone Usage in Tallinn, Estonia. *Urban Studies* 52(14): 2680–2698.

Law J (2004) *After Method: Mess in Social Science Research*. London and New York: Routledge.

Liefbroer AC, Poortman A-R and Seltzer J (2015) Why Do Intimate Partners Live Apart? Evidence on LAT Relationships Across Europe. *Demographic Research* S19(8): 251–286.

Nieuwenhuis M (2019) Aantal Daklozen in Nederland Verdubbeld: 'Iedereen Kan Dakloos Worden.' *AD.nl*, 23 August. Available at: https://bityl.co/556G (accessed 16 December 2019).

OED Online (2018) allocate, v. Oxford University Press.

Pairaudeau N (2016) *Mobile Citizens: French Indians in Indochina, 1858–1954*. Amsterdam: NIAS Press.

Potter R and Champion T (2014) *The Concept of Usual Residence: Has it Reached its Sell-by Date? Report on a BSPS Day Meeting, 24 October 2014*. LSE: British Population Studies Association.

Prins K (2017) *Population Register Data, Basis for the Netherlands Population Statistics*. Bevolkingstrends, The Hague: Statistics Netherlands.

Sager A (2016) Methodological Nationalism, Migration and Political Theory. *Political Studies* 64(1): 42–59.

Sager A (2018) *Toward a Cosmopolitan Ethics of Mobility: The Migrant's-Eye View of the World*. Basingstoke: Palgrave Macmillan.

Scott JC (1998) *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven and London: Yale University Press.

Sheller M and Urry J (2006) The New Mobilities Paradigm. *Environment and Planning A: Economy and Space* 38(2): 207–226.

Statistics Netherlands (2016) *Usual Residence Population Definition: Feasibility Study, The Netherlands*. Den Haag: CBS.

Statistics Sweden (2016) *Register for Mapping Circular Migration*. Paper presented at the Conference of European Statisticians. Work Session on Migration Statistics. Geneva, CH, 18 May.

Ustek-Spilda F (2019) Statisticians as Back-office Policy-makers: Counting Asylum-Seekers and Refugees in Europe. *Science, Technology, & Human Values* 45(2): 289–316.

Van Maarseveen JGSJ (ed.) (2002) *Algemene Tellingen in de Twintigste Eeuw: De Methode van Onderzoek Bij Personen En Bedrijven*. Voorburg: Centraal Bureau voor de Statistiek.

4

Refugees and Homeless People: Coordinating and Narrating

Evelyn Ruppert and Funda Ustek-Spilda

Introduction

There are many 'persons between refugee and asylum status [that] end up hanging in the air.'[1]

The task force member stated that while there are difficulties harmonising the categorising and counting of homeless people they need to be included somewhere in the census.[2]

These opening quotes highlight some of the issues that come to constitute refugees and homeless people as special cases and exceptions. As we argue in this chapter, references to refugee statistics often conflate refugees, asylum seekers, and internationally displaced persons (IDPs), each of whom has a specific and distinct legal status. Furthermore, the category of refugee may (or may not) include people who are in the process of applying for or appealing a decision about their status, which is one reason why they are sometimes deemed hard-to-count or 'end up hanging in the air' as the opening quote states. Hence, when we use the term 'refugee' we do so recognising the problematic use of the category. As argued in Chapter 3, being constituted as hard-to-count is in part a consequence of the harmonised definition of the usually resident population, which is provided as a solution to the question of which bodies to count and where to locate them. Here we analyse how that definition contributes to the designation

of refugees and homeless people hard-to-count for different reasons. For refugees, these reasons begin with the complexity and multiplicity of national and international definitions, methods, and data sources such that the statistical category of refugee has come to variably include people with different legal statuses: they may have been granted refugee status, could be asylum seekers in the process of applying for refugee status, appealing the rejection of their application, or deemed deportable but their deportation cannot be enforced due to existing international agreements or human rights concerns. For these and other reasons, international conventions have been adopted (and which we detail below) that refer to 'refugees and refugee-related populations' (EGRIS, 2018: 19).[3]

For all these reasons, we recognise that referring to the category of refugees is problematic because it collapses myriad life situations and legal statuses. Rather than a matter of convenience we refer to the statistical category of refugee as it has become the predominant referent in government documents and media reports concerning mobile people fleeing their country of nationality and seeking international protection. Similar definitional complexities apply to the category of the homeless, which also collapses life situations including people living without a shelter but also who move frequently between various forms of temporary accommodation. These complexities are reflected in terms used to describe their status such as rough sleeper, street person, vagrant, transient, or people of no fixed abode. Each refers to a person on the move and problematises this form of mobility. We also recognise that the category is problematic by grouping people in relation to their 'lack' of a recognised residence, which underpins the categories of 'the homeless' or 'homeless people' that have come to dominate statistics.

Hard-to-count is a term generally applied to these and other special cases or exceptions such as higher education students, and circular and short-term movers who are identified as hard to locate, contact, interview, and persuade to participate in data collection methods (Tourangeau, Edwards, and Johnson, 2014: 5). For refugees, some of the cited reasons are their unstable and uncertain legal status, language barriers, an unwillingness to have contact with government authorities altogether or suspicions about the reasons for data collection (EGRIS, 2018: 49). In comparison, homeless people live under various conditions that are often erratic including sleeping rough, living in temporary accommodation, or in insecure or inadequate housing, which challenge definitions of usual residence because of the difficulty of locating and placing them at a defined address (Serme-Morin, 2017). For both groups, national governments thus adjust and develop special methods to make them countable and legible. However, these national methods vary considerably and often involve different definitions, data sources, and technologies, which lead to problems of international comparability. Data practices that seek to harmonise national data are thus introduced by international bodies such as the UN, UNECE, and EU and it is these data practices of 'output harmonisation' that we focus on in this chapter.

More specifically, this chapter addresses how the data practices adopted by international authorities come to enact refugees and homeless people as 'excess populations', a concept we adopt from Agier's (2018) interpretations of a migrant camp as a space that signals and simultaneously conceals refugees. Agier considers a migrant camp as an excess in three senses: it is extraterritorial (a delimited special physical space), exceptional (a legal and political regime that suspends citizenship), and exclusionary (it contains or repels people to

the borders of a society). Data practices similarly signal refugees and homeless people as 'above and left over from the sum of states' and in this sense excess populations (2018: 137). Like a migrant camp, data practices do this by first delimiting them as special, hard-to-count cases and exceptions in the definition of the usually resident population. That is, they are enacted as an excess by definition, an excess that nevertheless needs to be tamed and managed. Second, exceptional methods, rules, and quality standards must be adopted for the enumeration of these excess populations in order to render them legible and visible. That is, they are also enacted as an excess in relation to methods that must then be deployed, because they are hard-to-count using existing methods that are based on the definition of usual residents. In both senses – by definition and by method – refugees and homeless people are marked at the margins of a population, but who must still be made legible for the purposes of governing. Herein lies what we later argue is the double edge of enumeration: being counted is simultaneously a precondition of recognition and in turn government support, but also makes possible intrusive and potentially harmful governing interventions such as eviction, detention, or deportation.

We arrived at this interpretation of refugees and homeless people as excess populations when considering the French Statistical Bureau's (INSEE) enumeration in early 2016 of the French city of Calais including the so-called 'Calais Jungle' (which we will refer to as the Calais camp).[4] The camp was created in April 2015 as part of a public strategy on the part of government authorities to manage migrant populations by transferring and regrouping them into a single 'tolerated' camp outside of the city of Calais. Located at the border zone between France and Britain, it served as a staging post for

refugees seeking asylum in France or as a place of transit to Britain (Agier, 2018). On the one hand, INSEE's report on the enumeration notes that people living in the camp were enumerated on the same basis as homeless persons or nomads, that is, the place of enumeration was considered as the place of usual residence.[5] At the same time there was some question around if people in the Calais camp were part of the French population, given that their main goal was to leave France for the UK. Yet, on the other hand, INSEE noted that this comparison is unsatisfactory because people living in the camp

> are not really lacking a place of usual residence, insofar as the containers, tents and sheet metal shacks that they occupy constitute their residence, contrary to homeless persons, persons without a fixed address or nomads, who regularly change the place where they spend the night (INSEE, 2016: 8).

Another difference reported is that whereas enumerators comb assigned areas to locate and interview homeless persons, such direct data collection was not feasible for the camp as 'requirements for ensuring the safety of the enumerators were not met, and the population is very distrustful of and reluctant to respond to interviewers, notwithstanding the appropriate assurances given about the confidential nature of the data collected' (INSEE, 2016: 6). As a result, enumerators simply counted the number of people without collecting further demographic characteristics about them such as sex, age, or nationality. The simple count was conducted by teams who were allocated to different sectors of the camp based on aerial survey maps and who 'all at once, combed the area on foot for half a day, opened each tent and each shelter and counted the persons who were staying there' (7).

We will return to discuss the relevance of the Calais camp for our argument later. Of note here is that the enumeration of the camp involved adjustments to usual practices of enumeration to produce data about asylum seekers. It necessitated making practical adjustments that could deal with the contingent and complex conditions and situated relations of the camp through data practices that exceeded the usual. The adjustments echo practices of European colonial experts who enumerated and mapped Egypt in the early twentieth century (Mitchell, 2002). To deal with language, opposition and other 'difficulties', Mitchell documents how statistical knowledge had to be reformatted by practices of translating, moving, shrinking, simplifying, and redrawing information in order to format social processes as a national economy (115–116).

It is in this sense of excess – of a population and practices that overflow the usual –that we consider how data practices that aim to harmonise the definition of usually resident come to enact numbers of refugees and homeless people. They are excess in part because of the sedentary bias of the 'framework of national thought and action' (Agier, 2018: 138). Chapter 3 considered the implications of a 'sedentary bias' within both state and non-state practices which posits sedentarism as the norm and that 'people want to remain in their place' (Bakewell, 2007: 10). It argued that this underpins the constitution of mobile people as special cases or exceptions in relation to the definition of usual residence. However, for some mobile people, such as EU citizens who exercise their legal right to free movement, their status as special cases or exceptions is not problematised. In fact, identifying them and knowing their numbers is crucial for monitoring the political objectives of the European project. In contrast, refugees are subjects whose mobility is problematised due to their variable and uncertain legal statuses.[6] In the case of homeless people, they are mobile

in the sense that they 'regularly change the place where they spend the night' (INSEE, 2016: 8) and problematised as they often cannot be emplaced in a recognised usual residence. It is in these ways that the mobility of both refugees and homeless people contributes to their enactment as excess populations.

While quality standards guide and regulate the production of data for all population categories, this chapter analyses how different standards are adopted for refugees and homeless people that exceed and overflow the usual. As documented below, and which the enumeration of the Calais camp poignantly illustrated, known differences in quality are tolerated if data is 'good enough' and/or 'fit for purpose'. While various data practices are developed to achieve this, we examine two that do so through adjustments to what is known as output harmonisation: data practices that *coordinate* international numbers on refugees; and data practices that *narrate* numbers of homeless people in the EU.

Output Harmonisation: Making Data 'Good Enough'

As noted in Chapter 3, countries are 'free to assess for themselves' how to conduct censuses including 'which data sources, method and technology are best in the context of their country' (Eurostat, 2011: 9).[7] In other words, making data internationally comparable is not achieved by harmonising national enumeration practices and methods or what is referred to as 'input harmonisation'. Rather than harmonising all elements of data production, international organisations focus on harmonising the 'final statistical' product, the output (Baldacci, Japec, and Stoop, 2016: 8). How then is output harmonisation achieved? For international organisations, it is through three 'quality dimensions' for achieving comparability (1): the adoption of harmonised definitions for census categories

such as usual residence discussed in Chapter 3; the production of metadata that documents all data sources, definitions, and methods; and the production of quality reports on data sources that assess data relevance, accuracy, timeliness, accessibility, clarity, comparability, and coherence.

Output harmonisation thus accepts there are differences in national methods which in turn can have an impact on the quality and comparability of data produced. In relation to data adopted by international development organisations, Rocha de Siqueira (2017) argues that much attention is paid to unpacking these differences as 'imperfections', but that such critiques miss how imperfect data are not only accepted but become authoritative and authorise governing interventions. One source of their authority in the field of international development is the claim of their 'ever-perfectability' – that it is only a matter of time, technique, resources, and so on, before otherwise hard-to-count populations can be more accurately made into data. One consequence Rocha de Siqueira draws attention to is how the acceptance of imperfections leads to the acceptance of 'good enough' data and methods and in turn 'good enough governance' especially in the case of fragile states. That is, claims of imperfection and ever-perfectability reinforce each other and have performative effects insofar as good enough data comes to shape governing interventions. Gabrys, Pritchard, and Barratt (2016) approach 'good enough' data from a different perspective in their analysis of data generated by citizens as alternative ways of creating, valuing, and interpreting environmental datasets. Their study of air pollution data collected by citizens shows how data that do not meet scientific criteria of legitimation and validation can be treated as good enough for purposes such as initiating conversations, making claims and developing new understandings and approaches.[8] Like Rocha de Siqueira, it is not

the 'truthfulness' of data, but what conditions of production are deemed good enough and acceptable and can then come to have effects through their take-up.

For these reasons, questions of data quality should not be reduced to whether data are imperfect or meet established scientific standards. Rather, how data comes to be legitimised and made authoritative, and in turn have performative effects, is what matters. Where we depart from especially Rocha de Sequiera's argument is that rather than the criterion of imperfection, the criterion of quality covers all of the sociotechnical relations that bring data into being. That is, the object of quality is not about the data per se, but, as the conditions of output harmonisation express, it is about the justification and legitimation of the practices, procedures, and relations that have produced them. From adopting harmonised definitions to documenting methods and practices in metadata and producing quality reports, the transparency of these conditions is the criterion through which data is evaluated and assessed and whether it is good enough for the purposes to which they might be put.

Our second departure is that quality captures what is generally accepted: that methods, and in turn data, can always be improved along numerous rather than any single dimension and made better – quality is variable and a matter of degree, which, as documented below, cannot be singularly measured but can be narratively and normatively justified.[9] And finally, for international organisations that usually rely on national data, quality enables differing national practices to coexist while at the same time achieve what is deemed good enough for European and international comparability. For the UNECE, establishing the dimensions of quality enables NSIs to answer the question 'What does good look like?' and also enables stakeholders to discuss 'How good is good enough?' (UNECE, 2015). As such, quality is about evaluating whether

data is good enough; it is a practical and pragmatic approach to addressing the relation between evaluations of data and the purposes to which they might be put.[10] Such an approach requires adjustments to quality standards and a pragmatic application of a range of different data practices to achieve output harmonisation. In the following section we detail one of those data practices, that of coordinating often disparate, ad hoc, and incommensurable data from various national sources to make refugees legible and data about them comparable.

Coordinating Refugee Numbers

> Numbers of refugees, asylum seekers and internally displaced persons (IDPs) have increased rapidly in recent years. Moreover, almost every country in the world is affected by forced displacement either as a source, point of transit, or host of refugees, asylum seekers or IDPs, making forced displacement a global phenomenon. There is also an increasing number of countries affected by large movements of people, often involving mixed flows of forcibly displaced people and migrants, who move for different reasons but use similar routes (EGRIS, 2018: 13).[11]

So begins a report on *International Guidelines on Refugee Statistics*: that increasing numbers of people are being displaced and moving for different reasons around the world. While basing this observation on estimates from 1997–2016 produced by the United Nations High Commissioner for Refugees (UNHCR), the report argues that national, and in turn internationally comparable, migration statistics are incomplete or inadequate. The report was produced by an Expert Group on Refugee and Internally Displaced Persons Statistics (EGRIS), which was formed by the United Nations Statistical Commission (UNSC) in 2016 in response to the so-called European refugee crisis. Prior to that, refugee

statistics were not high on the European agenda. A series of events led to the establishment of EGRIS beginning with a joint report prepared by Statistics Norway and the UNHCR and presented to the UNSC's 2015 meeting. The report addressed the challenges of collecting, compiling, and disseminating statistics on refugees, asylum seekers, and IDPs and highlighted problems of data quality and international harmonisation (UN Statistical Commission, 2014). It was at this meeting that the UNSC acknowledged the need for a handbook on statistics on refugees and IDPs, which would serve as a practical guide for achieving data comparability. An international conference held in Turkey was convened to consider the report in October of the same year, which brought together NSIs, international statistical organisations, and humanitarian organisations as well as other technical experts working in the field of humanitarian protection.

At that meeting, the Turkish Statistical Institute (TurkStat) reported on a survey that revealed that almost all countries use their own national definitions for displaced populations; that while refugees are enumerated in population censuses, their status as such is not specified; and that asylum seekers are often excluded altogether from enumeration. One statistician who participated in the meeting noted what is also more generally reported in reviews of various state practices that '[t]here are multiple forms of refugees, and multiple definitions of who is a refugee, there are also different ways of referring to them.'[12] Statisticians from many NSIs noted other challenges such as language and translation issues, negative perceptions, and whether or not refugees are housed separately in camps, for example. These and other issues such as refugees' reluctance to have contact with government authorities and their suspicions about the uses of data were cited as reasons why they are deemed hard-to-count (EGRIS, 2018: 49).

At this and other meetings leading to the formation of EGRIS, a major reason for these issues and the lack of internationally comparable data was that international census regulations and guidelines do not include refugee status as a required or core category of data collection. While constituting one of the special cases in the definition of the usually resident population to be enumerated, the status of refugees (and the other listed groups) is not:

> Persons who may be illegal, irregular or undocumented migrants, as well as asylum seekers and persons who have applied for, or been granted, refugee status or similar types of international protections, provided that they meet the criteria for the usual residence in the country. The intention is not to distinguish these persons separately, but rather to ensure that they are not missed from the enumeration (EC, 2017a; UNECE, 2015: 75).

For the UNECE this distinction is manifest in its guidelines that differentiate between 'core' and 'non-core' topics (in other words, population categories such as age and educational qualifications respectively).[13] While providing recommended standards (e.g., definitions) for both, core topics are deemed 'essential' for international comparability while non-core topics 'less vital' and 'optional' (UNECE, 2015: 9). Because the topic 'reason for migration' – the motivation for a person's most recent migratory move such as to seek asylum or refugee status – is designated non-core, many states do not collect this data as part of their censuses (see Table 4.1). Furthermore, if they choose to do so, the recommendations specify that the topic of 'refugee background' be derived from existing data sources (such as administrative registers), rather than through direct enumeration (e.g., through methods of

Table 4.1 Core and Non-core Migration Topics[a]

Core Topics	Non-core Topics
Country of birth	Country of birth of parents
Country of citizenship	Citizenship acquisition
Ever resided abroad and year of arrival in the country	Country of previous usual residence abroad
Place of birth	Total duration of residence in the country
Previous place of usual residence and date of arrival in the current place	Reason for migration
	Place of usual residence five years prior to the census
	Population with refugee background (derived)
	Internally Displaced Persons (IDPs) (derived)

[a]Source: UNECE 2015, 197

self-identification on a questionnaire).[14] Participants at the first EGRIS meeting noted that the designation of 'reason for migration' as non-core and the absence of clear guidance on how to derive data are some of the reasons standing in the way of producing data good enough for international comparison.

While recommending that 'reason for migration' become a core topic of international guidelines in the future, the EGRIS report recognised that the different data sources and methods of NSIs constitute a major challenge. The report meticulously documented the strengths and weaknesses of these various methods for producing internationally comparable

refugee statistics. However, unlike other population categories for which output harmonisation is the solution to the multiplicity of methods, in the case of refugees, major differences were noted that come to enact them as excess populations.

For one, each method was noted to have 'refugee specific limitations' that exceed issues identified for other population categories. Many of these reflect the previously noted reasons for why refugees are deemed hard-to-count and which were identified as limitations to achieving a workable and harmonised definition (EGRIS, 2018: 29). Given the impossibility of harmonisation, the data practice of coordinating data was thus developed. It entails conceptual and classification activities that operate *post facto* on existing data to enable disparate definitions to be 'held together'. Desrosières asserts that this is a fundamental challenge and objective of statistics: how can aggregates of individual be made to hold in objective categories? (1998: 101). For refugee statistics, this involved the development of 'standard statistical concepts' and a 'workable classification' framework to 'enable identification of the refugee-related populations in data sources in ways that are both practical and cost effective to apply' (ibid.) (Table 4.2).

Each of these categories is described in detail in over two-and-a-half pages of the report that basically consists of legal administrative classifications such as those in Table 4.2. What this framework conveys is how one statistical category – refugee – subsumes numerous life situations and legal statuses that are variously recorded for administrative purposes in relation to national operational definitions (e.g., immigration or population registers). Such operational definitions are distinct and serve different purposes from those adopted, for example, by the UNHCR for the registration of refugees for the purposes of international protection. Relatedly, the statuses of refugees are variable and changing and involve more

Table 4.2 Classifications of 'Refugee and Refugee-related Populations'[a]

Persons in the country needing international protection	1. Prospective asylum seekers 2. Asylum seekers 3. Persons with determined protection status a. Refugees b. Admitted for complementary and subsidiary forms of protection c. Admitted for temporary protection 4. Others in refugee-like situations
Persons with a refugee background	5. Naturalized former refugees 6. Children born of refugee parents 7. Reunified refugee family members from abroad 8. Others with a refugee background
Persons who have returned to their home country after seeking international protection abroad	9. Repatriating refugees 10. Repatriating asylum seekers 11. Returning from international protection abroad 12. Others returning from seeking international protection abroad

[a]Source: EGRIS 2018, 30

than crossing a national border; they can be the result of birth, death, migration, being granted citizenship or changes in a person's international protection status (EGRIS, 2018: 34). Additionally, how those different legal statuses are operationalised, by which governing authority (border agency, government department), and according to what criteria and definitions are also variable. The production of refugee statistics can involve several ministries, departments, or agencies,

which hold different relevant data and who may or may not cooperate and share data with NSIs or make it fit for statistical purposes (EGRIS, 2018: 113).

For instance, in Norway, border agencies register asylum applications at the case level. This means that members of the same family are registered with the same case number and are not given individual case numbers. Consequently, at the data compilation level, the total number of persons who have registered for an asylum application might not be visible in data, as only the case numbers get enumerated. For refugees, Germany produces statistics based on initial applications while the UK does so on the basis of granted status.[15] This can result in major differences; in the case of Germany, after applying the definition of granted status, this amounted to a difference of 350,000 people.[16] Finally, many NSIs do not report data on people who applied for asylum but got rejected and there are many 'persons between refugee and asylum status [that] end up hanging in the air.'[17] Moreover, time lags between applications, decisions, appeals, and other legal and administrative processes mean that even when these data are reported, they do not necessarily pertain to the year when they are released or the same groups of people can be reported multiple times, depending on the status of their application.

There are several other challenges that are too numerous to summarise here and which occupy most of the 150 pages of the EGRIS report. They are in part the consequence of the uncertain and changing legal status of refugees and how they are (or are not) accounted for by national statistical systems, where administrative definitions are set in relation to their legal context but operationalised by methods developed for statistical purposes. In this regard, besides their physical mobility being at odds with a sedentary bias that in turn makes them invisible in population censuses, the changing and

variable legal status of refugees further renders them as hard-to-count (cf. Ustek-Spilda, 2019) and as excess populations. The data practice of coordinating manages this excess through a 'pragmatic' solution for holding disparate data together. It does not involve standardising or harmonising the statistical definition of refugees, but coordinating data produced by different methods and administrative classifications within and between national systems *post facto* to achieve 'consistency and efficiency' (EGRIS, 2018: 122). That is, it involves adjusting the quality dimensions of output harmonisation so that data can be good enough for the purposes of international comparability.

Such a pragmatic strategy is evident in the practices of many professional fields, such as those of medicine documented by Mol. In relation to disease, 'objects handled in practice are not the same from one site to another' and it is through specific practices that differences are coordinated so that objects can go under a single name and not clash (Mol, 2002: 6). Similarly, Bowker and Star (1999) have documented how the International Classification of Diseases (ICD) facilitates the coordination of classification work distributed amongst multiple agencies (135). They argue that rather than making different classifications interoperable, large-scale coordination entails satisfying conflicting requirements of distributed work. This, they argue, is achieved by compiling heterogenous lists much like that of the EGRIS framework. This enables the framework to then function as a 'centre of coordination' for intensive work practices that are distributed but which need to be coordinated to maintain an institutionally accountable order to which practitioners can orient and refer (Suchman, 1997: 41–42).

But, while the data practice of coordinating may be pragmatic, it has performative effects. This is especially so when

statistics come to be compiled in indicators to measure states' relative compliance with facilitating the integration of refugees as required by Article 34 of the 1951 Convention, which grants them the same rights as permanent residents or nationals (EGRIS, 2018: 81). Indicators of integration include variables such as legal and civil rights, demographic and migration characteristics, educational attainment, employment, social inclusion, and health status. Such indicators not only assess compliance but shape state and international organisation policies such as those related to the assessment of the characteristics, needs, and supports to refugee populations in comparison to other mobile populations.

In these ways, the data practice of coordinating is a solution to the problem of counting refugees, an excess population that needs to be made legible for the purposes of governing. It is in relation to this point that we return to the discussion of the Calais camp, where pragmatic strategies of enumeration can also have direct consequences for people seeking international protection. This was made powerfully evident with the eviction of residents and complete destruction of the camp in October 2016, some six months after the INSEE enumeration.[18] In the documentary *Calais Children: A Case to Answer*, Director Sue Clayton sheds light on the consequences for asylum-seeking children stuck in the border between France and the UK of being counted and not counted by statistical authorities (Clayton, 2017). On the one hand, being counted as part of the population of Calais meant that the French and UK governments could not ignore their presence and were responsible for providing shelter, food, and care to residents and especially unaccompanied children. On the other hand, being counted did not mean unaccompanied children could exercise their right to claim asylum – even when they had a right to apply and receive asylum, their claims were often ignored and/or rejected – leaving them

with the only option to disappear and not be counted.[19] *Calais Children* shows the struggle of volunteers to seek, find, and identify unaccompanied children who might qualify for asylum when it became public knowledge that the camp was going to be demolished.[20]

While official results of the enumeration are difficult to locate, one report states that the population was estimated to be 6,901 in September 2016.[21] Although the Calais camp made 'migrants' invisible by containing them some distance from the city, it also made them visible to not only government authorities, but also to voluntary organisations, artists, campaigners, researchers (Agier, 2018). From providing humanitarian care and legal support to raising awareness of the plight of its people, the Calais camp sparked political solidarities and mobilisations. It became a 'community' where inhabitants 'invented the hospitable town in France that the government refused them' (143). However, by becoming 'too visible, autonomous and political' the state then reacted with its destruction. That the census was conducted half a year before the eviction and destruction of the Calais camp suggests that enumeration can precede, prepare, and justify governing interventions that are not necessarily advantageous from the perspective of hard-to-count people who may have good reason to escape and evade governing authorities. Herein resides what we call the double edge of enumeration: to be counted may be connected to the exercise of social and political rights, but at the same time render people subjects of governing interventions.

The double edge of enumeration also exists for homeless people whose movements and lack of a recognised residence render them hard-to-count. Like refugees, they are enacted as excess populations about whom adjustments need to be made to quality standards to achieve output harmonisation and comparability within the EU.

Narrating Homeless People

While the data practice of coordinating is a pragmatic solution that addresses the varying definitions and methods of counting refugees, in the case of homeless people the data practice of narrating is a pragmatic solution to their absence in many EU enumerations. While also a response to the problem of different definitions and quality standards, it is a solution that not only marks homeless people as special cases, exceptions, and hard-to-count but in turn comes to enact them as an 'absent presence' in population statistics. As Callon and Law (2004) offer, an absent presence refers to something like a category of people that may be visible but then disappears, that is made absent, through specific practices. Yet, the category continues to have what M'Charek, Schramm, and Skinner (2014) refer to in relation to race as a 'ghostly' presence. We take up this understanding to analyse how homeless people are both visible but rendered a ghostly presence through the data practice of narrating.

Definitions of homeless people vary greatly across national statistical systems. While often reduced to the idiosyncrasies of national practices, variations in definitions are again in part a consequence of the internationally adopted definition of usual residence and its sedentary bias. That is particularly evident for census categories such as household status or housing arrangements and how homeless people exceed those definitions. These issues emerged in discussions of an ESS task force that reviewed the implementation of 2011 EU census regulations towards identifying changes for the conduct of the 2021 enumerations.[22] Their review evaluated harmonised census data which for the first time in EU history member states were mandated to provide for dissemination via the ESS Census Hub.[23] The Hub is promoted as providing consistently

classified, structured, standardised and methodologically comparable data produced by EU NSIs so that a census of Europe can be centrally assembled and accessed. Search queries enable users to produce tables that aggregate and relate population data from different countries according to combinations of three to eight topics (e.g., age, gender, marital status, citizenship) and at varying levels of aggregation.

The task force review, amongst other things, identified problems such as data irregularities and gaps in tables that can be generated via the Hub on various combinations of topics.[24] It then identified various solutions to these problems that could be adopted for the 2021 enumerations. In the case of homeless people, the solution to irregularities and gaps was the data practice of *narrating* homeless people in generic categories and descriptions and relegating their numbers to metadata reports. This entailed adjusting the quality dimensions of output harmonisation so that good enough data on homeless people could be produced. As we argue below, as a consequence homeless people will become an absent presence in population statistics as they can only be identified in narrative accounts and unevenly so due to variations in their inclusion or exclusion in the methods and data of member states.

Narrating Through Categories

A query to the Hub on household status for 12 member states illustrates some of the gaps and irregularities in the data that were discussed by the ESS task force. The query results are depicted in Figure 4.1 where a category defined as 'primary homeless people' appears in only four states, a flag indicates data is 'temporarily unavailable' for the UK, and a flag for Sweden – 'd' – states that 'Data on primary homelessness are not available' (though it is also not available for seven other states).

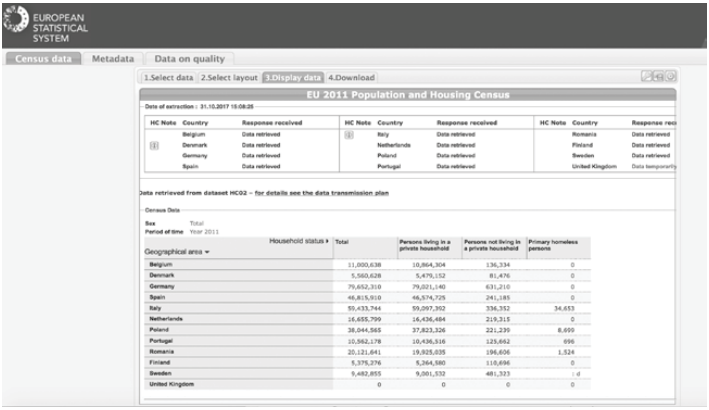

Figure 4.1 Census Hub Query Result[a]

[a]Source: Screenshot from ESS Census Hub query. See note 23. Retrieved 2 November 2017

Stepping back from the example of Sweden – which is not exceptional – and examining the data in Figure 4.1, variations in the counting of homeless people are impossible to evaluate. They may have been counted – or not – but by which states, why and how is not evident. Rather, a table with some empty cells is generated and possibly explained by metadata. However, the tab on metadata leads to a complex table of 21 textual fields. In relation to comparability, the metadata notes that 'Sweden has done a complete register-based census. This can impair the comparability of the data with censuses conducted in a traditional or a combined way.' Why and how so are not elaborated. Other queries generate additional flags and notes: break in time series; not available; confidential; definition differs; estimated; forecast; see metadata; not significant; provisional; revised; Eurostat estimate; low reliability; and not applicable.

Further investigation of the metadata reveals that the description of the topic of household status does not refer to homeless people but states that 'Persons not possible to link to a dwelling cannot form a household and are classified as "Persons not living in a private household, but category not stated".' In other words, homeless people may be part of this category though the reasons why and their numbers are not provided. Yet, data on the total 'primary homeless' people in the EU can be generated: 116,510. The number is underpinned by innumerable provisos, missing data, variations in methods, and so on that would be practically impossible to assemble and interpret.

The Hub made visible these gaps in data and variations in how member states defined, counted, or did not count homeless people in 2011. These variations became especially evident in discussions on the topic of household status (referred to above) which was defined according to two categories: people living in a private household (as a family, living alone or living with others) and people not living in a private household (in an institution or primary homeless). The category of primary homeless referred to 'persons living in the streets without a shelter that would fall within the scope of living quarters,' which excluded what is sometimes defined as secondary homeless: 'persons moving frequently between temporary accommodation.'[25] In 2011, only the 'primary homeless' category was included in the EU regulations and this was identified as one cause of missing data for some member states.[26]

A second cause noted in the task force discussions is that the data sources and census methods used by NSIs often do not include homeless people at all.[27] For example, countries that conduct register-based censuses cannot report this category because homeless persons do not have a registered

address. Given this, some statisticians argued that it does not make sense to require this data, otherwise some states would continue to report zeros and make it look like they do not have homeless people. Yet, other NSIs use a combination of methods including surveys or data collected by social agencies such as hostels. Indeed, there are many examples of the immense effort on the part of some statistical institutes and social agencies to generate data on homeless people but they do so according to different definitions and methods.[28] In the face of these issues, one statistician offered that the numbers of homeless people in most countries are negligible and that including the category would give a false impression that NSIs are able to count homeless people or can do so in a harmonised way. As in the case of refugees, due to these differences a practice had to be introduced to achieve output harmonisation that could produce good enough data on homeless people.

The solution for the 2021 regulations was the recommendation that the category of primary homeless be removed and homeless persons – however defined and counted – be subsumed in the generic category of 'Persons not living in a private household, but category not stated', and that 'including homeless people' be added to the description. In other words, the solution to variations was to do away with the separate category of homeless people and name and describe them as part of a generic category rather than number them separately. It is in this way that narrating practices enact homeless people as an absent presence by naming them as part of a category but not identifying and numbering them separately. Each of the terms that make up the category is defined and described further (e.g., private household) and in relation to other categories that make up household status (e.g., family nucleus), all of which do not apply to homeless people whose lives and conditions exceed all of these

categories. This is evident in the revised and final description of the topic of household status illustrated in Table 4.3. While describing the many life situations that constitute a household, subcategory 2.2 exemplifies how the social position of homeless people even exceeds description. The same can be said for the second form narrating that we discuss next, that of metadata.

Table 4.3 Technical Specification of the Topic of Household Status[a]

Household status	HST.L.	HST.M.	HST.H.
O. Total	0.	0.	0.
1. Persons living in a private household	1.	1.	1.
1.1. Persons in a family nucleus		1.1.	1.1.
1.2. Persons not in a family nucleus		1.2.	1.2.
1.2.1. Living alone			1.2.1.
1.2.2. Not living alone			1.2.2.
1.3. Persons living in a private household, but category not stated		1.3.	1.3.
2. Persons not living in a private household	2.	2.	2.
2.1. Persons in an institutional household		2.1.	2.1.
2.2. Persons not living in a private Household (including homeless persons), but category not stated		2.2.	2.2.

[a]From implementing regulation (EC) No 763/2008 of the European Parliament and of the Council on population and housing censuses as regards the technical specifications of the topics and of their breakdowns

Narrating Through Metadata

Sweden's metadata on household status: Persons not possible to link to a dwelling can not form a household and are classified as 'Persons not living in a private household, but category not stated'. Data on primary homelessness can not be classified. But according to a report from the National Board of Health and Welfare the number of primary homeless persons in 2011 is estimated to 4500.

UK's metadata on household status: Primary homeless persons are those that are identified as 'absolutely homeless', that is: people sleeping, or bedded down, in the open air (such as on the streets, or in doorways, parks or bus shelters), and people in buildings or other places not designed for habitation. These data have been flagged as unreliable due to the transient and hard to count nature of this population.[29]

A second way that gaps and irregularities in data on homeless people are managed is through the data practice of narrating through metadata. Typically defined as 'data about data' (Pomerantz, 2015: 19), metadata narratively documents the when, where, and how of data production and is integral to enabling data and individuals to be treated as equivalent.[30] It includes all of the descriptions included in the metadata tab and the flags, footnotes, and so on discussed previously. Metadata not only documents the conditions under which data was produced within an NSI but also that of other government administrative departments, which can differ considerably in their definitions and methods. Metadata makes data equivalent by smoothing out and accounting for (some of) the partiality of and differences between data. However, rather than resolving, metadata enables data to be treated as equivalent despite these differences.

As discussed previously, metadata is one quality dimension of output harmonisation. It is required for all population

statistics and prescribed in a separate EU regulation. In relation to homeless people, member states must include the specific definition used (such as who is included, how the estimate was derived, and from what sources), and definitions of primary and secondary homeless people, if applicable.[31] Yet, while metadata is prescribed to resolve differences, it is also contested. Discussions at one meeting reported that metadata was either too long (60 pages or more) or too short (not very informative for users) and some countries did not make full use of footnotes. Various discussions thus took place on how to revise the metadata regulation towards achieving greater standardisation. However, the different methods and practices of member states stood in the way of achieving this. One example was the requirement to report on all data sources, which is problematic for register-based countries which may use ten or more data sources and 'behind those there are about 100 that are used indirectly'. Even though the draft regulation defined a data source[32], this did not account for indirect sources. That is, numerous sites of production are part of the method assemblages that make up national statistical systems such as other departments and agencies of member states.

A further concern was that the quality of a source must be assessed. As one member reported, registers of external organisations are not harmonised in terms of data definitions, methods, architecture, and metadata. They are thus difficult to combine without considerable manual labour, decisions, and judgements, and information about the way the data is collected and treated is often not available. What the members' comments highlighted is that it is not only difficult to account for different methods; it is also impossible to account for the myriad practices, technologies, and decisions involved in the production of data. An excess of differences could not be contained or accounted for in metadata narratives and so the agreed solution

was that the regulation should state that only direct data sources be accounted for and assessed.[33] As Chapter 5 will argue, such excesses constitute a form of non-knowledge which, while consequential for the data produced, is often ignored and omitted.

Narrating data is thus negotiated and governed by agreed-to conventions about what and how conditions of production in the making of data can and must be recounted. Just as the making of data involves explicit decisions about what to make present and absent, metadata also involves decisions about what practices can and must be accounted for and described (Pomerantz, 2015). However, as in the case of data, differences are recognised and allowed to exist, are accepted, and what statisticians prioritise is accounting for difference, and being seen to do so, in relation to established protocols and standards.

But metadata serves as a way to deal with excess in another sense. The task force recommendation to subsume homeless persons in a generic category was eventually accepted by the ESS, which asserted that the practice ensured that homeless people would be included in the total population of a country. However, the ESS decided that member states should still be required to provide a 'best estimate' of homeless persons separately as part of the metadata and optionally break down this estimate into primary and secondary homeless persons. In doing so, the numbers of homeless people are relegated to metadata, which becomes not just 'data about data', but data in-and-of itself. So, while metadata is a container of differences, when such differences are excessive and output harmonisation is not possible, then data must also be relegated to metadata. In this way, metadata can be considered a placeholder which enables overlooking something by operating as a 'tool of forgetting, of putting to one side' (Riles, 2010: 803) and involving 'strategic ignorance' about the known limits of

quantification (Scheel and Ustek-Spilda, 2019). Metadata thus also establishes which social relations – such as being part of a household – can be made explicit. As Marquardt (2016) argues, homelessness is not only a social issue ignored by governmental data production, but an 'obstacle to conventional ways of data collection on the population' (301). It is a social condition that is also an obstacle to conventional forms of harmonisation, which the data practice of narrating is introduced to address.

This example of data on homeless people, while seemingly exceptional, involves data practices that are part of establishing equivalences between bodies that enable data to be good enough in the interests of output harmonisation. For example, same-sex marriages or consensual partnerships often get folded into opposite-sex categories of population data to address harmonisation (Grommé and Ruppert, 2019). These examples highlight the inseparability of data and social and political relations. To say so is not to suggest that data is a simple reflection. Rather, it suggests that harmonising data follows norms and values of dominant cultures such as people being part of a household. Just as the social existence of marginalised groups often exceed social and political recognition, so too are they a statistical excess. In this way, data practices such as narrating further push marginalised groups into the shadows. Race, for example, is well documented as a population category that has been excluded, othered, and silenced in various data practices that while not visible, does not go away but constitutes an absent presence (M'charek, Schramm, and Skinner, 2014). For Karkakis and Jordan-Young (2020), race is a 'ghost variable' enmeshed in science, medicine, and technology; to 'not notice' this is to do violence and thus methods for 'sensing ghosts' are integral to pursuing postcolonial, anti-racist, feminist science (775). So too is

sensing and acknowledging the absent presence of homeless people in population censuses.

It is in these ways that practices that make data good enough have knowledge effects such as enacting homeless people as an excess population and absent presence. This happens not only through the data practice analysed here but in other ways such as the explicit exclusion of vulnerable groups as a consequence of political agendas. Marquardt (2016), for instance, notes that since the 1980s the German national government has refused to collect statistical data on homeless people, which has led to advocates fighting for quantitative assessments as a crucial form of recognition. Such recognition (or lack thereof) can have many governing consequences as the data can inform policy decisions of the EU. One policy referred to by the task force is the distribution of social cohesion funding, which makes up the lion's share of EU spending; in 2014–2020 this amounted to €351.8 bn. However, by rendering one of Europe's most socially excluded groups an absent presence in population statistics, these data practices could lead to resource allocations that do not meet the relative needs of homeless people across member states.

Conclusion

The foregoing analysis and critique of data practices to make refugees and homeless people legible does not deny the immense effort and work of statisticians who recognise weaknesses and seek improvements in the production of statistics about them. Those efforts are well evidenced in the formation and work of international task forces and expert groups, in the appeals made at international meetings, and in the analyses and recommendations documented in reports that this chapter has considered. They are efforts that are usually well

intentioned and often underpinned by humanitarian aims. Indeed, the production of robust statistics is advocated in recognition of the consequences for the lives of vulnerable populations. This includes appreciation of the importance of statistics for informing policy and decision-making on programmes to support refugee populations and contributing to public debate and advocacy through more effective monitoring, evaluation and accountability (EGRIS, 2018: 13).

However, such efforts misrecognise two major political implications of the enumeration of marginalised populations. The first concerns what we have interpreted as the double edge of enumeration, which especially affects vulnerable populations on the move such as refugees and homeless people, and which the enumeration of the Calais camp poignantly illustrated. Being counted is simultaneously a precondition of recognition and government support but also makes possible often intrusive and life-changing governing interventions. As the pandemic of 2020 has underscored, legibility and visibility are also a precondition of care, which has been made vivid in countries failing to account for the health conditions of people on the move and whose exclusion may mean death (Milan, Pelizza, and Lausberg, 2020). Milan et al. note that while people on the move may prefer not to be counted so as to remain outside of government surveillance and action, under COVID-19 their exclusion from testing, tracing, and data collection on the part of many European countries has hindered access to care and relief services and also exacerbated the spread of the disease. One remedy offered by Milan et al. is that people on the move should be counted and incorporated in data practices but afforded the same data protection rights as citizens.

However, their remedy points to a second implication which is usually ignored and which we have underscored in this and the previous chapter. The enactment of refugees

and homeless people as excess populations 'above and left over from the sum of states' happens through data practices founded on a sedentary bias and framework of national thought and action. Such remedies, including those noted above, do not question this. Instead, relegating vulnerable groups that exceed norms such as having a usual residence to the status of hard-to-count only reinforces the sedentary bias of statistics. Rather than lower standards of quality and harmonisation or rendering the existence of people as an absent presence, our analysis suggests the re-evaluation of the sedentary and nationalist assumptions of population statistics. The identified weaknesses of population statistics about vulnerable and marginalised populations, the immense efforts to manage and sustain those statistics, and the social and humanitarian values they are intended to support rarely acknowledge that part of the statistical 'problem' resides here. That is, the problem does not originate in so-called hard-to-count populations but from systems of thought and practice that exclude, other, and silence particular groups whose social existence exceed the norms of dominant societies. If data justice means the right to be or not to be statistically present, then these fundamental assumptions of population statistics ought to be questioned.

References

Agier M (2018) *The Jungle: Calais's Camps and Migrants*. Cambridge, UK: Polity.

Aradau C and Huysmans J (2018) Assembling credibility: Knowledge, method and critique in times of 'post-truth.' *Security Dialogue* 50(1): 40–58.

Bakewell O (2007) Keeping Them in Their Place: the ambivalent relationship between development and migration in Africa. *International Migration Institute* Paper 8. Working Papers.

Baldacci E, Japec L, and Stoop I (2016) Improving the Comparability Dimension of European Statistics by Minimizing Unnecessary Variation. In: *European Conference on Quality in Official Statistics (Q2016)*, Madrid, 31 June 2016.

Bowker GC and Star SL (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Bulman M (2018) Government's Treatment of Child Refugees Under Dubs Scheme Broke Law, Court of Appeal Rules. Available at: http://bit.ly/2PTeCBr (accessed 16 December 2019).

Callon M and Law J (2004) Introduction: Absence – Presence, Circulation, and Encountering in Complex Space. *Environment and Planning D: Society and Space* 22(1): 3–11.

CLANDESTINO Project (2009) *Final Report*. Undocumented Migration: Counting the Uncountable. Data and Trends Across Europe, 23 November.

Clayton S (2017) *Calais Children: A Case to Answer*. Eastwest Pictures. Available at: http://bit.ly/36GlM2I.

Desrosières A (1998) *The Politics of Large Numbers: A History of Statistical Reasoning* (ed. RD Whitley). Cambridge, MA and London: Harvard University Press.

EC (2017a) *Commission Implementing Regulation (EU) Laying Down Rules for the Application of Regulation (EC) No 763/2008 of the European Parliament and of the Council on Population and Housing Censuses as Regards the Technical Specifications of the Topics and of Their Breakdowns*. COM(2017) 1728 final.

EC (2017b) *Community Statistics on Migration and International Protection*. Regulation (EC) No 862/2007.

EGRIS (2018) *International Recommendations on Refugee Statistics*. European Union and the United Nations. Available at: https://bityl.co/568q (accessed 4 November 2019).

Eurostat (2011) *EU Legislation on the 2011 Population and Housing Censuses. Explanatory Notes*. Luxembourg: Eurostat Methodologies and Working Papers.

Gabrys J, Pritchard H, and Barratt B (2016) Just Good Enough Data: Figuring Data Citizenships Through Air Pollution Sensing and Data Stories. *Big Data & Society* 3(2): 1–14.

Gentleman A (2016) 'It Is a Disgrace to Europe': Former Child Refugee Lord Dubs on the Calais Camp. *The Guardian*, 5 September. Available at: http://bit.ly/2sBqqjt (accessed 16 December 2019).

Goodwill R (2017) *Immigration: Written Statement by The Minister of State for Immigration*. UK House of Lords. Available at: http://bit.ly/37Nvvos.

Grommé F and Ruppert E (2019) Population Geometries of Europe: The Topologies of Data Cubes and Grids. *Science, Technology & Human Values* 45(2): 235–261.

Help Refugees (2016) Latest Calais Census. Available at: https://bit.ly/33MBQBv (accessed 11 August 2020).

Home Office (2018) Dubs Policy Statement: Section 67 of the Immigration Act 2016. UK Parliament. Available at: http://bit.ly/2RU7RSe (accessed 9 December 2019).

INSEE (2016) Some Thoughts on Taking a Census of Hard-to-count Populations Based on a Few Examples in France. In: *Conference of European Statisticians Group of Experts on Population and Housing Censuses*, Geneva, 28 September 2016.

Jandl M, Vogal D, and Iglicka K (2008) *Report on Methodological Issues*. CLANDESTINO Project. European Commission.

Karkazis K and Jordan-Young R (2020) Sensing Race as a Ghost Variable in Science, Technology, and Medicine. *Science, Technology & Human Values* 45(5): 763–778.

Marquardt N (2016) Counting the Countless: Statistics on Homelessness and the Spatial Ontology of Political Numbers. *Environment and Planning D: Society and Space* 34(2): 301–318.

M'charek A, Schramm K, and Skinner D (2014) Topologies of Race: Doing Territory, Population and Identity in Europe. *Science, Technology & Human Values* 39(4): 468–487.

Milan S, Pelizza A, and Lausberg Y (2020) Making Migrants Visible to COVID-19 Counting: The Dilemma. *openDemocracy*, 28 April. Available at: https://bit.ly/33jvxVQ (accessed 30 April 2020).

Mitchell T (2002) *Rule of Experts. Egypt, Techno-Politics, Modernity*. Berkeley, Los Angeles and London: University of California Press.

Mol A (2002) *The Body Multiple: Ontology in Medical Practice*. Durham, NC: Duke University Press.

Pomerantz J (ed.) (2015) Definitions. In: *Metadata*. Cambridge, MA: MIT Press, pp. 19–64.

Prestwood D (2019) The Emerging Picture of UK Homelessness and Rough Sleeping. In: *National Statistical: News and insight from the Office for National Statistics (UK)*. Available at: https://bit.ly/2DAmMMw (accessed 3 May 2020).

Ratner H and Ruppert E (2019) Producing and Projecting Data: Aesthetic Practices of Government Data Portals. *Big Data & Society* (July–Dec): 1–16.

Riles A (2010) Collateral Expertise: Legal Knowledge in the Global Financial Markets. *Current Anthropology* 51(6): 795–818.

Rocha de Siqueira I (2017) Development by Trial and Error: The Authority of Good Enough Numbers. *International Political Sociology* 11: 166–184.

Scheel S and Ustek-Spilda F (2019) On the Politics of Expertise and Ignorance in the Field of Migration Management. *Environment and Planning D: Society and Space* 37(4): 579–588.

Serme-Morin C (2017) Increases in Homelessness. *Homeless in Europe: The Magazine of FEANTSA*. Available at: https://bit.ly/2ATOBhP (accessed 15 November 2020).

Solletty M (2016) INFOGRAPHIE. L'explosion du nombre de migrants à Calais en un graphique. Available at: https://bit.ly/2Ch47VH (accessed 11 August 2020).

Suchman L (1997) Centers of Coordination: A Case and Some Themes. In: Resnick LB, Säljö R, Pontecorvo C et al. (eds) *Discourse, Tools, and Reasoning: Essays on Situated Cognition*. Berlin: Springer-Verlag, pp. 41–62.

Tourangeau R, Edwards B, and Johnson TP (2014) *Hard-to-Survey Populations*. Cambridge: Cambridge University Press.

UN Statistical Commission (2014) *Report of Statistics Norway and the Office of the United Nations High Commissioner for Refugees on statistics on refugees and internally displaced persons*. E/CN.3/2015/9. United Nations Social and Economic Council.

UNECE (2015) *Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing. Draft 4 March 2015*. Geneva: UNECE.

Ustek-Spilda F (2019) Statisticians as Back-office Policy-makers: Counting Asylum-Seekers and Refugees in Europe. *Science, Technology, & Human Values* 45(2): 289–316.

5

Migrants: Omitting and Recalibrating

Stephan Scheel and Funda Ustek-Spilda

Introduction[1]

In this chapter, we show that the enactment of the people of Europe, as well as their 'Others', features data practices that produce various types of non-knowledge. Theoretically, we combine material-semiotic approaches and the concept of data practices with insights and concepts from literatures on *agnotology*[2] (Proctor and Schiebinger, 2008) and ignorance studies (Gross and McGoey, 2015; McGoey, 2019). This framework allows us to show that, in the making of the people of Europe through data practices, the production of knowledge and non-knowledge are intertwined in multifarious and complex ways.

More specifically, we attend to how the production of non-knowledge features in the production of migration statistics and how statistical data are taken up and circulated in the field of migration management. This allows us to show that visualisations of statistical data in the field of migration management enact migrations, through the production of non-knowledge about the known limitations of existing migration statistics, as coherent, stable, and intelligible realities that can be managed according to certain policy priorities because they can be precisely quantified. To illustrate this point, we elaborate on two data practices namely: *omitting and recalibrating*. Shifting

the analytical focus to data practices allows us to attend to the specific activities through which this non-knowledge is produced and sustained. We show that the production of this non-knowledge happens both through data practices employed in the transnational field of statistics, data practices employed in the transnational field of migration management as well as during the transfer of knowledge between these different fields of practice.[3]

To develop and illustrate these points, we elaborate and reflect on our studies of one particular visualisation tool of statistical data on migration, the 'Global Migration Flows Interactive App' (GMFIA).[4] The GMFIA was created by the International Organisation for Migration (IOM), which is certainly the most influential actor in the transnational field of migration management. Although the GMFIA was deactivated in 2019, we are convinced that, given the similar attempts to develop visualisation tools for migration by various national and international organisations,[5] GMFIA continues to offer an emblematic example of how statistical data about migration gets taken up and is used in the field of migration management. In this chapter, we argue that the GMFIA and these other tools exemplify that actors in the field of migration management assemble seemingly exact and remarkably precise figures about stocks and flows of migration into visualisations to perform themselves as knowledgeable, competent actors with the expertise needed for successfully implementing projects of migration management.

One important effect of these visualisations is that migration is enacted, through the display of seemingly exact numbers about stocks and flows of migrants, as a reality that can be managed according to certain predefined policy objectives.

The production, visualisation, and display of precise numerical facts about migration is, however, intertwined with the production of non-knowledge about the known limits and challenges of quantifying migration through various data practices. This non-knowledge is produced in overlapping fields of practice as the reliance of actors in the field of migration management on exact and accurate numerical facts about migration as a source of expertise, legitimacy, and authority creates pressure for actors in the field of statistics to provide figures satisfying these requirements, that is, the requirements of the users of official statistics.

In the following we develop this argument in three sections. In the next section we elaborate our conceptual framework, most notably how we think about and conceptualise the relationship between expertise, data practices, and the production of non-knowledge. We also briefly discuss our understanding of the field of statistics and the field of migration management as distinct, but nevertheless overlapping and interacting social spaces. Subsequently, in the second section, we illustrate how methodological issues and limitations, such as the known divergence between reported emigration and immigration events or known inconsistencies of migration data resulting from methodological heterogeneity across NSIs, are ignored in migration visualisation tools, such as the GMFIA through the data practice of *omitting*. In the third section we explain, in turn, how statisticians enact migration as a coherent and stable reality by *recalibrating* migration data in time series after the implementation of methodological changes. Here, we pay attention to how non-knowledge about the known limits of quantifying migration is, to a large extent, generated through dynamics in and between different fields of practice.

On the Production of Expertise and Ignorance in Fields of Practice

As noted in Chapter 1, we understand fields of practice, such as the field of statistics or the field of migration management, as transnational fields of struggle where different actors compete over influence, funding, and agendas (Bourdieu and Wacquant, 1992). The demonstration of expertise, understood as highly technical, specialised knowledge, plays a central role in the competition over authority, funding, and influence in both fields. Put differently, the production and display of expertise operates as a stake in these struggles, a stake that could be described, in Bourdieu's (1991) terminology, as forms of cultural and symbolic capital.

In the field of migration management, the quest for expertise is not necessarily about improving the outputs of migration policy. A recent study of the EU's policy responses to the 2015 migration crisis notes, for instance, 'an ongoing and substantial "gap" between a significant body of evidence examining migration processes … and the policy response' (Baldwin-Edwards, Blitz, and Crawley, 2019: 2147). This conclusion supports Boswell's observation that actors of the field of migration management produce and display expert knowledge to increase their legitimacy and bolster their 'claim to resources or jurisdiction over particular policy areas' (Boswell, 2009: 7). In Bourdieusian terms, expertise then resembles a form of cultural capital that may allow an actor to attain the status of 'epistemic authority' in a given field. Epistemic authority affords the holder the capacity to shape agendas and attract funding in that field 'by virtue of possessing theoretical knowledge [and] being [regarded as] a reliable source of information or a skilful practitioner of a certain craft' (Geuss, 2001: 38). Importantly, expertise is not

a fixed attribute. Instead, organisations like the IOM need to constantly perform themselves as knowledgeable, competent actors through the publication of reports, maintenance of research units, and development of digital devices like the GMFIA. Hence, expertise emerges as an always mutable and precarious outcome of context-specific struggles and practices, rather than a lasting accomplishment or stable attribute (Jasanoff, 2003: 159).

To perform themselves as competent and showcase their expertise, many actors in the field of migration management engage in quantification practices thanks to the authority granted to numerical facts and their producers. The production of migration statistics is part of these struggles and belongs to the 'numerical operations' that promise to make certain aspects of social life – in this case migration – 'transparent and governable' (Hansen, 2015: 204). Hence, '[a] shift to numbers implies … a shift towards accuracy and truth, and this plays an important role in the legitimation and control of power' (Hansen and Porter, 2012: 415). In general, numbers and statistics, and the colourful charts and neat tables in which they are presented, resemble veritable 'technologies of truth production' (Urla, 1993: 819). They endow otherwise diffuse social processes like migration with the quality of quantifiable objectivity by constituting them as countable 'matters of fact'. The authority attributed to numerical facts rests, to a large extent, on the epistemological register of 'statistical realism'. The latter suggests that statistics measure realities that exist independently of statistical practices and that they do so more or less accurately, as explained in Chapter 1. Thus, the reality-status of abstract and often diffuse social processes like migration is 'presumed to be as permanent and real as any physical object' (Espeland and Stevens, 2008: 417).

In the field of migration management, the mobilisation of numerical facts as a source of expertise has been intensified by

recent calls for 'strengthening evidence-based policy making' (e.g. ICMPD, 2013). These calls translate, in turn, into calls for more and 'better' – that is: more accurate, more detailed and more timely – statistics on migration. This dynamic is well-reflected by the *Global Compact for Migration*, which is widely understood as a 'milestone' in global migration governance (Pécoud, 2021: 16). In the Compact's first objective, signatory parties 'commit to ensure that this data … guides coherent and evidence-based policymaking and well-informed public discourse, and allows for effective monitoring and evaluation of the implementation of commitments over time' (UN General Assembly, 2019: 7). These calls for better data for better policy (Willekens et al., 2016) have received an additional boost by the promise of producing more reliable and timely migration statistics thanks to new methodologies featuring various types of big data such as Google searches or mobile positioning data (Scheel and Ustek-Spilda, 2018). The thirst for ever more detailed and timely data on migration is also evident in the establishment of a growing number of 'migration knowledge hubs', such as the IOM's Global Migration Data Analysis Centre (founded 2015, Berlin), the European Commission's Knowledge Centre on Migration and Demography (2016, Brussels) and the UNHCR-World Bank's Joint Data Center on Forced Displacement (2018, Copenhagen).

The point is that this thirst for more and better numerical facts in the field of migration management translates into growing pressure on statisticians to provide more detailed, timely, and accurate data on migration. Moreover, their own expertise is increasingly measured by their ability to come up with innovative methods that can satisfy the most important users of migration statistics for precise numerical facts. While government agencies like FRONTEX, international organisations like the UNHCR, IOM, of the ICMPD as well as large

NGOs have their own statistical departments, they often rely on data provided by NSIs within the field of statistics. Hence, actors of the field of migration management are amongst some of the most important users of migration statistics. From this follows, in turn, that statisticians and NSIs are confronted with the former's expectations. It is through the growing demand for precise numerical facts as a source of expertise that the doxa[6] and inner dynamics of the field of migration management affect the stakes, struggles, and practices in the field of statistics. Especially in a context in which new sources and producers of big data compete with statisticians' methodologies and outputs, the latter are eager to respond to the call for better and more data for better migration policies by 'mak[ing] data on international migration flows more accurate and more "fit-for-use" ' (UNECE, 2008: 9).

This example shows that fields of practice do not constitute completely autonomous social spheres each of which follows a distinct inner logic, as some readings of Bourdieu suggest. Rather, fields of practices partly overlap and intersect, thus influencing the stakes and means of struggles, which in turn determine the boundaries and composition of neighbouring fields (for such a reading see for instance: C.A.S.E., 2006: 459). This is also because some actors move between and sometimes work in different fields of practice simultaneously.

Hence, the call for and use of seemingly exact numerical facts as a source of expertise in the field of migration management translates into the production of non-knowledge about the known limits of quantifying migration in both fields. To account for and study this production of non-knowledge we combine our understanding of data practices, and in particular their performativity, with insights and concepts from ignorance studies (Gross and McGoey, 2015; Proctor and Schiebinger, 2008) to show that various forms of non-knowledge play an

important role in the enactment of realities. This growing field of research is not only concerned with classical questions of epistemology such as how knowledge is generated, what qualifies knowledge as scientific or credible, and what kinds of effects this knowledge has. Scholars in ignorance studies also ask what we do not know, why we do not know it, how this non-knowledge is produced and sustained, and what kind of effects different types of non-knowledge have (Proctor, 2008). Importantly, non-knowledge is not simply conceptualised as the negative of knowledge; rather – just like knowledge – it must be actively produced, and various types of non-knowledge exist, such as uncertainty, doubt, secrecy, or 'undone science' (Hess, 2015). This is how it differs from a state of not-knowing or missing information. Moreover, the relationship between knowledge and non-knowledge is not understood in terms of a zero-sum game. Instead, non-knowledge is often intertwined with the production of knowledge, as it is also thought of as productive (Gross and McGoey, 2015). The point for the following analysis is that the production of non-knowledge also comes to shape how migration is enacted as an object of government.

To avoid allusions to any 'conspirational logic', which Frickel and Edwards (2014: 216) warn against, we stress the following two conceptual points. First, we build on McGoey's (2012) notion of 'strategic ignorance'. McGoey emphasises that, while the production of ignorance may be strategic and deliberate in the sense that individual actors may actively try to nurture and preserve ignorance to use it as a resource to advance their interests, it can also be tacit and distributed, when, for instance, it would be more advantageous to avoid troubling knowledge, in the face of social taboos or organisational and professional pressures (557, 559). These organisational and professional pressures – such as the demand for

more timely and more accurate migration statistics that are 'fit-for-use' – shape the working practices and methodological decisions of statisticians who may use simplifications as opportunities to omit inconsistencies in data or to obscure methodological uncertainties. Instances of simplification (also known as 'black-boxing' in the jargon of actor-network theory) are moments when the infinite complexity of reality is simplified through processes of translation and the creation of spokespersons acting as stand-ins for more complex processes or entities (Callon, 1986). These instances offer statisticians opportunities to produce non-knowledge about methodological issues such as inconsistencies in data. For any statistical figure – such as a five-digit number acting as a stand-in for the number of deportable migrants in year X that could not be returned to their country of origin although they had been issued a return order – is the product of multiple processes of translation that inevitably involve simplification (cf. Scheel, 2021). However, this does not mean that simplification has to be used in ways that produce allegedly solid numerical 'matters of fact' through the omission of inconsistencies in data and other methodological challenges. It still requires an actor to turn these necessary reductions of complexity into opportunities for the production of non-knowledge in order to satisfy certain professional demands and organisational pressures. The task of the following analysis is then precisely, to paraphrase Callon (1986, 29), to reveal 'the reality represented by these simplifications as an impoverished betrayal' by attending to the data practices through which such simplifications are accomplished. The benefit of such an analysis is that it can highlight the dispersed nature of the production of non-knowledge.

Second, we stress how non-knowledge is generated during the transfer of knowledge – in our case statistical data about

migration in Europe – between different fields of practice each of which follows a distinct logic and harbours distinct epistemic communities, but nevertheless interact and overlap in various ways, as noted above. This non-transfer results partly from the fact that different fields of practice deploy – and are partly defined through – distinct epistemic forms, understood as 'the suite of concepts, methods, measures and interpretations that shapes the ways in which actors produce knowledge and ignorance in their professional/ intellectual fields of practice' (Kleinman and Suryanarayanan, 2013: 492). The crucial point is that the conventions, methods, and practices which make up epistemic forms in one field are not easily transferable to other fields of practice. This is why different fields of practice develop distinct 'epistemic cultures of non-knowledge' over time (Böschen et al., 2010), which may nevertheless be affected by the expectations and epistemic cultures of related fields of practice. In our case, the demand of actors in the field of migration management for ever more detailed and timely data on migration as a source of expertise translates, for instance, into pressure for actors in the field of statistics to produce and deliver such numerical facts. These conceptual precautions are important in order to avoid the impression that our arguments follow the logic of a conspiracy theory, which attributes the production of non-knowledge about the known limits of quantifying migration to a secret, orchestrated plan of powerful lobbies or organisations like the IOM.

In sum, the production of non-knowledge emerges as the combined effect of deliberate actions and strategies of individual actors immersed in struggles over influence and resources, as mostly tacit attempts to accommodate and satisfy organisational and professional pressures in processes of translation and simplification as well as instances of the non-transfer of knowledge between different fields of practice. In

what follows, we mobilise this conceptual framework to trace the data informing the GMFIA back to sites of their statistical production. In this way we show how non-knowledge about the known limits of quantifying migration is produced at multiple sites in different fields through various data practices.

Omitting I: Ignoring Methodological Differences and Limitations

The GMFIA was launched by the IOM in 2016 as a 'migration visualisation tool' which 'tracks migrants around the world.'[7] It had a prominent position on IOM's homepage, where it could be accessed under the tab 'migration.' The GMFIA was widely circulated and used until it was deactivated by the IOM in the first half of 2019. Until then, it was ranked as the first link of any online search engine query for 'world migration' or 'migration in the world.' This was also why we chose the GMFIA as a case study.

At first sight, the GMFIA showed a conventional geopolitical map of the world (see Figure 5.1). When a particular country was selected, the quantity and composition of in- and outward migration to or from that country appeared on the screen. If inward migration to the UK was selected, a circle of coloured clusters emerged, each cluster visualising the number of immigrants from another country. By hovering over one of the coloured clusters, the respective country of origin was highlighted in the same colour, and the number of immigrants from that country was displayed. The circles showed, for instance, in 2017, 703,050 migrants from Poland and 9,361 migrants from Estonia resided in the UK.

In sum, the set-up of the GMFIA confirmed that 'data are mobilised graphically' (Gitelman and Jackson, 2013: 12). The visual design and interface of the GMFIA illustrated the

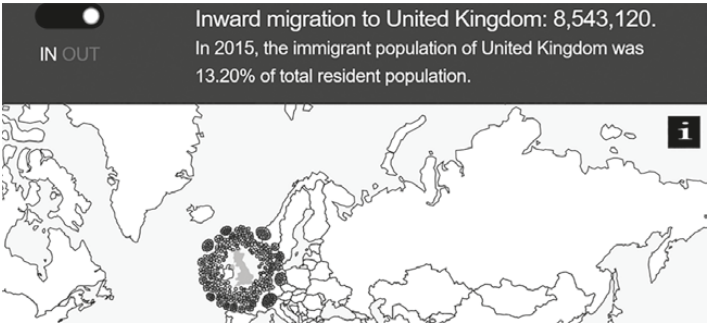

Figure 5.1 Screenshot of GMFIA[a]
[a]Screenshot from: https://web.archive.org/web/20180706142057/https://www.iom.int/world-migration (accessed 09 October 2017)

existence of certain conventions of data visualisation which, if followed, 'imbue visualisations with the quality of objectivity (which brings together other qualities such as transparency, scientific-ness and facticity)' (Kennedy et al., 2016: 716). These conventions underline the emblematic character of the GMFIA, which follows the four most important conventions of data visualisation more generally cited by Kennedy et al. (2016): (1) two-dimensional viewpoints (in this case: a map); (2) clean layouts without any decoration (only the most relevant data in in and out-flows are provided in the GMFIA with minimal text); (3) use of simple geometric shapes and lines (in the GMFIA circles and clusters); (4) inclusion of reference to data sources (in case of the GMFIA data from UN institutions, as we discuss below in detail). By following these conventions data visualisations like the GMFIA are creating a reality-effect, giving the impression that they are just 'showing the facts, telling it like it is, offering windows onto data' (Kennedy et al., 2016: 716).

Regarding the data displayed by the GMFIA, another feature was striking. The GMFIA provided perfectly matching

figures for inward and corresponding outward flows. If one compared the recorded emigration events of a given country, like Latvia, with the recorded immigration events of the corresponding destination country, like the UK, the numbers would match perfectly on the GMFIA: 66,046 people. This perfect correspondence may seem logical insofar as any immigration event (arrival in a destination country) is, by definition, tied to an emigration event (departure from previous country of residence). This is reflected in the definition of an international migrant of the 1998 'United Nations Recommendations on Statistics of International Migration' which is used by most UNECE countries and EU member states. It is based on the 12-month rule that also informs the definition of 'usual residents' to be counted as part of the resident population (see Chapter 3). The UN defines a migrant as

> a person who moves to a country other than that of his or her usual residence for a period of at least a year (12 months), so that the country of residence effectively becomes his or her new country of usual residence. From the perspective of the country of departure, the person will be a long-term emigrant and from that of the country of arrival, the person will be a long-term immigrant (see also: UNECE, 2015: 137; UNSD, 1998: 10).

This definition conforms to conventions adopted after the First World War for the statistical category of 'international migration', as developed by the ILO (International Labour Organisation), which merges the categories of emigration and immigration (Stricker, 2019). While this merging raises the expectation of perfectly matching figures on emigration and corresponding immigration flows, in practice 'emigration numbers reported by sending countries tend to differ from the corresponding immigration numbers reported by receiving

countries' (de Beer et al., 2010: 459; cf. UNECE, 2014). This mismatch is often too significant to overlook. According to Eurostat figures, for example, the UK reported 42,403 immigrants from Poland in 2015, while Poland reported sending only 11,682 emigrants to the UK.[8] This example confirms the general observation that figures on emigration tend to be lower than reported immigration events in receiving countries (UNECE, 2008). These mismatches are not only the effect of known methodological issues. They include different operationalisations of the definition cited above by NSIs or the use of different methods of data production (questionnaire-based, register-based, etc.) and encoding (i.e., the allocation of mobile people to the category international migrant across and within NSIs). One important reason for the general mismatch between numbers on immigration and corresponding emigration events is that migrating individuals usually have little incentive to inform authorities about their departure, while there may be some benefits for informing authorities in a destination country about one's arrival (UNECE, 2008).

At European and international levels, increased levels of cross-border mobility and the non-reporting of departures by emigrants raise another methodological issue known as 'double-counting' (Valente, 2014). In this case, the same person may be counted twice in population statistics and censuses: once in their country of departure and once in their destination country. Such double-counting might occur if at least one NSI, either in the country of departure and the new country of residence, conducts a register-based census. Due to the ongoing move towards register-based statistics and increasing levels of cross-border mobility the methodological issue of double-counting is likely to intensify in coming years.

However, the GMFIA omits this known inconsistency in statistical data on migration events by providing perfectly

matching figures. This omission is accomplished through a simplification that happens during the transfer of knowledge between the field of statistics and that of migration management. Regarding the origin of the data informing the visualisation, the GMFIA provided only scarce information to users. Only if they followed a link to an external webpage, were users able to find out that the data were taken from the 2015 edition of a database on 'Trends in International Migrant Stock' by the United Nations' Department of Economic and Social Affairs (UNDESA). The point is that the UN database only provides data on stocks of immigrants in a given country and that the data had simply been equated in the GMFIA with the number of emigration events in corresponding countries of origin. By establishing this equivalence, the known inconsistencies between reported emigration and immigration events were simply omitted. Instead, the IOM's data visualisation tool provided perfectly matching figures for immigration and emigration, thus enacting migration as a coherent, precisely measurable reality.

One reason that may justify this data practice of omission is that data on immigration events in receiving countries are usually regarded as 'more reliable' than data on emigration events in sending countries (UNECE & Eurostat, 2010: 7).[9] This is why the UNECE recommends the use of so-called 'mirror statistics' in which data on emigration are validated and possibly adjusted with the help of immigration data obtained from NSIs in countries of destination. However, this recommendation does not mean that receiving country data should always be considered as better or more reliable than sending country data (de Beer et al., 2010: 471) as the former might also suffer from methodological limitations.

Moreover, data on immigration – and in particular data on stocks of migrants used in the GMFIA – is haunted by a similar

problem as data on emigration events. This is because immigrants may also re-emigrate (either to another country or their country of origin) without notifying authorities about their departure. Hence, as renowned migration studies scholars observe, '[w]ith arrivals reported and departures not reported, the number of immigrants is overestimated' (Willekens et al., 2016: 897). Thus, 'the mirror [recorded immigration events] reflects biased images [of emigration events in countries of origin]', as a statistician of Eurostat put it in a presentation on migration statistics.[10] Hence, the usage of data on stocks of international migrants as a stand-in for emigration events in the GMFIA is, in methodological terms, a highly questionable and misleading simplification as it effectively produces non-knowledge about one of the most important known weaknesses of migration statistics.

Moreover, if users traced back the origin of the data informing the GMFIA they would learn in the documentation on the UN webpage that data from various sources have been combined in the dataset, including censuses, surveys, and administrative registers (UNDESA 2015, 7). Yet, this assembling of different data sources into the GMFIA could not be done without omitting important methodological differences. To ignore this methodological heterogeneity was necessary because the use of different definitions, methods, and data sources makes comparison of migration data 'difficult and confusing' (Wisniowski et al. 2013, 460). This concerns both spatial comparisons between countries and temporal comparisons in the migration time series of one country (Herm and Poulain, 2012). Consequently, the GMFIA's representation of 'world migration' as a series of seemingly precise, stable, and comparable figures emerged as an accomplishment which relied on the production of non-knowledge of methodological differences and the implications of this heterogeneity.

Omitting II: On the Dispersed Production of Non-knowledge

Non-knowledge about methodological differences and their implications was, in case of the GMFIA, to a significant extent produced by efficiently placing the burden of tracing the origin and assessing the quality of data used in the GMFIA on the individual user. A short notice below the GMFIA's 'world migration map' informed users that the figures displayed by the visualisation were based on UNDESA's 2015 dataset. If users did, however, retrieve relevant metadata from the two linked UN webpages, they would have discovered that information provided on the sources of UNDESA's dataset were generic and incomplete. Hence, they were compelled to visit the webpages of individual NSIs to search for and retrieve relevant metadata on migration statistics. This turned the assessment of the origin and quality of data informing the GMFIA into a cumbersome, time-consuming and, in many instances, futile forensic search. Moreover, users 'experience of agency in this respect is reliant upon technological competence' (Birchall 2015, 190), most notably statistical literacy. This is, however, a skill that not all users of the GMFIA may have had. What the example shows is that omission also works by excluding any explicit reference to the known limitations of migration statistics, most notably their incoherence and incomparability across space and time which results, in turn, from the pronounced methodological heterogeneity in the production of migration statistics across as well as within NSIs.

In the following we describe methodological changes and choices in the production of migration statistics within one NSI in order to illustrate the complexity of the methodological issues, decisions, and controversies that are omitted by the GMFIA. What we seek to highlight in particular is

that the production of non-knowledge about methodological challenges and limitations, related changes, and controversies, as well as their implications, is dispersed across different sites and fields of practice. To this end we analyse how different versions of migration have been produced and negotiated at Statistics Estonia (SE) in the course of implementing methodological changes in migration statistics. This allows us to show that non-knowledge about methodological differences and limitations is not only produced by actors of the field of migration management, but also by actors in the field of statistics as well as during the transfer of knowledge between these fields.

According to the GMFIA, Estonia hosted 202,348 immigrants in 2015. If we access the GMFIA via the internet archive and hover over the coloured circles visualising the composition of Estonia's immigrant population, we learn that 143,677 people from Russia, 1,271 people from Germany, 1 person from Sudan, and 13 people from Nigeria resided in Estonia in 2015. Again, through the provision of these very exact figures, also for very small groups of migrants, the GMFIA enacts migration as something that can be precisely quantified.

Figure 5.2 Screenshot of GMFIA[a]

[a]Screenshot from: https://web.archive.org/web/20180706142057/https://www.iom.int/world-migration (accessed 27 May 2020)

However, this account required omitting multiple efforts and methodological changes implemented by SE's statisticians in preceding years to better quantify migration. In 2013, SE reported 6,740 emigrants and 4,098 immigrants and in 2014, 4,637 emigrants and 3,904 immigrants (SE, 2015).[11] In 2015, however, 15,413 immigrants and 13,003 emigrants were reported – an increase of nearly 400 per cent in immigration and 300 percent in emigration. In press releases, Estonian statisticians attributed this jump to a change in methodology (SE, 2016a, 2016b). Until 2016, SE mainly relied on recorded migration events in the Estonian population register (RR). However, the low numbers of reported migration events were increasingly regarded as implausible by statisticians, policy makers and demographers.[12]

Statisticians cited unreliable RR data as the principal reason for the low figures. Many individuals would simply not notify authorities about their departures, a known problem mentioned above. Furthermore, statisticians pointed out an issue with the computer software used for producing migration statistics. The software required statisticians to enter a person's previous country of usual residence to include that person in the immigrant population. While introduced with the intention of obtaining as detailed information on the immigrant/emigrant population as possible, this requirement produced a significant 'undercount' of immigration from EU member states.[13]

Hence, statisticians developed a new method, the 'residency index' (RI), for producing migration statistics. The RI is based on a relatively simple idea. In brief, it infers an individual's residence status from recorded activities of that person in various government registers. If a person with a record in the RR does actually live in Estonia, it is assumed they will engage in more transactions with government institutions than a

person who does not reside in the country. These transactions leave behind traces (so-called 'signs of life') in administrative registers (Tiit and Maasing, 2016). To illustrate, if a person studies in Estonia, they will have a record in the education register. If they are employed, they will pay taxes. If they are retired, they will receive a pension and so forth. Thanks to the unique personal identification number that is used across all administrative registers, Estonian statisticians can link data from 14 different registers to calculate a residency index for each person with a record in the RR. The value of a person's residency index ranges between 0 and 1, depending on the number of signs of life they accumulate across all government registers in a given year. The higher the value of the index, the higher is the probability that they are a usual resident of Estonia. To be considered a resident of Estonia, a person's residency index has to be above the threshold of 0.7 (ibid.).

In press releases SE promoted the RI-model as 'reflect[ing] reality more accurately' (SE, 2016a). The change in methods also made it necessary to address the software issue as the RI-model 'discovered' many EU immigrants whose previous country of residence was unknown.[14] Hence, Estonian statisticians claim that the new methodology better accounts for the 'immigration of European citizens, which the previous methodology reflected to a smaller extent' (SE, 2016b).

This framing of the RI-model as 'more accurate' suggests that statisticians assume – as suggested by the epistemological register of statistical realism – an objective, external reality that can be measured – more or less accurately – with different methods. Since statisticians have no direct access to this external reality, the challenge is then to assess which of the methods used provides the most accurate measurements. Hence, statisticians deploy various tactics to assess the reliability and limitations of their methods in order to privilege

one over the other. If we take seriously the performativity of data practices, this orchestration work illustrates, however, that different methods enact different versions of the real – in this instance different accounts of migration to Estonia. The orchestration work needed to maintain the notion of a singular, stable, and coherent reality illustrates Mol's (2002: 47) observation that 'if two objects that go under the same name clash, in practice one of them will be privileged over the other'. In the following we describe how this privileging of one version of the real over other versions is done through the concepts of 'over-' and 'under-coverage'.

In Estonia the original RR methodology for migration statistics was problematised after the 2011 census as entailing a significant under-coverage of emigration. The census questionnaire included a question on the emigration of any household members. Results suggested that more people emigrated than previously calculated with the cohort component method. This method adapts the results of the last census (PHC2000) on an annual basis by the number of births, deaths, and net migration figures recorded in the RR. By diagnosing an under-coverage in this RR-based methodology, the census was elevated to the position of the superior method (Tiit, 2012). However, since the census is only conducted once per decade, it is unable to provide a methodology for SE's annual migration statistics. Hence, statisticians developed the RI-model and subsequently declared this new methodology as the new gold standard. An article published in SE's house journal assesses, for instance, the registration behaviour of people in Estonia, based on a comparison between RR data and calculations with the RI-model. The article stresses that 'in the case of a discrepancy between the two datasets, the data according to the index, and not the population register, are considered accurate' (Meres 2017, 72).

What this orchestration work indicates is that over- and under-coverage function as important devices for privileging one method as 'more accurate' than another. While over- and under-coverage suggest a relation between a statistical account of migration and an objective, external migration reality, they do actually express relations between different versions of migration that have been produced with different methods. Hence, over- and under-coverage are the outcome of assessments in which statisticians compare one statistical method in relation to another one in order to privilege one over the other in terms of accuracy. The under-coverage of emigration events by the RR-based methodology only emerged as a problem when a second methodology – the 2011 census – was brought into play. Hence, the diagnosed under-coverage in the RR data expresses a relation to census data and not to an external reality. Ultimately, the relational character of over- and under-coverage shows that there is no objective migration reality that could provide a standard to assess the quality of statistical methods since any migration reality only exists in relation to the practices and methods that are used to know it.

However, the invocation of over- or under-coverage permits statisticians to establish hierarchies between different versions of migration that have been produced with different methods. This hierarchisation is important groundwork for other activities needed to comply with the convention of common-sense realism that there can only be one, more or less coherent migration reality: the discontinuation of any method that has been established as inferior and the compression of essentially different versions of migration into one time series. For all that remains of essentially different versions of migration is a significant increase in migration events in 2015, a jump in the time series of Estonia's migration statistics

which may be, if interpreted carefully, attributed to a change in methodology (Herm and Poulain, 2012).

The point concerning the GMFIA is that changes in methodology within one NSI, as well as methodological heterogeneity across NSIs, imply that migration is not comparable across time and space (UNECE, 2014). Users of the GMFIA are essentially dealing with different versions of migration. For what migration is – and how large or small in terms of numbers – 'depends on how "it" is being done in practice' (Law and Lien, 2012: 366). Depending on the methodology used, emigration from Estonia can be a 'yes' to the question 'has any close relative of yourself or of a member of your household … left Estonia in 2000 or later and is currently living abroad?' (Tiit, 2014a: 85); or it could be a recorded emigration event in the population register or a residency index value of less than 0.7 in two consecutive years. And other NSIs use different methodologies like estimations based on administrative data or sample surveys. These are all different methods that enact not only different accounts, but different versions of migration. The GMFIA omits these methodological differences by compressing essentially different versions of migration into one 'world migration map'. In this way knowledge of these methodological differences and their implications is omitted as all migration data are treated as equivalent.

In sum, the analysis of GMFIA and SE's methodological changes show that the production of non-knowledge about the known limits and challenges of quantifying migration is dispersed across various sites in different fields of practice. As illustrated by the use of data on stocks of international migrants as a stand-in for emigration events, instances of simplification (or 'black-boxing') are often turned into opportunities for the production of non-knowledge about methodological challenges and inconsistencies in data. Furthermore, the

production of non-knowledge is also located in the dynamics in-between different fields of practice. This can be illustrated through the use of metadata reports as devices for omitting the messiness of statistical data production which may compromise the authority of published figures, or the legitimacy and comparability of the methodologies by which they were produced. In brief, metadata reports are meant to make the statistical production process transparent and, at least potentially, replicable. In practice, metadata reports are, however, often incomplete. The use of metadata reports as means of omitting messy, potentially compromising moments of the statistical production process is possible, because a detailed description of the latter is potentially infinite. Hence, statisticians can – and often do – provide metadata in such a way that it does not compromise the authority and methodological soundness of their statistical products. For example, in the case of SE discussed in this chapter, metadata reports on migration do not ever mention the computer problem as one cause for the known under-coverage of immigration events. The selective provision of metadata is not surprising for a field of practice in which methodological rigour and a commitment to diligence and accuracy are part of its professional 'code of honour', that is, part of its epistemic form and authority. However, in doing so, metadata reports also produce strategic ignorance about the messy moments of statistical production processes, and result in the partial transfer of knowledge about the problems of quantifying migration from the field of statistics to that of migration management.

Furthermore, the production of non-knowledge about the known limits of quantifying migration stems, to a significant degree from dynamics in-between the field of statistics and the field of migration management and related non-transfer of knowledge between these two professional

fields of practice. Statisticians do, for instance, not only omit the mess of the statistical production process to satisfy the expectations and demands of peers or to gain influence and authority within the field of statistics. They also do so to satisfy the expectations and demands of their 'customers' which are, in this case, users of migration statistics, mostly located in the field of migration management. And these users are, due to the requirements, inner logics and 'rules of the game' of their field, primarily interested in precise numerical facts that they can use to perform themselves as knowledgeable actors and for legitimising their preferred policy options. What they are not interested in, are lengthy accounts of methodological issues that may cast doubts on, and provoke discussions about the reliability of their published figures.[15] These are thus also the requirements and inner logics of a field of practice in which quantitative precision and the provision of precise numerical facts function as stake in struggles over authority and resources which constitute one of the reasons why omitting messy moments of statistical production in metadata reports has become part of the 'epistemic culture of non-knowledge' (Böschen et al., 2010) in the field of statistics.

At the same time, actors in the field of migration management are not so much interested in statisticians' careful methodological considerations and changes, as well as related debates and controversies, which we understand as an important part of the epistemic culture of the field of statistics. These elements of this epistemic culture are, however, largely ignored by the field of migration management when migration statistics are used and taken up by actors in that field as a resource for assembling credibility and showcasing expertise. What these examples show is that the production of non-knowledge about the known limits of quantifying migration is not only strategic in the sense of McGoey (2012).

It is also the combined effect of the dynamics in-between overlapping fields of practice.

Recalibrating: Enacting Migration as a Singular, Coherent Reality

In this section we describe another data practice by which statisticians try to satisfy the demands of users of statistics for coherent and precise figures about migration. We call this data practice recalibration. It refers to the activities statisticians regularly engage in to adjust data on migration in order to even out inconsistencies in time series resulting from changes in methodology. Such ex-post adjustments are a widely accepted practice in official statistics. After population censuses are conducted statisticians often adjust their population figures, as well as related migration statistics, because censuses are generally regarded as a superior methodology that can provide a more accurate account of the population than other statistical methods such as surveys or register-based statistics. In the case discussed in this section, the adjustment of migration data for the years preceding the 2011 census was related to the recalculation of the population size of Estonia for the intercensal years. As we show, the main purpose of these adjustments was to displace the only indication of methodological changes – jumps and bumps in the time series of migration and demographic statistics – to a distant, less controversial, or even inaccessible past.

In the following, we refer to the activities facilitating such adjustments as recalibration. According to the Cambridge dictionary the verb 'to recalibrate' carries two connotations: (1) 'to make small changes to an instrument so that it measures accurately' and (2) 'to change the way you do or think about something.'[16] Hence, the verb recalibrate emphasises that

something is remade or redone, albeit in an adjusted, modified way. Yet, one important aspect we seek to highlight with the notion of recalibration is not captured by the definition above. Rather than just changes in instruments of measurement (the methodology used to 'measure' migration), recalibration also involves adjusting migration data for the years preceding these methodological changes. Put simply, recalibration features in not only changes to instruments of measurement, but also changes to the data about the object to be measured. In this way recalibration aligns the results of what are supposedly only 'measurements' – in this case data produced on migration – with new instruments of measurement. Thus, inconsistencies in data resulting from changes in methodology are efficiently displaced and ignored. Hence, recalibration is about adjusting statistical data to the outputs of a new methodology that is declared as the new gold standard.

Statisticians decided to adjust the population size of Estonia after the 2011 census (PHC2011) following reports in the media about people who claimed they had not been enumerated (Tiit, Meres, and Vaehi, 2012). When statisticians investigated the matter, they were confronted with three different population estimates based on three different methods. First, the population size calculated annually by SE using the cohort component method, as described in the previous section. According to this method, Estonia's population had declined, comprising 1,320,000 residents in 2011. Second, there was the even lower census result of 1,294,455 people enumerated through online and door-to-door enumeration. Third, there was the largest figure of 1,365,000 people officially recorded as residents of Estonia in the population register (RR) (Tiit, 2014b). When a working group compared these different population estimates, they discovered that about 71,000 people had a record in the RR but had not been enumerated in

the census. For all these people statisticians had to determine – in the terms of statistical realism – if they had not been enumerated in the census due to under-coverage or if they were part of the over-coverage in the RR, that is, emigrants who had left Estonia without notifying authorities about their departure.

Statisticians solved this conundrum by developing a method that allowed them to determine the residency status of the 71,000 people with a record in the RR that had not been enumerated in the census (Tiit, Meres, and Vaehi, 2012). To this end, they formed different sex-age groups and determined the residency status of individuals assigned to these groups by assessing the 'signs of life' these people had left behind in different registers through their transactions with state institutions. Based on this methodology, which was an earlier version of the RI-model described in the previous section, statisticians concluded that 30,760 of these 71,000 people were actually residents of Estonia. Hence, SE quantified the population of Estonia with an 'adjusted figure' of 1,325,217 people on its website (for a more detailed account of this methodology see: Scheel, 2020).

The crucial point for our argument on recalibration is that SE's statisticians decided to adjust the population size of Estonia for the intercensal period from the year 2000 onwards. Thereby statisticians sought to even out the sudden population increase implicated by the adjustment of the PHC2011 results in the time series of Estonia's demographic statistics. Following statisticians, this adjustment was supposed to provide a more accurate account of the development of the decreasing Estonian population (Tiit, Meres, and Vaehi, 2012: 100). This claim was justified with two interrelated assumptions. First, that 'throughout the past 12 years, Estonia's reported population size has been slightly smaller than the actual population size' (ibid.) because the PHC2000 results – the basis

of the cohort component methodology – also comprised an under-coverage according to a follow-up survey.[17] Second, statisticians assumed high levels of unregistered emigration that was not reflected in the migration statistics for the intercensal years, making it necessary 'to *distribute the population decline* caused by emigration evenly between the years' (SE, 2014; emphasis added). Based on these assumptions, statisticians decided to 'correct' the PHC2000 results upwards by 31,200 people (ibid.). Furthermore, they decided to adjust the reported population size for the intercensal years downwards, based on estimated levels of unregistered emigration. These were calculated on basis of the following assumptions:

> First, the volume and distribution [of unregistered migration] by age is the same; second, the distribution by sex is different, assuming that women register their migration more correctly than men, with a difference of 5%. With these assumptions we built the model of migration, also considering the citizenship of migrants. We calculated for each year how migration affected the population size.[18]

In other words, the recalibration of Estonia's population size in demographic statistics required to also recalibrate migration data for the intercensal years.

This second recalibration led, however, to inconsistent data on international migration in SE's statistical database. In the demographic statistics shown below (Figure 5.3), statisticians added between 1,181 and 2,203 emigrants to the annual emigration rate in the column 'statistical correction', based on the assumption that these people had not informed authorities about their departure. Since SE's statisticians did not change the numbers in Estonia's migration statistics for the same period (see Figure 5.3 below), statisticians produced and worked with two different accounts of international migration for the period 2000–2011.

PO0213: COMPONENTS OF CHANGE IN POPULATION FIGURE by County, Year, Sex and Indicator

	Males and females									
	Population at the beginning of the year	Natural increase	Live births	Deaths	Net migration	Immigration	Emigration	Statistical correction	Population at the end of the year	Change in population figure
Whole country										
2000	1 401 250	-5 336	13 067	18 403	-1 749	35	1 784	-1 445	1 392 720	-8 530
2001	1 392 720	-5 884	12 632	18 516	-1 934	241	2 175	-1 392	1 383 510	-9 210
2002	1 383 510	-5 354	13 001	18 355	-1 463	575	2 038	-1 503	1 375 190	-8 320
2003	1 375 190	-5 116	13 036	18 152	-2 106	967	3 073	-1 718	1 366 250	-8 940
2004	1 366 250	-3 693	13 992	17 685	-1 830	1 097	2 927	-1 877	1 358 850	-7 400
2005	1 358 850	-2 966	14 350	17 316	-3 174	1 436	4 610	-2 010	1 350 700	-8 150
2006	1 350 700	-2 439	14 877	17 316	-3 293	2 234	5 527	-2 048	1 342 920	-7 780
2007	1 342 920	-1 634	15 775	17 409	-643	3 741	4 384	-2 203	1 338 440	-4 480
2008	1 338 440	-647	16 028	16 675	-735	3 671	4 406	-1 318	1 335 740	-2 700
2009	1 335 740	-318	15 763	16 081	-774	3 884	4 658	-1 358	1 333 290	-2 450
2010	1 333 290	35	15 825	15 790	-2 484	2 810	5 294	-1 181	1 329 860	-3 630
2011	1 329 660	-565	14 679	15 244	-2 505	3 709	6 214	-1 373	1 325 217	-4 443
2012	1 325 217	-1 394	14 056	15 450	-3 682	2 639	6 321	33	1 320 174	-5 043
2013	1 320 174	-1 713	13 531	15 244	-2 642	4 098	6 740	0	1 315 819	-4 355
2014	1 315 819	-1 933	13 551	15 484	-733	3 904	4 637	118	1 313 271	-2 548
2015	1 313 271	-1 336	13 907	15 243	2 410	15 413	13 003	1 599	1 315 944	2 673
2016	1 315 944	-1 339	14 053	15 392	1 030	14 822	13 792	0	1 315 635	-309

Footnote:
As of 1 January 2016, Statistics Estonia uses a new source of residence data and methodology for calculating the population figure, which have to be taken into account upon analysing changes. The place of residence is the place of residence stated in the Population Register, if it is left unmarked, the persons will be categorised under "County unknown". More information can be found under Definitions and Methodology.

Indicator
Statistical correction
Up to 2014 (incl.), can be regarded as unregistered external migration; 2015 – the difference due to the methodological change and transition from the place of residence recorded in the census to the place of residence recorded in the Population Register.

Figure 5.3 Population Size of Estonia with 'Statistical Correction'[a]
[a]Screenshot from: SE's statistical database at https://andmed.stat.ee/et/stat (accessed 16 May 2019)

This inconsistency is, however, only visible to experts, such as demographers, who carefully analyse population and migration statistics. A demographer from the University of Tallinn thus concluded in an interview that 'the numbers do not add up.'[19] The average user will hardly notice the inconsistency between the numbers on migration provided in demographic and migration statistics. Nor will most users notice the sudden increase in Estonia's population size resulting from the correction of census results since the jump in the data resulting from the recalculation of the PHC2011 results has been effectively displaced – through the adjustment of Estonia's population size for the intercensal period – to a distant past which is

* Latest update: 09.05.2019

POR05: EXTERNAL MIGRATION by Country, Year, Sex and Indicator

	Males and females		
	Immigration	Emigration	Net migration
Total			
2004	1 097	2 927	-1 830
2005	1 436	4 610	-3 174
2006	2 234	5 527	-3 293
2007	3 741	4 384	-643
2008	3 671	4 406	-735
2009	3 884	4 658	-774
2010	2 810	5 294	-2 484
2011	3 709	6 214	-2 505
2012	2 639	6 321	-3 682
2013	4 098	6 740	-2 642
2014	3 904	4 637	-733
2015	15 413	13 003	2 410
2016	14 822	13 792	1 030
2017	17 616	12 358	5 258
2018	17 547	10 476	7 071

Footnote:
The structure of the table has been changed on 23.05.2016.
As of 2016, Statistics Estonia calculates external migration (the transition of a person from a resident to a non-resident and vice versa) using the residency index.

Figure 5.4 Statistics on International Migration by SE for the Period 2004–18[a]

[a]Screenshot from: SE's statistical database at https://andmed.stat.ee/et/stat (accessed 29 September 2017)

politically less relevant and controversial. The sudden increase in the time series of Estonia's population occurs now between the years 1999 and 2000 but this jump is invisible for users as data on the population size of Estonia can only be retrieved up to the year 2000 from SE's statistical database.[20] By recalibrating migration data for the intercensal years statisticians have thus efficiently displaced the jumps in the time series of SE's population statistics resulting from methodological changes to an inaccessible past (see Figure 5.4).

The important point for our argument about the inconsistencies and limitations of migration statistics is that recalibration comes at a price. Statisticians' attempts to even out and dislocate jumps and bumps in time series resulting from

changes in methodology to a distant, less visible or even inaccessible past has one important implication: data on migration become mutable and changeable. Rather than solid 'numerical facts' that provide exact and reliable accounts of stable migration realities, the results of statisticians' attempts to quantify migration emerge as contingent, fragile accomplishments that may be subject to change, adjustment, and recalculation. Hence, migration emerges as a reality that is multiple, slippery, and ghostly, and thus frequently escapes and betrays statisticians' efforts to quantify it.

Importantly, the production of non-knowledge about the limitations and inconsistencies of migration statistics through recalibration cannot be attributed to the inner logics and epistemic form of the field of statistics alone. Rather, statisticians engage in the immense work that recalibration requires – in the case of Estonia it preoccupied an entire working group for a half year – to satisfy the demands of the users of migration statistics for precise numerical facts as well as the expectations of the wider public to provide just one consistent statistical account of migration. One statistician summarised this imperative to produce precise numbers as follows:

> If you think of the population of Estonia, there are in fact many grey areas … who should be part of population and who not? We can say plus-minus 10,000 [people], and it would be a very exact number from a statistical point of view. But our customers want even more precise numbers, to the point, and we work for our customers.[21]

What this quote illustrates is that the doxa of related fields of practice, and the interests and expectations of the actors inhabiting these fields, influence the logics and shape the data practices within the field of statistics. And since individual statisticians and institutional actors like NSIs work in relation

to the logic of this field, they engage in data practices like omitting or recalibrating to produce non-knowledge about the known limits, methodological issues, and inconsistencies of migration statistics. By engaging in such practices, statisticians try to satisfy, as much as possible, users' demands for precise numerical facts that can be mobilised as stakes in struggles over expertise, authority, and legitimacy in related professional fields. For even if statisticians were absolutely transparent about the limitations, uncertainties, and inconsistencies haunting their statistical outputs, these methodological intricacies would be ignored by users of statistics in other fields, as the example of the GMFIA and similar visualisations of migration data illustrate.

Conclusion

This chapter has focused on two data practices: omitting and recalibrating. What the analysis has shown is how statistical practices that enact the people of Europe involve the production of various types of non-knowledge (McGoey, 2012; Gross and McGoey, 2015; Aradau, 2017). A central insight of our analysis is that the data practices through which this non-knowledge is produced – in our case, non-knowledge about the known limitations and methodological heterogeneity of migration statistics – are distributed across different sites and fields of practice. Hence, the production of non-knowledge about limits, uncertainties, and inconsistencies of population statistics is not only strategic in the sense that it serves individual actors in the field of statistics or related fields to improve their relative positions. Nor is it only strategic in the sense that it is a mostly unconscious strategy that actors of a given field deploy to accommodate and reproduce the doxa of that field. While the production of non-knowledge about the

limitations and inconsistencies of migration and other statistics is strategic in these two senses, it is also the combined effect of the non-transfer of knowledge in-between different fields of practice. This non-transfer of knowledge, as well as the production of strategic ignorance within the fields of statistics, are in turn related to dynamics between different fields of practice, such as the field of statistics and the field of migration management.

Hence, the production of non-knowledge emerges indeed as an integral feature of the enactment of the people of Europe and their 'Others'. Consequently, the enactment of the people of Europe through the co-production of their 'Others' is not only situated in regimes of knowledge, as Said (2003) and many others have shown (Todorov, 1999; Honig, 2001; de Genova, 2016). It is also situated in and accomplished through regimes of non-knowledge that are enmeshed with the former and comprise manifold forms and types of non-knowledge as well as diverse sets of practices and devices through which this non-knowledge is produced. This is where we see ample opportunities for future research: a range of studies that attends to how the production of various forms of non-knowledge features in the making of Europe, understood as a polity and a people.

References

Aradau C (2017) Assembling (Non)Knowledge: Security, Law, and Surveillance in a Digital World. *International Political Sociology* 11(4): 327–342.

Baldwin-Edwards M, Blitz BK, and Crawley H (2019) The Politics of Evidence-Based Policy in Europe's 'Migration Crisis.' *Journal of Ethnic and Migration Studies* 45(12): 2139–2155.

Bigo D (2011) Pierre Bourdieu and International Relations: Power of Practices, Practices of Power. *International Political Sociology* 5(3): 225–258.

Birchall C (2015) 'data.gov-in-a-Box': Delimiting Transparency. *European Journal of Social Theory* 18(2): 185–202.

Böschen S, Kastenhofer K, Rust I, et al. (2010) Scientific Nonknowledge and Its Political Dynamics: The Cases of Agri-Biotechnology and Mobile Phoning. *Science, Technology & Human Values* 35(6): 783–811.

Boswell C (2009) *The Political Uses of Expert Knowledge: Immigration Policy and Social Research*. Cambridge: Cambridge University Press.

Bourdieu P (1977) *Outline of a Theory of Practice*. Cambridge: Cambridge University Press.

Bourdieu P (1991) *Language and Symbolic Power*. Cambridge and Malden, MA: Polity Press.

Bourdieu P and Wacquant L (1992) *An Invitation to Reflexive Sociology*. Cambridge: Polity Press.

Callon M (1986) The Sociology of an Actor Network: The Case of the Electric Vehicle. In: Callon M, Law J, and Rip A (eds) *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World*. London: Macmillan.

C.A.S.E. (2006) Critical Approaches to Security in Europe: A Network Manifesto. *Security Dialogue* 37(4): 443–487.

de Beer J, Raymer J, van der Erf R, et al. (2010) Overcoming the Problems of Inconsistent International Migration data: A New Method Applied to Flows in Europe. *European Journal of Population* 26(4): 459–481.

de Genova NP (2016) The European Question. Migration, Race, and Postcoloniality in Europe. *Social Text* 34(3): 75–102.

Espeland WN and Stevens ML (2008) A Sociology of Quantification. *European Journal of Sociology* 49(03): 401–436.

Frickel S and Edwards M (2014) Untangling Risk in Environmental Risk Assessment. In: Soraya B and Jas N (eds) *Powerless Science? Science and Politics in a Toxic World*. New York: Berghahn Books, pp. 215–233.

Geuss R (2001) *History and Illusion in Politics*. Cambridge: Cambridge University Press.

Gitelman L and Jackson (2013) Introduction. In: Bowker GC, Brine KR. Garvey EG et al. (eds) *'Raw Data' Is an Oxymoron*. Cambridge, MA and London: MIT Press, pp. 1–14.

Gross M and McGoey L (2015) Introduction. In: *Routledge International Handbook of Ignorance Studies*. London and New York: Routledge, pp. 1–14.

Hansen HK (2015) Numerical Operations, Transparency Illusions and the Datafication of Governance. *European Journal of Social Theory* 18(2): 203–220.

Hansen HK and Porter T (2012) What Do Numbers Do in Transnational Governance? *International Political Sociology* 6(4): 409–426.

Herm A and Poulain M (2012) Economic Crisis and International Migration. What the EU Data Reveal? *Revue européenne de migrations internationales* 28(4): 145–169.

Hess DJ (2015) Undone Science and Social Movements: A Review and Typology. In: Gross M and McGoey L (eds) *Routledge International Handbook of Ignorance Studies*. London and New York: Routledge, pp. 141–154.

Honig B (2001) *Democracy and the Foreigner*. Princeton, NJ and Oxford: Princeton University Press.

ICMPD (2013) *Summary of Debates and Conclusions of the Co-Chairs*. Meeting on the Cross-cutting Topic: Strengthening Evidence-Based Policy Making in the Field of Migration. 11-12 September 2013, Dakar. Available at: https://bityl.co/6TTT (accessed 19 July 2018).

Jasanoff S (2003) (No?) Accounting for Expertise. *Science and Public Policy* 30(3): 157–162.

Kennedy H, Hill RL, Aiello G, et al. (2016) The Work That Visualisation Conventions Do. *Information, Communication & Society* 19(6): 715–735.

Kleinman DL and Suryanarayanan S (2013) Dying Bees and the Social Production of Ignorance. *Science, Technology & Human Values* 38(4): 492–517.

Law J and Lien ME (2012) Slippery: Field Notes in Empirical Ontology. *Social Studies of Science* 43(3): 363–378.

McGoey L (2012) The Logic of Strategic Ignorance. *The British Journal of Sociology* 63(6): 553–576.

McGoey L (2019) *The Unknowers: How Strategic Ignorance Rules the World*. London and New York: Zed Books.

Meres K (2017) Calculation of Population Size: Residency Index vs Population Register. *Quarterly Bulletin of Statistics Estonia* 2017(1): 67–72.

Mol A (2002) *The Body Multiple: Ontology in Medical Practice*. Durham, NC: Duke University Press.

Pécoud A (2021) Narrating an Ideal Migration World? An Analysis of the Global Compact for Safe, Orderly and Regular Migration. *Third World Quarterly* 42(1): 16–33.

Proctor R and Schiebinger LL (2008) *Agnotology: The Making and Unmaking of Ignorance*. Redwood City, CA: Stanford University Press.

Proctor RN (2008) Agnotology. A Missing Term to Describe the Cultural Production of Ignorance (and Its Study). In: Proctor RN and Schiebinger LL (eds) *Agnotology: The Making and Unmaking of Ignorance*. Redwood City, CA: Stanford University Press, pp. 1–34.

Said EW (2003) *Orientalism*. 25th Anniversary Edition. London: Penguin Books.

Scheel S (2020) Biopolitical Bordering: Enacting Populations as Intelligible Objects of Government. *European Journal of Social Theory* 23(4): 571–590.

Scheel S (2021) The Politics of (Non)Knowledge in the (Un)Making of Migration. *Journal of Migration Studies* 1(2): 1–33.

Scheel S and Ustek-Spilda F (2018) Big Data, Big Promises: Revisiting Migration Statistics in Context of the Datafication of Everything. *Border Criminologies* Blog. Available at: https://bityl.co/57xd (accessed 24 January 2019).

Scheel S and Ustek-Spilda F (2019) On the Politics of Expertise and Ignorance in the Field of Migration Management. *Environment and Planning D: Society and Space* 37(4): 579–588.

SE (2014) Aastate 2000–2013 rahvaarvu ümberarvutuste metoodika [Methodology of Revision of Population Figures for the Years 2000–2013]. Statistics Estonia. Available at: www.stat.ee/75541 (accessed 17 May 2019).

SE (2015) *Emigration from Estonia Decreased Last Year.* News Press Release No. 57. 22 May. Tallinn: Statistics Estonia. Available at: www.stat.ee.

SE (2016a) *Last Year Immigration Exceeded Emigration for the First Time after the Restoration of Independence in Estonia*. News Press Release No. 59. 23 May. Tallinn: Statistics Estonia. Available at: www.stat.ee.

SE (2016b) *The Population Increase Was Influenced by External Migration.* News Press Release No. 48. 4 May. Tallinn: Statistics Estonia. Available at: www.stat.ee.

Stricker Y (2019) 'International Migration' Between Empire and Nation. The Statistical Construction of an Ambiguous Global Category in the International Labour Office in the 1920s. *Ethnicities* 19(3): 469–485.

Tiit E-M (2012) Assessment of Under-Coverage in the 2011 Population and Housing Census. *Quarterly Bulletin of Statistics Estonia* 2012(4): 116–119.

Tiit E-M (2014a) *2011 Population and Housing Census. Methodology*. Tallinn: Statistics Estonia. Available at: stat.ee/dokumendid/77907 (accessed 30 March 2015).

Tiit E-M (2014b) On the Methodology of Demographic Statistics. *Quarterly Bulletin of Statistics Estonia* 2014(2): 93–99.

Tiit E-M and Maasing E (2016) *Implementation of the Residency Index in Demographic Statistics*. Tallinn: Statistics Estonia. Available at: www.stat.ee.

Tiit E-M, Meres K, and Vaehi M (2012) Assessment of the Target Population of the Census. *Quarterly Bulletin of Statistics Estonia* 2012(3): 96–108.

Todorov T (1999) *The Conquest of America: The Question of the Other*. Norman, OK: University of Oklahoma Press.

UN General Assembly (2019) *Global Compact for Safe, Orderly and Regular Migration*. Resolution adopted by the General Assembly on 19 December 2018. United Nations.

UNECE (2008) *Improving Statistics on International Migration.* Note by the Secretariat presented at the Seminar on Measuring Population Movement and Integration in a Globalized World. Geneva: Conference of European Statisticians.

UNECE (2014) *Filling the 'Migration Gaps' – Can Research Outcomes Help Us Improve Migration Statistics?* Note by the Statistical Office of the European Union (Eurostat). Geneva: Conference of European Statisticians.

UNECE (2015) *CES Recommendations for the 2020 Censuses of Population and Housing*. ECE/CES/41. Geneva: UNECE.

UNECE & Eurostat (2010) *Guidelines for Exchanging Data to Improve Emigration Statistics.* Prepared by the Task Force on Measuring Emigration Using Data Collected by the Receiving Country. Geneva: UNECE.

UNSD (1998) *Recommendations on Statistics of International Migration, Revision 1*. 98.XVII.14. United Nations Statistics Division. Available at: https://bityl.co/57xx (accessed 26 January 2016).

Urla J (1993) Cultural Politics in an Age of Statistics: Numbers, Nations, and the Making of Basque Identity. *American Ethnologist* 20(4): 818–843.

Valente P (2014) Innovative Approaches to Census-Taking: Overview of the 2011 Census Round in Europe. In: Crescenzi F and Mignani S (eds) *Statistical Methods and Applications from a Historical Perspective: Selected Issues*. Cham, CH: Springer International Publishing, pp. 187–200.

Willekens F, Massey D, Raymer J, et al. (2016) International Migration Under the Microscope. *Science* 352(6288): 897–899.

Wisniowski A, Bijak J, Christiansen S, Forster, J, Keilman, N, Raymer, J, Smith, and PWF (2013) Utilising Expert Opinion to Improve the Measurement of International Migration in Europe. *Journal of Official Statistics* 29(4): 583–607.

6

Foreigners: Inferring and Assigning

Francisca Grommé and Stephan Scheel

Introduction

Categories play a key role in the enactment of the people of Europe and their (migrant) others. In this chapter we attend to the performativity of categories (Grommé and Scheel, 2020).[1] To this end, we analyse how two statistical identity categories have been introduced and used in register-based population statistics in Estonia and the Netherlands. We focus on two data practices – inferring and assigning – through which people are allocated to these categories with the effect that some people are enacted as 'foreign', while others are enacted as 'native'. This allows us to show that statistical identity categories used for the classification and quantification of minorities and groups of migrants do not only enact the people they name as foreign. They also help to enact the majoritarian populations of the host country as an imagined community (Anderson, 2006) of 'native people'.

The two categories we attend to are the 'third generation migrant' category, which was introduced by statisticians at Statistics Estonia (SE) in 2015, and the 'Caribbean Netherlands origin' category used by Statistics Netherlands (SN). Before analysing these categories and related data practices in the second and third sections, we first explain our conceptual framework.

We locate the performative powers of categories in their sociotechnical composition. Categories used in population

statistics are social in that they are invested with and carry particular historical narratives and imaginaries of the nation. In the case of the categories analysed here, these narratives concern histories of occupation and colonisation, albeit with one important difference: the category of the 'third generation migrant' concerns the offspring of Russian-speaking inhabitants of Estonia, a group of people associated with the Soviet occupation of the Estonian nation-state (in this case the former 'colonizers' are construed as migrants). The Caribbean Netherlands origin categories are, in contrast, rooted in narratives of the Dutch colonial state (people living in colonised territories are construed as migrants). These narratives and imaginaries, we argue, operate as self-fulfilling prophecies, and help to enact the people that are allocated to the respective category as 'foreign'.

Importantly, the social ingredients of categories acquire their performativity through their relations with technical aspects of categories. In brief, statistical categories are elements of wider method assemblages whose composition shapes how these categories are *done* – that is, used and operationalised – in practice. The two statistical identity categories analysed in this chapter are used in register-based statistics which reuse administrative data stored in government registers for operational and statistical purposes. The move from traditional questionnaire-based towards register-based statistics is promoted by international statistical organisations like UNECE or Eurostat as a way to save costs, reduce the response burden and produce more timely statistics (Schulte Nordholt, 2018). SE and especially SN belong to the 'early adopters' of register-based statistics in Europe and try to promote themselves as innovative, leading NSIs in the transnational field of statistics. However, a central implication of the move towards register-based statistics is that people no longer allocate themselves

to these categories through practices of self-identification (for instance by ticking a box on a questionnaire as discussed in Chapter 7). Rather, they are allocated to these categories by statisticians through particular data practices that make use of data held in various administrative registers about the people in question. We analyse two of those data practices: inferring and assigning. Our analysis suggests that the methodological move towards register-based statistics implicates a shift towards origin-based categories used for the classification and quantification of the people of Europe, with significant consequences for related politics of (national) belonging.

To fully appreciate the significance of this observation, it is necessary to consider the use of statistical identity categories in Europe more generally. With 'identity categories' we refer to categories used by statisticians and various other actors to assign people a collective identity, often defined along lines of cultural background, ethnicity, or origin (Kertzer and Arel, 2002).[2] Identity categories may serve to monitor the 'integration' of minority groups into the majoritarian culture, but also to monitor discrimination against such groups. In some countries they are also vital in the distribution of rights and resources to minority groups (Hoh, 2018; Simon, 2012). Many of the migrant and minority identity categories used in western European countries today have been developed to record immigration after decolonisation and post-WWII labour migration. In many eastern European countries identity categories have, in contrast, been inherited from the Soviet period (Hirsch, 2000) and then developed in close relation to border changes and state-building efforts after the collapse of the Soviet Union and former Yugoslavia (Hoh, 2018; Kertzer and Arel, 2002). Even though notions about culture, ethnicity, and origin are central to identity categories, they are not part of the core EU census programme.[3]

Moreover, most NSIs in EU member states do not collect statistics on ethnicity (with the exception of Ireland and the UK). Reasons for this include data protection, but also the absence of a consensus about how ethnicity (or related terms) should be defined and measured. Nevertheless, as a response to the lack of data about ethnicity, statisticians often turn to proxies in population statistics, such as citizenship, first language, or country of birth (Simon, 2012). The move to register-based statistics promotes the use of birthplace – and even the use of the place of birth of a person's parents and grandparents – as a proxy for ethnicity, as our analysis suggests. This trend expresses a resurgence of what Kertzer and Arel call *census primordialism*: 'the equation of ethnic or national identity with ancestral identity' (2002: 27), and thereby, the entrenchment of notions of ethnic purity in official statistics.

With the conceptual framing and arguments developed in this chapter, we seek to contribute to existing works and debates on the performativity of categories. Bourdieu noted these performative effects when he aptly observed that categories used to name and classify groups along lines of ethnic origin or cultural background 'may *contribute to producing* what they apparently describe or designate' (Bourdieu 1991, 220; italics in original); a point also emphasised by Hacking (2007) and Brubaker (2002) in the context of ethnicity and nationalism. One aspect highlighted by our analysis is that identity categories used for classifying and quantifying migrant and minority populations enact much more than the people that they name. They also help to enact the majoritarian population of the host country as an imagined community of shared history and values, thus contributing to the shaping and reproduction of particular national identities and related politics of belonging.

The following analysis is driven by the question of how and through what kind of data practices statistical categories

enact certain groups as native or foreign. Our interest is how statistical identity categories are put to use in statistics and with what kind of data practices people are allocated to the categories in question. We engage with this question by analysing two data practices: *inferring* and *assigning*. First, we show, based on an analysis of the category of the 'third generation' in Estonia, how the 'foreignness' or 'nativeness' of an individual is inferred from data about the place of birth of their grandparents. Subsequently, we attend to how individual subjects are assigned to the origin-based category 'Caribbean Netherlands' through the data practice of assigning.

Before engaging in these analyses, we first discuss our understanding of the performativity of categories and how data practices feature in this. We explore and illustrate this understanding in relation to the Estonian case, or more specifically, in relation to an attempt of government bodies involved in migration and integration policies to develop shared definitions of key concepts and categories. We provide this illustration to convey a rich picture of how statistical categories are embedded in intricate policy hinterlands, and how these categories help to sustain these hinterlands. Even though we develop this conceptual frame in relation to the Estonian case, many themes and issues (e.g. integration and assimilation) are central in discussions across Europe (see, for instance, Bovens et al., 2016 for a similar discussion in the Netherlands).

On the Performativity of Statistical Categories

In 2012 a cross-institutional working group in Estonia attempted to standardise the terms and definitions of concepts used by government agencies for population and migration-related issues and policies.[4] The working group's discussions were seen as preparatory work for developing the government's 'Strategy

of Integration and Social Cohesion in Estonia 2020'. The latter was supposed to outline the government's policies vis-à-vis the countries' migrant populations, in particular Estonia's Russian-speaking minority. According to official statistics, this minority accounts for about a third of the country's population. Hence, the working group's task also included the challenge to harmonise the translation of relevant concepts into English and Russian. The task was further complicated by the fact that the working group comprised representatives from all government agencies considered as stakeholders in the policy fields of migration and integration. Hence, the working group featured representatives from the Estonian Institute of Language, which was chairing the meetings, demographers from the Institute of Population Studies, representatives from the Ministries of Culture, Interior, as well as Foreign Affairs, statisticians from SE and so forth. Unsurprisingly, the working group meetings were fraught with controversies and at times heated debates occurred about the wording and definition of particular migration-related concepts and categories.

As the comments and notes in a working document with the discussed terminology highlight, one of these controversies concerned the meaning and definition of the term integration – the core concept of the entire government strategy.[5] A representative of the Estonian Ministry of Culture proposed, for example, to define integration as 'a multilateral … process towards creation of social cohesion in society between persons with different linguistic and cultural backgrounds' (Estonian Institute of Language, 2012: 2). In a parenthesis the representative added 'we could also say "different ethnic backgrounds"', but nevertheless advised against the use of ethnicity because 'a Russian-speaking resident could also identify himself/herself as an ethnic Estonian' (ibid.). With this additional remark the ministry's representative was referring to the statistical

category 'ethnic nationality', which Estonia inherited from the Soviet period (cf. Hirsch, 2000).

While the working group emphasised that the definition of integration would require 'further discussion', it nevertheless reached agreement on the following points. First, the definition 'should definitely specify that it [integration] is a continuous two-way process encompassing immigrants and the host society' (ibid.). Second, the new government strategy on integration 'should include a differentiation of stages of integration' and provide concrete 'descriptions of the stages' (ibid.). In brackets, the working group also made proposals for these different stages of integration: '(facilitating adaptation of new arrivals; creating ties between adapted immigrants and society, fostering social cohesion)'. The decision to adopt a relatively liberal conception of integration as a 'two-way process' was motivated by the requirement to align national definitions with definitions suggested by EU institutions in order to assure cross-national comparability and policy dialogue across member states. This is indicated by a direct reference to the 2012 version of a glossary on migration and integration-related terms that is continuously developed by the European Migration Network, a network of national migration administrations (for the EMN's definition of integration see: European Commission, 2012: 103).

The second decision, to specify different stages of integration, as well as the suggestions for these stages, indicate in turn, that there were also forces lobbying for more traditional notions of integration that follow an assimilationist model, which prioritises the gradual adaptation of people considered as immigrants to the cultural and linguistical norms and values of the host society. The proposal to include these stages aimed at the gradual assimilation of migrants in the name of 'social cohesion' confirms many of the criticisms that

have been raised in regards to the notion of integration (e.g. Boersma, 2018; Korteweg and Yurdakul, 2009; Schinkel, 2013).

Conceptualising Categories: Three Aspects

The working group's discussions illustrate three points in our conception of categories and related data practices identified in Chapter 2. First, statistical categories emerge as elements of wider method assemblages, which comprise human actors, bodies of knowledge, institutions as well as material and technical devices. Thus, neither data practices, nor the categories to which they refer, can be studied in isolation of other and often enduring practices, devices, and conventions of the method assemblage to which they are related in various and changing ways. In the example above, the definition of integration is, for instance, related to (and influenced by) already existing definitions (of the EMN), institutional pressures for 'harmonisation', the assimilationist agenda of some members of the working group as well as the existence of other statistical identity categories inherited from the Soviet period (such as ethnic nationality or mother tongue).

From this follows, second, that categories and related data practices, rely on 'a large hinterland' (Law, 2004: 31) of infrastructures, conventions and other, often routinised practices which enable and configure particular data practices and how these practices are *done*. Hence, the notion of the hinterland highlights that data practices are relational, as explained in Chapter 2. Investigating these relations is important in order to take into account a set of pre-existing social and material realities (Law, 2004: 13) and how these shape and inform the data practices under study. Categories and related data practices that do not cohere with a set of established theories, influential statements, inscription devices, authorised communities of

practice, institutionalised forms of expertise and so forth will not be successful in enacting the realities that they describe, name, classify, or enumerate (Law, 2009). To appreciate the performativity of categories we therefore need to study related material-semiotic practices and the assemblages and infrastructures of which they are part (Bowker and Star, 1999).

The controversies about the definition of integration shows, third, that categories and related data practices are *of* the social but also work *on* the social world. The working group's controversies about the definition of integration illustrate that categories are of the social in the sense that they have particular advocates as well as adversaries that pursue certain political agendas which shape (and are carried by) these categories. However, categories and related definitions also work on and help to shape the social world we live in as they influence and guide the doings and makings of actors. How integration is defined and in what kind of stages it is subdivided will influence concrete interventions of government in this policy field as well as related expectations vis-à-vis the targets of these interventions and calculations of government (for the Estonian case see: Cianetti, 2015). This highlights the performativity of categories and related data practices: both are not only of the social but also work on (and help to enact) the social (Law and Urry, 2004).

Importantly, attending to the performativity of categories allows for a more powerful and radical critique of identity categories. Social constructivist studies of censuses and statistical practices highlight that ethnic and racial classifications are the result of the discussions and negotiations of a multiplicity of actors vying 'over that most basic of powers, the power to name, to categorise, and thus to create social reality' (Kertzer and Arel 2002: 36; Starr 1992; Yanow 2003). An influential critique of identity categories builds on this

observation. It stresses that statistical identity categories are variable and contingent as they change over space and time (Jenkins, 1994; Kertzer and Arel, 2002; Yanow, 2003). From this, it follows that ethnic and racial categorisation schemes are, essentially, 'human inventions, created to *impose* some sense of order on the *surrounding* social world' (Yanow 2003: vii; emphasis added). Likewise, Loveman regards ethnic and 'racial categorization schemes as cultural *impositions* on human diversity, not merely descriptive of that diversity' (2014: 14; emphasis added). Although these criticisms point out crucial issues, they remain within the realist register of representation: they assume 'the social world' as an external reality existing independently of the categories and enumeration practices mobilised to classify and quantify it. Thus, they leave open the possibility of an adequate categorisation scheme that could capture and do justice to the immense human diversity 'out there'. Hence, these criticisms implicitly confirm the very assumption they seek to abandon, namely that 'identities can be reduced to an essential core within each individual, a core that exists outside of politics' (Kertzer and Arel 2002: 19).

Enacting Foreignness, Nativeness, and Nationhood

A conception of statistical categories that accounts for their performativity as well as their sociotechnical dimension moves the critique of categories from a question of offering a more or less accurate representation of the real to the question of the very constitution of that real. This move is enabled by material-semiotic approaches which assume that categories do not just describe or represent an already existing social fabric 'out there', but rather help to constitute and sustain it in particular ways. The names of particular identity categories may already exist in everyday use and social repertoires, but it is through

their use in official statistics that identity categories formalise, restructure, and organise everyday experience, and therefore become complicit in enacting the very social realities they allegedly only represent and describe. This is why 'statistical categorizations both *reflect* and *affect* the structural divisions of societies' (Simon, 2012: 1368). The performativity of identity categories is particularly pronounced when the power of naming is combined with the authority of numbers, which rests in turn on the dominant dogma of 'statistical realism'. Yanow (2003: 11) aptly summarises this as follows: 'Naming a category asserts its importance; counting its members further underscores this.'

The performativity of statistical identity categories is, however, not reducible to the authority attributed to official statistics. It also resides in a set of mostly tacit assumptions that are ingrained in the categories in question and that carry and reproduce certain premises about the character of the social world, but also tacit political agendas, economic interests, and technical affordances. While these assumptions are often not made explicit, they operate, nevertheless, as self-fulfilling prophecies that bring into being and shape the realities to which they refer. This is well illustrated by the working group's discussion on how to name Estonia's 'original population'. As the following account shows, the eventual decision on this question was heavily influenced by the implications of (national) migration histories.

A central point of contention was whether 'majority population' or 'native population' (in Estonian: põlisrahvastik) would be the more adequate term for the 'original population of Estonia' (Estonian Institute of Language, 2012: 12). Eventually, the term majority population was rejected because it just describes a 'numerical majority', as one working group member emphasised. Accordingly, native population is the

more adequate term because it 'emphasises origin' and therefore better corresponds to its opposite, 'the foreign-origin population' (ibid.). One assumption that is carried by the category native population is that Estonia's 'original population' is not only a numerical majority, but also embodies a qualitative difference to groups considered as foreign, a difference that is framed in terms of origin.

The implications for the enactment of the Estonian nation as an imagined community become apparent in discussions about how to name what was considered as the non-native. The latter were eventually subsumed under the term 'foreign-origin population'. In this instance a debate ensued that if 'foreign-born population' – a widely used international concept – was not a better term. However, foreign-born population was discarded because of one group not considered as foreign, namely the third generation of Estonian emigrants who return to their 'home country', that is, Estonia. While the people concerned were foreign-born in the sense that they were – just like their parents – born outside of the territory of Estonia, they were considered to be part of the 'native population' (11). By rendering foreignness not as a question of country of birth, but as a question of ancestry it was possible to include 'some foreign-born persons [as] part of native population' (ibid.). What is not made explicit in the notes of the working document is that through this move to foreign-origin thousands of people born and raised in Estonia were excluded from the native population because they were the children or grand-children of Russian-speaking residents of Estonia.

The foregoing highlights that categories used for migrants and minorities in statistics carry particular imaginaries and narratives of the nation and related notions of belonging. This is why categories help to enact much more than the people that they name. Categories are relational and also entail

tacit assumptions about groups and individuals that are not considered to belong to the category in question. Hence, categories enact what Law calls 'collateral realities', that is, 'realities that get done incidentally and along the way' (Law 2012: 156).

For example, the Eurobarometer survey measures attitudes on a particular issue (its main aim), but also enacts – through a method in which a sample functions as a stand-in for a larger entity – a European public as a collateral reality (Law, 2009). Likewise, categories like foreign-origin population do not just name and enact the people to which they refer: they also enact a particular vision of the Estonian nation as an imagined community. As the case of Estonia shows, just because collateral realities get done 'incidentally', their effects are not negligible or of minor importance. For identity categories used in censuses and official statistics do not just help 'to *construct* and *constitute* the groups they ostensibly *describe*' (Brubaker 2009: 33; italics in original). They also help to enact – as a collateral reality – the national identity of the supposed host country. This is because identity categories used for migrants and minorities can enact the 'other', thus 'marking negatively what "we" are not' (Honig 2001: 3; cf. Said 2003; Wekker 2016). Hence, categories used for migrants and minorities also carry more or less tacit assumptions about 'the nation' and national belonging.[6] This means that, rather than representing the characteristics of minority groups, these statistical identity categories are instructive about the enactment of majoritarian groups and notions of belonging in the host countries (cf. Grommé and Scheel, 2020).

We recognise that our approach does not account for how statistical categories come to circulate and how they may come to be identified with. This would require us to study the 'double social process' (Ruppert, 2012) of how 'names interact with the named' (Hacking 2007, 294; cf. Bowker and Star

1999; Loveman 2014). At the same time, our conceptual frame underscores that the performativity of categories cannot be reduced to 'feedback loops' (Hacking, 2007) between names and the named. Rather, the performativity of categories also resides in their sociotechnical dimensions, most notably the often-tacit imaginaries, historical narratives, technical affordances, and political agendas that are carried by categories in the form of in-built assumptions which work as self-fulfilling prophecies. In this way official narratives and everyday discourses about national identity and nationhood are taken up, reified, and amplified by statistical categories used for the naming, classification, and quantification migrants and minorities in official statistics.

As indicated in the introduction, the move towards register-based statistics raises an important question regarding statistical categories. While traditional, questionnaire-based methods call on subjects to self-identify with pre-given or self-chosen categories, thus efficiently operating as forces of subjectivation (Cakici and Ruppert, 2020; see also Chapter 7), this self-allocation to categories is no longer possible in the context of register-based statistics. Consequently, the move to register-based statistics raises the question how – and through what kind of data practices – individual subjects are allocated to particular statistical categories, if this allocation is no longer based on the individual's self-identification.

In his seminal analysis account of population statistics, Desrosières (1998) refers to the allocation of individual subjects to statistical categories and related classification practices as encoding. What Desrosières emphasises is how such practices contribute to the stability of categories. First, he points out that, even if a category is tentative at first, each 'basic act of recognition and designation ("this is a …")' gives it 'new life by reactivating a category – just as a path only survives if it is taken' (1998: 277).

Second, to endure, a category must be supported by bodies of knowledge, experience, and expertise built around it. References to it make it increasingly solid, and stories develop around categories as they become common bases for research, policy, and government interventions. However, besides these valuable insights on categories and related classification practices, Desrosières' account does little to explain how encoding is done in practice. In the following analysis we therefore attend to two data practices that we understand as particular modes of encoding: inferring and assigning. We first continue with our analysis of the Estonian case to learn how inferring is done, after which we turn to the case of the Caribbean Netherlands for an analysis of assigning. Together, these cases demonstrate different, but related, ways of doing foreignness through statistical categories.

Estonia: Inferring Foreignness

Since December 2015 data can be retrieved from SE's statistical database about a new category of people: the 'third generation of the foreign-origin population'. In December 2015 two statisticians were busy with calculating tables on this new category of people in relation to various characteristics like sex, age, spatial distribution in Estonia by county, educational background, and unemployment rates.[7] These tables were uploaded to the new 'integration indicator database' (IID) which is meant to provide 'a single information point for finding and monitoring data on integration of different ethnic groups in Estonian society'.[8]

At first glance, the new category appears as a neutral denominator of the 'third generation of the foreign-origin population' (hereafter: 'third generation') free of any distinctions along lines of race or ethnicity. In practice, the category refers, however, for the most part, to the offspring of Estonia's

* Latest update: 14.12.2016

PO07: NATIVE AND FOREIGN-ORIGIN POPULATION, 1 JANUARY by County, Year, Native / foreign-origin population, Sex and Age group

	Males and females Age groups total
Whole country	
2015	
Native population	956 653
Foreign-origin population	354 565
..first generation of foreign-origin population	171 886
..second generation of foreign-origin population	105 036
..third generation of foreign-origin population	77 643
Origin unknown	2 053
2016	
Native population	955 471
Foreign-origin population	358 843
..first generation of foreign-origin population	178 200
..second generation of foreign-origin population	103 434
..third generation of foreign-origin population	77 209
Origin unknown	1 630

Footnote:
The data for 2015 have been revised on 14.12.2016.

Figure 6.1 SE Table of the 'Native and Foreign-origin Population', 1 January 2015 and 2016[a]

[a]Screenshot from SE's statistical database: https://andmed.stat.ee/et/stat (accessed 4 May 2017)

Russian-speaking inhabitants which account, according to official statistics, for up to one third of Estonia's population (Poleshchuk, 2009; Tammur, 2017; Vetik, 2011).

It is only possible to retrieve information on the third generation from SE's homepage from the year 2012 onwards (see Figure 6.1). The reason is simple: the construction of this category of people relies on data that was not collected prior to the population and housing census (PHC) in 2011. In the PHC 2011 it was decided to include an additional question in the census questionnaire which inquired about the place of birth of grandparents.[9] Responses to this question are used to determine whether an individual is part of the third generation, which is based on ancestry as set out in the official definition: any person 'permanently living in Estonia of whose parents at least one was born in Estonia but whose grandparents were all born abroad'.[10]

What the definition highlights is that implementing categories such as third generation require specific data practices

for encoding individuals. In the case of the third generation it is not an individual's citizenship, their language capacities or self-identification with a particular ethnicity or nationality that determines their foreignness or nativeness. Instead, an individual's status as native or foreign is *inferred* from the place of birth of their grandparents. If a person's grandparents were all born abroad, the person is encoded as of 'foreign-origin'. Conversely, a person will be categorised 'native', if their grandparents were born in Estonia, even if they as well their parents were born abroad. In this way foreignness and nativeness are enacted on the basis of ancestry, as a feature inherent to a person that is inherited and cannot be altered as a result of their actions, beliefs, or practices of self-identification (see Chapter 7). Rather, inferring is a data practice that draws on data that cannot be influenced by the person concerned – the place of birth of their grandparents. The main accomplishment of inferring nativeness and foreignness from the place of birth of a person's ancestors is thus the capacity to subdivide the resident population into two stable subgroups – the foreign and the native – that can be assessed, compared, and monitored with all kinds of statistical indicators, such as educational background, fertility rates, income, employment rate, age distribution and so forth. It is the inferring of nativeness and foreignness which provides the epistemic basis for integration monitoring of all sorts.

Inferring thus highlights both the performativity and the sociotechnical characteristics of data practices. What inferring helps to perform is a particular version of the twin-concepts foreignness and nativeness, which are not only enacted as a mutually exclusive opposition, but also as dependent variables of ancestry. That such inferring relies on the inclusion of a new question on the census questionnaire highlights in turn the sociotechnical character of this data practice.

Inferring nativeness and foreignness relies on a hinterland of other actors, devices, and data practices that make up the questionnaire-based Estonian census that asks subjects to report the place of birth of their grandparents without revealing the reasons for and uses to which their answers will be put. Through the responses of subjects, statisticians can infer their status as native or foreign and encode them into the category of the third generation.

The importance of ancestry in the definition of the third generation highlights how this statistical category carries a particular historical narrative about the Estonian nation-state. This historical narrative becomes apparent if one considers that the definition of the third generation features Estonia as a spatial reference point for the place of birth of the parents and grandparents. The narrative is peculiar insofar as Estonia did – de facto – not exist as an independent nation-state when most of the parents and grandparents of the people labelled as third generation were born. Rather, the territory of what is today known as Estonia was part of the Soviet Union between 1939 and 1991. The category of the third generation thus enacts a central element of the official historical narrative of Estonia: while the Estonian nation-state did not exist de facto during the Soviet period, it never ceased to exist de jure, leading a virtual existence of legal continuity during a period officially known as occupation that lasted more than 60 years.

Hence, the category of the third generation enacts much more than the people it names: it carries a particular version of the history of the Estonian nation, a history imagined in terms of both (de jure) legal continuity and (de facto) rupture of Estonian nationhood. The notion of the rupture 'stands for the interruption and deterioration of the harmonious national development [of the Estonian nation-state] of the pre-war independence era' (Jõesalu and Kõresaar, 2013: 183). This is

the dominant script in re-independent Estonia for interpreting the Soviet era, which is disavowed as a brutal occupation characterised by violent repression, ideological pressure and political persecution (184). The script frames the communist regime of the Soviet period as the occupation by an external force that is construed as foreign in both ideological and ethnic terms (Troebst, 2006: 79–80). In this way the script of the rupture 'has developed a strong ethnic and national repertoire [… that] differentiates [among the inhabitants of re-independent Estonia] between the carriers of "our own" national history (the Estonian middle class and farmers) and the carriers of "alien" history (communists and Russians)' (Jõesalu and Kõresaar, 2013: 184). This stark distinction along ethnic lines is carried by the category of the third generation. It is enacted by the data practice of inferring when people whose grandparents were born outside Estonia are allocated to this category as of foreign origin – despite the fact that people defined as such were born and have grown up in re-independent Estonia.

An implication of inferring foreignness and nativeness from the place of birth of a subject's grandparents is that Russian-speaking inhabitants of Estonia are enacted as immigrants despite the fact that they may have never crossed an international border: whereas members of the third generation have been born in re-independent Estonia, members of the second generation have mostly been born in a part of the Soviet Union that became Estonia in 1991, while members of the first generation initially settled in that corner of the Soviet Union, mostly during the 1960s and 1970s. By making ancestry the central criterion for national belonging, origin-based categories and related data practices like inferring essentialise alleged cultural differences, enacting them as immutable. This was, in fact, the impetus driving the introduction of the category. A demographer who lobbied in the scientific council of

the 2011 census to add a question on the grandparents' place of birth to the census questionnaire summarised the rationale for the new identity category as follows: 'You can change your mother tongue, you can decide to identify as Estonian; you can even change your citizenship. But you cannot change the place of birth of your grandparents!'.[11]

The demographer's reference to mother tongue and self-identification with a particular nationality point to alternative statistical identity categories used in Estonian population statistics. These are the categories of 'mother tongue' (first language) and 'ethnic nationality' which have been inherited from the Soviet period. Both categories played a central role in the state-building of the Soviet Union as a multinational socialist federation (Hirsch 1997: 2000). Furthermore, both categories are based on self-identification with a particular national culture.[12] The methodology of self-identification is, however, precisely the reason why these categories are dismissed as subjective by statisticians who contrast these 'unreliable indicators' with the 'objectivity' of information on the place of birth of the grandparents.[13]

The crucial point is that the allegedly objective criteria of place of birth of grandparents, and the ancestry-based distinction between native and foreign-origin population accomplished by the data practice of inferring, enact Estonia as a decisively *ethnic* nation (Poleshchuk, 2009). With the help of the data practice of inferring Estonia is enacted as a nation that is built around a 'myth of a common origin or shared blood/genes' (Yuval-Davis, 2007: 21). Whether a person is considered to be member of the 'foreign' [*välispäritolu*] or of the 'native' [*pölis*] population, and thus as a member of the imagined national community of Estonia, depends not on their legal citizenship, place of birth, language capacities

or self-identification, but on their 'cultural background'.[14] This cultural background is, however, essentialised as it is inferred from a person's 'roots', that is, the cultural background of their biological parents and grandparents which is, in turn, territorialised as it is inferred from their respective place of birth.

In sum, the data practice of inferring thus enacts what Alba calls a 'bright boundary' – a boundary between a native and a foreign population that is unambiguous and difficult, if not impossible, to transgress since this boundary is drawn by an 'objective' criterion that cannot be influenced by an individual: the parents' and grandparents' place of birth. Hence, belonging to the imagined community of Estonia becomes not a question of self-identification, citizenship or language faculty but a question of ancestry. This emphasis on ancestry enacts a form of nationalism that imagines the nation in terms of ethnic purity (cf. Kertzer and Arel, 2002). By declaring the place of birth of the parents and grandparents as the central criterion for the definition of the native population, national belonging is fixed to a distant past that determines the (non-)belonging of an individual to the imagined national collective in the present. Ultimately, belonging to the national community becomes an exclusive affair that is protected by an insurmountable hurdle of ethnic origin. In the following section we analyse how citizens from former colonised territories are enacted as foreign through the category 'Caribbean Netherlands' and the related data practice of assigning.

The Netherlands: Assigning Foreignness

A statistician sitting in front of a screen was checking online tables about the country of origin of residents of the Netherlands on SN's website. He was looking for the origin

group of the Caribbean Netherlands – people with at least one parent born in the Caribbean Netherlands:

> Can you come from Bonaire? No. From the Caribbean Netherlands? No. [he mumbles] People cannot come from the Caribbean Netherlands. Sorry, sometimes I do not understand the statistics made by SN [laughs]. [He finds the figures] This is possible from 2012, 10 people. But this is nonsense! It is not a new country. I am going to e-mail someone about this, it is confusing.[15]

The source of the statistician's surprise was the reference on the webpage of SN to the Caribbean Netherlands – the islands of Bonaire, Saint Eustatius, and Saba – as a new country (see Figure 6.2).[16] Colonised in the 17th century, the islands became part of a country in 1954 – the Netherlands Antilles. The 'status aparte', as it is referred to, made the Netherlands Antilles a partly self-governed entity within the Kingdom of the Netherlands (which also includes the continental Netherlands, see Figure 6.2). However, like many Caribbean countries, the islands have not been following a linear path to full independence (Bonilla 2015; Oostindie 2006). After a period of rising government debts and poverty, Bonaire, Saint Eustatius and Saba voted for closer ties with the continental Netherlands in 2006. Although not uncontested, they changed status in 2010 and became 'special municipalities' of the continental Netherlands.[17]

To refer to the Caribbean Netherlands as a country is a rather common mistake, and therefore a telling one. In Dutch population statistics, people born in the Caribbean Netherlands (and the former Netherlands Antilles) are categorised as having a 'migrant background', even though they have a legal status as citizens of the Dutch Kingdom and the EU.[18] This has significant implications for the Dutch discourse on integration, where population statistics are widely used and where

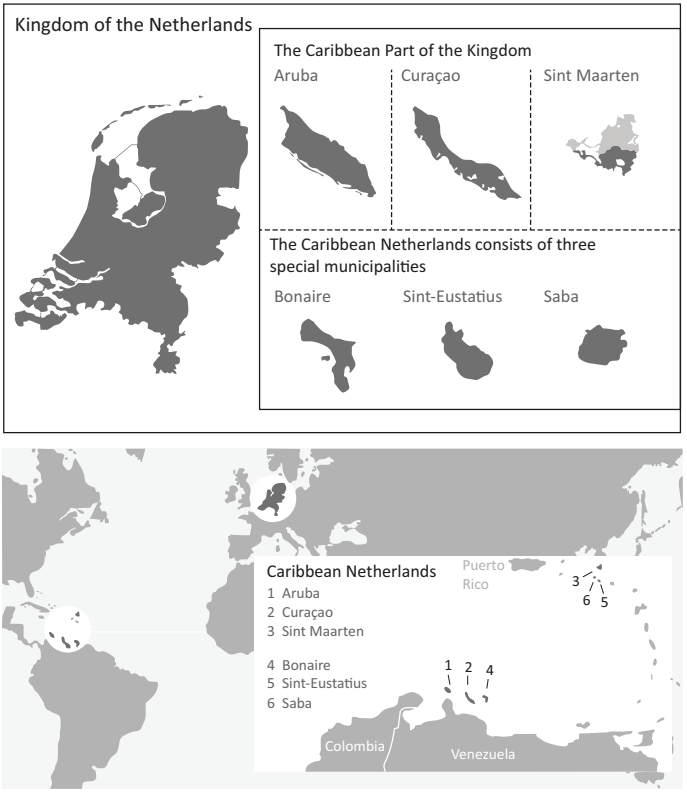

Figure 6.2 The Kingdom of the Netherlands: Continental and Caribbean Part
Source: Screenshot from Netherlands Ministry of the Interior and Kingdom Relations at: https://www.werkenvoornederland.nl/. Accessed: 18 June 2021. Translated by authors.

the foreignness of people from the Caribbean Netherlands is, partly due to this statistical categorisation practice, taken as self-evident. It is therefore of interest to inquire into the data practices that sustain taken-for-granted categories like migration background or Caribbean Netherlands and related

notions of foreignness. In this section we particularly investigate how the data practice of *assigning* sustains the Caribbean Netherlands as an origin category, thus contributing to the enactment of particular notions of foreignness, nativeness, and nationhood.

Assigning is a data practice that allocates particular individuals to a specific identity category. The case analysed in this section is especially instructive about the data practice of assigning because SN is one of the NSIs at the forefront in the move towards register-based statistics. As the following analysis shows, to sustain the category of the Caribbean Netherlands in register-based statistics new entries to the population register need to be assigned a country code specifying their country of birth that is different from that of the Netherlands. It is thus the data practice of assigning which enacts people concerned as foreign. We understand assigning as one of many possible data practices of encoding required to put categories to use, in this case for producing 'origin group' statistics at SN.

Following these insights, we examine two aspects of assigning that 'hold together' a category (Desrosières, 1998: 277): first, assigning people from the Caribbean Netherlands to country codes in population registers. We show that this aspect relies on administrative hinterlands that extend to international organisations and nation-building processes. Second, the further processing of these data for the production of demographic statistics relying on notions of origin embedded in national minority policies. Both aspects, we suggest, enact people born in the Caribbean Netherlands as foreign through essentialised notions of culture and ethnicity.

In contrast to other countries producing register-based statistics, Dutch demographic statistics are entirely based on population register data.[19] To learn more about the hinterland of the category of the Caribbean Netherlands, we first turn to

how people are assigned to particular country of birth categories in the population register. When a person settles in the Netherlands from abroad, they are required to register with a municipal population register. Most people do so because almost every aspect of life in the Netherlands requires such registration, including access to benefits, insurance, education, and employment. Registration involves a visit to the local city hall and an interview with a civil servant who enters a variety of personal data into the registration system, including their country of birth. This process is guided by an extensive set of guidelines prescribing how a claimed identity can be ascertained, how to verify data, how to correctly enter data in various fields, and so on. The guidelines also specify a country code for each country of birth, which follows an international standard, as we explain below.

The key point is that, according to these instructions, people arriving from the Caribbean Netherlands have to be registered as having arrived from a different country. Civil servants thus have to assign people from the Caribbean Netherlands to a different country code than that of the Netherlands. Moreover, the country code assigned to them will differ, depending on the island they are coming from. According to the guidelines, underlying this routine is 'the fact that all (is)lands in the Kingdom see each other as foreign in the context of the population register' (Basisadministratie Persoonsgegevens en Reisdocumenten, 2016: 22).[20]

The practice of administratively considering the Caribbean islands as separate 'nation-state(s)' is not new. While the motivations behind this decision are never explicitly stated in manuals and policy documents, it is likely that this practice is a continuation of practices that started with the partial independence of the Netherlands Antilles in 1954. First, financial transactions, IT systems and personal records

were given a distinct country code in this year. Although the islands of the Caribbean Netherlands lost their status as a part of a country (i.e., the Netherlands Antilles) when they became special municipalities after the 2010 vote, politicians from the islands and from the continental Netherlands agreed to retain the islands' economic and financial autonomy and distinctiveness as much as possible (Oostindie and Klinkers, 2012). In this sense a special municipality functions differently than a regular municipality. Second, people moving between different parts of the Kingdom had been referred to as migrants since 1954, when the direction of postcolonial reform was still more clearly envisioned as a state form close to full independence. In 2010, the use of the migrant category was still supported as civil servants and politicians in the continental Netherlands wished to continue to monitor the movement of people between the Caribbean part of the Kingdom and the continental Netherlands as migration.[21] Furthermore, after 2010, migration became a prominent theme in the politics in the Caribbean Netherlands as well. For instance, Bonairian politicians have put on the agenda increased migration from the continental Netherlands using SN statistics (Grommé, forthcoming). The implementation and continuation of these decisions and priorities required in turn data practices that mark people from the Caribbean Netherlands as distinct, such as assigning them to a different country code than that of the Netherlands.

To produce demographic statistics, each municipality shares parts of the population register with SN (cf. Prins, 2017; also see Chapter 3). In the subsequent statistical production process, the submitted register data are then subject to numerous data practices, such as the partially automated cleaning and categorising of data. These practices are based on pre-set algorithms (called 'business rules') that prescribe

how data is to be corrected and converted. For country of birth, the practice of assigning involves grouping people with country codes specific to one of the three islands to a singular country code: the Caribbean Netherlands. Although no formal explanation is provided, this practice coheres to the Eurostat country list, which draws on internationally recognised country codes drawn up by the International Organization for Standardization (ISO). In this list, the three countries that make up the Caribbean Netherlands are made singular by being designated by the country code BQ.

In sum, assigning people from the three islands to the category of Caribbean Netherlands happens through data practices at both the registration and statistical processing stages and is bound up with a hinterland of relations that extend to the ISO. Both stages contribute to the category's authority and endurance as a standard and convention and respond to ongoing bureaucratic conventions and political priorities of continental and Caribbean politicians to maintain a distinction between people from the Netherlands and its former colonies. Thus, the data practice of assigning contributes to marking people born in the Caribbean Netherlands as migrants, and thereby as foreign.

As a consequence of using standardised country codes that can be integrated in drop-down lists on a civil servant's screen, assigning people to this category can be done relatively automatically without much friction or thinking work, as routinised bureaucratic procedures based on past decisions have been inscribed into the software interface. In this way, people born in the Caribbean Netherlands are enacted as foreign although they hold Dutch citizenship and despite their municipalities being considered as part of the continental Netherlands state and administration. However, despite this black-boxing past of decisions and political priorities,

many experts, including statisticians at SN, are aware that the Caribbean Netherlands is an administrative construct referring to three island states with large distances between them.[22]

SN often publishes statistics about people born in the Caribbean Netherlands living in the continental Netherlands under a different name: 'origin group' (herkomstgroep) statistics. In what follows we will attend to two relevant Caribbean origin groups: the Caribbean Netherlands and the former Netherlands Antilles (for people born before 2010, see Table 6.1). Statistical origin groups serve the policy aim of measuring how people with a 'migration background' adjust to (what is assumed to be) Dutch culture. In Dutch minority policies this is called integration: the state needs to help people integrate

Table 6.1 The Caribbean Origin Categories[a]

Year	Islands	Name	State form	SN origin category
1954	Aruba, Curaçao, St Maarten, Bonaire, St Eustatius, Saba	Netherlands Antilles	Country within the Kingdom ('status aparte')	(former) Netherlands Antilles origin category (born before 10/10/2010)
2010	Bonaire, St Eustatius, Saba	Caribbean Netherlands or BES-islands	Special municipalities	Caribbean Netherlands origin category (born after 10/10/2010)

[a]For readability, we have omitted two events from this table. In 1986, Aruba gained *status aparte* and became a country in the Kingdom. In 2010, Curaçao and St Maarten gained *status aparte*

into society, supported by statistics that monitor this process. So not only does assigning rely on administrative hinterlands that extend to international standards and agreements between politicians across the Atlantic Ocean, it is also rooted in institutional understandings of culture and ethnicity.

Origin group statistics are about the first and second generations of people with a migration background. As in the previously discussed SN practices, making origin group statistics is a partially automated processes in which pre-set business rules categorise entries based on country codes designating the country of birth of an individual. It first requires inferring foreignness from the parents' country of birth, a data practice that we have described in the previous section. For instance, people with at least one parent born outside of the Netherlands are allocated to the second generation of people with a migration background. Furthermore, and this underlines that origin groups are not natural but informed by policies and assumptions about culture, assigning takes place along the lines of the people expected to require policy intervention to support their integration into Dutch society. Statistical tables published by SN on origin groups therefore do not make people from different countries equally visible (see Figure 6.3). Instead, first, western and non-western origin groups are distinguished and next, among the non-western origin groups, four main groups are identified and usually highlighted in publications regarding integration: Moroccan, Surinamese, Turkish, *and* the former Netherlands Antilles (a composite group including people born in the Caribbean Netherlands).

As critics have stated, this taxonomy inscribes origin categories with varying degrees of difference from a Dutch norm, thus constructing a hierarchy of geographically ranked cultures (Schinkel, 2013; Yanow and van der Haar, 2013).

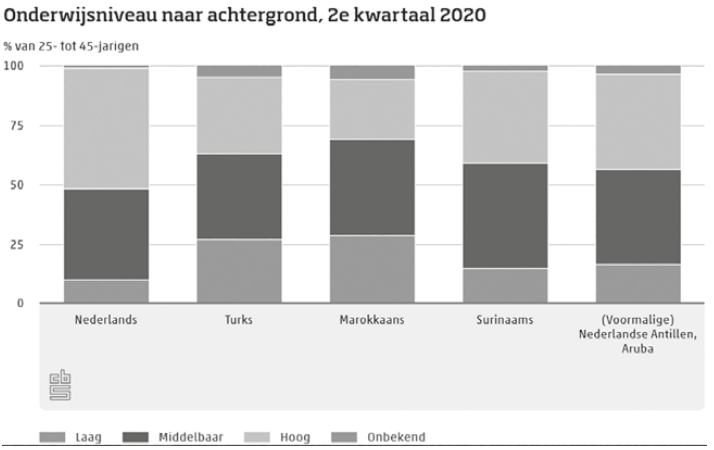

Figure 6.3 Example of Origin Group Statistics in the Context of Migration Policies[a]

[a]This figure presents the level of education of 25- to 45-year-olds from a Dutch background and from the four 'largest non-western origin groups.' Grey-shading from left to right (indicating level of education): low, middle, high, unknown. Columns from left to right: Dutch, Turkish, Moroccan, Surinamese, (former) Netherlands Antilles. Source: CBS, 2020

Here we highlight how this data practice enacts essentialised differences between people born in the Netherlands and abroad. The recent history of the terminology of the western and non-western categories helps to understand this. Until 2016, the groups were referred to as 'western allochthonous' and 'non-western allochthonous.' Allochthonous can be literally translated as 'not from the soil.' SN distinguished western and non-western allochthonous countries of birth on the basis of whether the population 'strongly resembles the Dutch population from a socio-economic or cultural perspective' (Keij, 2000: 24). Embedded in these distinctions thus is a fixed relation between place of birth (the 'soil') and the social, economic, and cultural traits of people.

Since 2016, allochthonous has been gradually replaced by migration background to avoid the negative connotations that have become associated with the term (Bovens et al., 2016). First, the term allochthonous had acquired the connotation of 'outsider', and somehow 'lagging behind'. Second, the term had accrued racial connotations as it had become a shorthand to negatively refer to people who are 'black, Muslim or both' (Groenendijk, 2007: 105). Nevertheless, the same taxonomy, relying on a distinction between western and non-western as defined in Keij (2000), is still in place. Routinely assigning people from the Caribbean Netherlands and the former Netherlands Antilles to the non-western origin group thus enacts an essentialised notion of difference based on place of birth.

Assigning people to origin groups therefore does not only enact a geographical and administrative separation as explained in the first part of this section. This data practice also carries and enacts essentialised notions about cultural difference along the lines of ethnicity and race (Abu-Lughod, 1991; Schinkel, 2013). Furthermore, a sizeable body of statistical publications now exists that both draws on and legitimates the country codes. To illustrate, a review of the last ten years of demographic publications by SN on persons of Caribbean Netherlands and Antillean origin shows that these exclusively concern the topics of urban residence, life expectancy, teenage motherhood, and single motherhood (Grommé and Scheel, 2020; cf. Krebbekx, Spronk, and M'charek, 2017).[23] When we asked about the latter two topics, a statistician responded 'we just know that this [a higher ratio of teenage mothers] is the case in the Caribbean Netherlands'.[24] The focus on gender and family norms echoes familiar colonial tropes where these aspects of life are a core area of boundary work between foreign and native populations (Bonjour and De Hart, 2013; Van

Reekum and Van den Berg, 2015). Our point is not that there are no differences between family life as practiced in the continental and the Caribbean Netherlands. The repeated choice to focus on the topics of motherhood and family relations rather illustrates how a category can hold together when a body of knowledge legitimates and supports it (Desrosières, 1998). In turn, using different country codes comes to seem like a more or less natural reflection of these boundaries between populations.

To conclude, the various aspects of assigning demonstrate the relevance of statistical practices to enacting groups of people as foreign or native. Even though the foreignness of people born in the Caribbean Netherlands (and the former Netherlands Antilles) is not a legal fact, it is enacted by assigning people to a migrant category. Assigning carries with it and mobilises various aspects that help to enact foreignness: conventions that can be traced back to international organisations such as the ISO; national policies and their embedded assumptions about the fixity of culture and ethnicity; and institutional bodies of knowledge. These aspects of the data practice can reinforce each other and make a category hold together as foreign. The automation of assigning practices has furthermore black-boxed this assemblage of agreements, conventions, bodies of knowledge, and assumptions.

As in the Estonian case, data practices not only contribute to enactments of foreignness, they also do this for nativeness. The use of the country codes complies with a notion of the continental Netherlands as an essentially stable territorial entity despite the 2010 changes. In fact, the adoption of the term 'the Caribbean Netherlands' co-occurred with the introduction of the term 'the European Netherlands' at SN, thereby enacting conceptual and symbolic boundaries between two territories that are both part of the nation-state

of the Kingdom of the Netherlands (one of the first instances of use can be traced back to the 2011 census publication, see CBS (2014)). The 'western' population of this territory, moreover, is analysed separately from the people that have moved to the Netherlands two or more generations ago but are not recognised as part of the 'original' population. Consequently, data practices contribute to attempts to fix a notion of Dutch origin, a notion that remains elusive and fragile even if it is continuously sought after in public debate (Geschiere, 2009). Nationhood, therefore, is not established once and for all by grand historical acts in the past but relies on the continuous operation of routinised data practices.

Conclusion

This chapter has attended to the performativity of statistical identity categories. We have developed a conceptual framing that locates the performativity of categories in tacit imaginaries, narratives, and political agendas that are ingrained in and carried by categories used for the classification of migrants and minorities. Central to our argument is that identity categories assigned to migrants and minorities help to enact more than the groups of people to which they refer. They also help to enact, in the form of collateral realities, notions of national identity and belonging of majoritarian groups. Hence, identity categories are analytical entry points to study the articulation of particular forms of nationalism and national belonging.

Based on this framing, we have investigated two statistical identity categories – the third generation migrant and the Caribbean Netherlands – and the data practices that are mobilised to allocate individual subjects to these categories in the context of register-based statistics. In case of the third generation category, individuals are allocated based on the place

of birth of their grandparents. Hence, people's status as foreign or native is *inferred* from data on the place of birth of the ancestors of the person in question. In case of the Caribbean Netherlands people are *assigned* to a statistical category based on standardised country codes and place of birth and thus enacted as foreigners.

In sum, our analysis allows us to make three observations regarding statistical identity categories and related politics of national belonging: first, it is neither an anonymous power nor clearly identifiable actors like political nationalists or 'ethnopolitical entrepreneurs' (Brubaker, 2002: 166) that reproduce exclusive understandings of national belonging and divisions along lines of ethnicity and nationality. Our analysis rather suggests myriad relational practices of various actors such as statisticians and the mobilisation of various sociotechnical devices like categories help to enact ethnic divisions and national identities. Hence, nation-building is not reducible to a one-time foundational act in a distant, mystified past. It emerges as an iterative process accomplished by the sociomaterial practices of actors like statisticians. Second, the move to origin-based categories that we diagnose in this chapter is related to a methodological shift – with consequences for the politics of (national) belonging. Both cases show a shift from questionnaire-based methodologies based on self-identification towards register-based methodologies that use existing administrative data. As part of this shift, data practices like inferring or assigning are mobilised to encode people to particular categories of ethnic or national belonging based on data about them (and their ancestors) in administrative registers. This chapter therefore demonstrates that the politics of belonging are intertwined with a politics of method. Finally, our analysis confirms how the reluctance of many EU member states to collect data on ethnicity and cultural background

is increasingly side-lined and superseded through the use of origin-based categories and related conceptions of migration as a feature that is inherent to an individual (cf. Elrick and Schwartzman, 2015; Renard, 2018).

In this way, the cases of the third generation and the Caribbean Netherlands categories are instructive about a central theme of this book: the dominant assumption that people are sedentary (or should be treated as such). This assumption does not only find its expression in efforts to locate people to a single address (see Chapter 3). It is also related to the notion that a 'people' is a bounded group that is historically linked to a single, stable, and delineated geographical location (their origin) and which is inextricably bound up with their identity. However, notwithstanding the growing salience of claims to origin, historical essence or authenticity, the nature of the essence itself typically remains elusive when experts or political advocates attempt to define it (Geschiere, 2009). We suggest that data practices such as inferring and assigning are part of efforts to find a (ultimately unfulfilled) solution to this conundrum.

References

Abu-Lughod L (1991) Writing Against Culture. In: Fox R (ed.) *Recapturing Anthropology: Working in the Present*. Santa Fé: School of American Research Press, pp. 137–162.

Anderson B (2006) *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.

Basisadministratie Persoonsgegevens en Reisdocumenten (2016) *Verschillen Persoonslijst GBA Versus PIVA*. LO GBA 3.10 / LO BES 1.3. The Hague: Ministerie van Binnenlandse Zaken en Koninkrijksrelaties.

Boersma S (2018) *Dissociating Society: Knowledge, Affect and Performativity in Immigrant Integration Monitoring*. PhD Thesis. Erasmus University Rotterdam, Rotterdam, NL.

Bonilla Y (2015) *Non-Sovereign Futures: French Caribbean Politics in the Wake of Disenchantment*. Chicago: University of Chicago Press.

Bonjour S and De Hart B (2013) A Proper Wife, a Proper Marriage: Constructions of 'Us' and 'Them' in Dutch Family Migration Policy. *European Journal of Women's Studies* 20(1): 61–76.

Bourdieu P (1991) Identity and Representation. Elements for a Critical Reflexion on the Idea of Region. In: *Language and Symbolic Power*. Paperback Edition. Cambridge: Polity Press, pp. 220–228.

Bovens M, Meike B, Jennissen R, et al. (2016) *Migratie en Classificatie: Naar een Meervoudig Migratie-Idioom*. WRR-verkenning nr. 34. The Hague: WRR. Available at: www.wrr.nl/publicaties/verkenningen/2016/11/01/migratie-en-classificatie-naar-een-meervoudig-migratie-idioom-34 (accessed 11 July 2017).

Bowker GC and Star SL (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Brubaker R (2002) Ethnicity without Groups. *European Journal of Sociology* 4(2): 163–189.

Brubaker R (2009) Ethnicity, Race, and Nationalism. *Annual Review of Sociology* 35: 21–42.

Cakici B and Ruppert E (2020) Methods as Forces of Subjectivation: Experiments in the Remaking of Official Statistics. *Journal of Cultural Economy* 13(2): 221–235.

CBS (2014) *Dutch Census 2011: Analysis and Methodology*. The Hague: Statistics Netherlands.

CBS (2020) Hoe Verschillen Opleiding en Schoolkeuze naar Migratieachtergrond? Available at: www.cbs.nl/nl-nl/dossier/dossier-asiel-migratie-en-integratie/hoe-verschillen-opleiding-en-schoolkeuze-naar-migratieachtergrond- (accessed 2 September 2020).

Cianetti L (2015) Integrating Minorities in Times of Crisis: Issues of Displacement in the Estonian and Latvian Integration Programs. *Nationalism and Ethnic Politics* 21(2): 191–212.

Desrosières A (1998) *The Politics of Large Numbers: A History of Statistical Reasoning* (ed. RD Whitley). Cambridge, MA and London: Harvard University Press.

Elrick J and Schwartzman LF (2015) From Statistical Category to Social Category: Organized Politics and Official Categorizations of 'persons with a Migration Background' in Germany. *Ethnic and Racial Studies* 38(9): 1539–1556.

Estonian Institute of Language (2012) *Summary of Integration Terminology Workshops of 9 May and 9 Sept (with the decisions and proposals of the steering group meeting of 29 October)*. Unpublished document. Tallinn: Estonian Institute of Language.

European Commission (2012) *Asylum and Migration Glossary 2.0. A Tool for Better Comparability Produced by the European Migration Network*. Luxembourg: European Migration Network.

Geschiere P (2009) *The Perils of Belonging: Autochthony, Citizenship, and Exclusion in Africa and Europe*. Chicago: University of Chicago Press.

Groenendijk K (2007) Allochtonen of Burgers: Definitiemacht in Debat en Wetgeving over Immigranten. In: Duyvendak JW, Engbersen G, Teeuwen M, et al. (eds) *Macht En Verantwoordelijkheid: Essays Voor Kees Schuyt*. Amsterdam: Amsterdam University Press, pp. 101–112.

Grommé F (forthcoming) Thinking, Seeing, and Doing Like a Kingdom: The Making of Caribbean Netherlands Statistics and the 'Native Bonairian.' In: Guadeloupe F and Pijl Y (eds) *Equaliberty in the Dutch Caribbean: Ways of Being Non/Sovereign*. New Brunswick, NJ: Rutgers University Press.

Grommé F and Scheel S (2020) Doing Statistics, Enacting the Nation: The Performative Powers of Categories. *Nations and Nationalism* 26(3): 576–593.

Hacking I (2007) Kinds of People: Moving Targets. *Proceedings of the British Academy* 151(1): 285–318.

Hirsch F (1997) The Soviet Union as a Work-in-Progress: Ethnographers and the Category Nationality in the 1926, 1937 and 1939 Censuses. *Slavic Revue* 56(2): 251–278.

Hirsch F (2000) Toward an Empire of Nations: Border-Making and the Formation of Soviet National Identities. *Russian Review* 59(2): 201–226.

Hoh A-L (2018) *Counting for EU enlargement? Census-taking in Croatia, Bosnia and Macedonia*. PhD Thesis. Maastricht: University of Maastricht.

Honig B (2001) *Democracy and the Foreigner*. Princeton, NJ and Oxford: Princeton University Press.

Jenkins R (1994) Rethinking Ethnicity: Identity, Categorization and Power. *Ethnic and Racial Studies* 17(2): 197–223.

Jõesalu K and Kõresaar E (2013) Continuity or Discontinuity: On the Dynamics of Remembering 'Mature Socialism' In Estonian Post-Soviet Remembrance Culture. *Journal of Baltic Studies* 44(2): 177–203.

Keij I (2000) Standaarddefinitie Allochtonen. *Index* (10): 24–25.

Kertzer DI and Arel D (2002) Censuses, Identity Formation, and the Struggle for Political Power. In: Kertzer DI and Arel D (eds) *Census and Identity: The Politics of Race, Ethnicity, and Language in National Censuses*. Cambridge and New York: Cambridge University Press, pp. 1–42.

Korteweg A and Yurdakul G (2009) Islam, Gender, and Immigrant Integration: Boundary Drawing in Discourses on Honour Killing in the Netherlands and Germany. *Ethnic and Racial Studies* 32(2): 218–238.

Krebbekx W, Spronk R and M'charek A (2017) Ethnicizing Sexuality: An Analysis of Research Practices in the Netherlands. *Ethnic and Racial Studies* 40(4): 636–655.

Law J (2004) *After Method: Mess in Social Science Research*. London and New York: Routledge.

Law J (2009) Seeing Like a Survey. *Cultural Sociology* 3(2): 239–256.

Law J (2012) Collateral Realities. In: Ludwig D, Boogaard B, Macnaghten P, and Leeuwis C (eds) *The Politics of Knowledge*. Oxon and New York: Routledge, pp. 156–178.

Law J and Urry J (2004) Enacting the Social. *Economy and Society* 33(3): 390–410.

Loveman M (2014) *National Colors Racial Classification and the State in Latin America*. Oxford: Oxford University Press.

Oostindie G (2006) Dependence and Autonomy in Sub-National Island Jurisdictions: The Case of the Kingdom of the Netherlands. *The Round Table* 95(386): 609–626.

Oostindie G and Klinkers I (2012) *Gedeeld Koninkrijk: De Ontmanteling van de Nederlandse Antillen en de Vernieuwing van De trans-atlantische relaties*. Amsterdam: Amsterdam University Press.

Poleshchuk V (2009) *The War of the Monuments in Estonia: The Challenges of History and the Minority Population*. Mariehamn, Åland, Finland: Åland Islands Peace Institute.

Prins K (2017) *Population Register Data, Basis for the Netherlands Population Statistics*. Bevolkingstrends. The Hague: Statistics Netherlands.

Renard L (2018) Mit den Augen der Statistiker. Deutsche Kategorisierungspraktiken von Migration im historischen Wandel. *Zeithistorische Forschungen* 15(3): 431–451.

Ruppert E (2012) Category. In: Lury C and Wakeford N (eds) *Inventive Methods: The Happening of the Social*. CRESC. London: Routledge, pp. 36–47.

Said EW (2003) *Orientalism*. 25th Anniversary Edition. London: Penguin Books.

Schinkel W (2013) The Imagination of 'Society' in Measurements of Immigrant Integration. *Ethnic and Racial Studies* 36(7): 1142–1161.

Schulte Nordholt E (2018) The Usability of Administrative Data for Register-Based Censuses. *Statistical Journal of the IAOS* (Preprint): 1–12.

Simon P (2012) Collecting Ethnic Statistics in Europe: A Review. *Ethnic and Racial Studies* 35(8): 1366–1391.

Starr P (1992) Social categories and claims in the liberal state. In: Douglas M and Hull D (eds) *How Classification Works*. Edinburgh: Edinburgh University Press, pp. 263–295.

Tammur A (2017) Native and Foreign-Origin Population in Estonia. *Statistical Bulletin of Statistics Estonia* 2017(1): 50–56.

Teulieres L (2007) Immigration and National Identity: Historiographical Experiences in France. In: Ellis SG and Klusáková L (eds) *Imagining Frontiers, Contesting Identities*. Pisa: Pisa University Press, pp. 43–58.

Troebst S (2006) Postkommunistische Erinnerungskulturen im östlichen Europa: Bestandsaufnahme, Kategorisierung, Periodisierung. In: *Kulturstudien Ostmitteleuropas. Aufsätze Und Essays*. Frankfurt am Main: Peter Lang Publishing, pp. 65–108.

Van der Pijl Y and Guadeloupe F (2015) Imagining the Nation in the Classroom: Belonging and Nationness in the Dutch Caribbean. *European Review of Latin American and Caribbean Studies / Revista Europea de Estudios Latinoamericanos y del Caribe* (98): 87–98.

Van Reekum R and Van den Berg M (2015) Performing Dialogical Dutchness: Negotiating a National Imaginary in Parenting Guidance. *Nations and Nationalism* 21(4): 741–760.

Vetik R (2011) Citizenship, statelessness and belonging in Estonia. Paper presented at the ECPR General Conference, Reykjavik.

Wekker G (2016) *White Innocence*. Durham, NC: Duke University Press.

Yanow D (2003) *Constructing 'Race' and 'Ethnicity' in America: Category-Making in Public Policy and Administration*. Abingdon and New York: Routledge.

Yanow D and van der Haar M (2013) People Out of Place: Allochthony and Autochthony in the Netherlands' Identity Discourse – Metaphors and Categories in Action. *Journal of International Relations and Development* 16(2): 227–261.

Yuval-Davis N (2007) *Gender & Nation*. London and New Delhi: Sage.

7

Data Subjects: Calibrating and Sieving

Baki Cakici and Evelyn Ruppert

Introduction[1]

The problem, however, is to get the respondent to answer these questions.[2]

Who are the subjects of data practices? How do data practices configure the capacities of subjects to engage and participate in their categorisation and become part of a population? These are questions this chapter turns to by first assuming that subjects do not pre-exist data practices but come into being through them (Ruppert, 2011). The data practices analysed in the foregoing chapters, such as those that make up administrative registers and surveys, involve different relations to what this chapter refers to as data subjects. Whether implicit or explicit, data practices that encode people into categories, for example, interact and engage with subjects in different ways. And, in doing so, data subjects come into being through varying relations, interactions, and dynamics between human and technological actors that make up data practices. This is distinct from usual understandings of data subjects, who are typically conceived as people who have a passive entitlement to their personal data and privacy, a right that is regulated by the state (Guild, 2019: 268). Similarly, it is different from an understanding

that conceives of data subjects as 'data doubles' (Haggerty and Ericson, 2000), which implies that data are simply digital duplicates rather than the products of subjects' relations with digital technologies.[3] Rather, this chapter explores how data subjects neither pre-exist nor are passive but shaped through data practices that configure their capacities to intervene, challenge, and influence how they are then categorised and become part of a population. Such configurations and capacities are variable and contingent because of different sociotechnical relations and data practices that make up methods; that is, data subjects are multiple, a point we demonstrate below through the analysis of two distinct data practices: calibrating and sieving.

A key aspect of the configuration of capacities that bring different data subjects into being concerns how data practices are organised and influenced by problematisations. For instance, as expressed in the opening quote, getting subjects to answer is a problem that is said to be evident in a general decline in response rates to paper questionnaires. This decline is usually attributed to people being overburdened by numerous state data collection activities or their concerns about privacy and confidentiality. Another cited cause explored in Chapter 4 is that certain groups, such as refugees and homeless people are identified as difficult to locate, contact, interview, and persuade to participate in data collection methods and thus 'hard-to-count'. However, even when subjects answer questionnaires, their responses can be a further source of critique. While expected to reveal themselves truthfully, subjects are also understood, in some cases, to answer strategically and subversively, for example, by claiming unrecognised or unauthoritative categories.[4] 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.[5]

One such example captured media attention in the UK in the wake of the 2001 census of England and Wales when more than 390,000 respondents declared 'Jedi' as their religion in response to a newly introduced optional question on religious beliefs. While the UK Office for National Statistics (ONS) categorised what was considered a subversive response under the 'no religion' category, the response was referenced in subsequent parliamentary debates on the future of population censuses and the inaccuracy of questionnaire-based methods. These are just a few of the problematisations of subjects whose self-elicited answers can also be influenced by how questions are worded or whether questions are self-completed or involve an enumerator.[6]

Such problematisations of data subjects come to inform and configure data practices that make up method experiments that engage with digital technologies as possible solutions. While also driven by problematisations of data quality, cost, and timeliness, it is how method experiments are offered as solutions to the (non)responsiveness and truthfulness of subjects that this chapter considers. That is, such problematisations of subjects' very capacity to act and influence (or subvert) how they are categorised, we argue, inform the development of data practices that are offered as solutions. We interpret two

such solutions – data practices that calibrate and sieve – and argue that they constitute different 'forces of subjectivation' (Cakici and Ruppert, 2019).

In brief, our conception builds on Foucault's (1982) formulation that subjects are capable of reflection, self-formation, and are engaged in struggles against direct domination that involves a tension between governing and technologies of the self. It is a power relationship that requires that a person is capable of acting and who, when faced with a relationship of power, engages with 'a whole field of responses, reactions, results, and possible inventions' (Foucault, 1982: 789). In this way, Foucault connected subjection and subjectivation to capture that power is not possessed but is a relation and process (Cremonesi et al., 2016). 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: 789).

The tension within a relationship of power is captured in a distinction suggested by Balibar (1991) between being a subject *to* power and a subject *of* power. 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. Regarding the latter possibility, this is what distinguishes being a citizen: one who is both a subject to and subject

of power, where obedience, submission, and subversion are always-present potentialities (Isin and Ruppert, 2015).

It is in this sense that the data practices that make up method experiments can be conceived of as forces of subjectivation. They are forces of power not in the sense that they determine but rather, through the different sociotechnical relations that make them up, differently configure the capacities of subjects to act in how they are categorised and become part of a population. For the data practices that make up 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 or the writing of a tweet – who participate in their subjectivation and categorisation. They can act in obedient and submissive ways and simply respond as expected and required or they can invent, subvert, and resist their subjectivation and perform as citizens including not participating or submitting to the data demands of governing authorities (Isin and Ruppert, 2015). As such, changes in data practices reconfigure the possibilities and potentials of acting and performing as citizens.

It is regarding this potential that method experiments can also be inventive of new forms of acting when they come into play and can also change initial problem formulations. For when put into action, the interactions and dynamics between human and technological actors are not determining but contingent. 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, problems, solutions, and interventions are entangled and dynamically reformulated.

This conception of forces of subjectivation is taken up in this chapter to analyse two method experiments. They are considered as experiments insofar as they involve pilot

projects and the testing of innovations in methods that need to be proven not through argumentation but demonstration (Ruppert and Scheel, 2019). The analyses interpret the data practices that make up method experiments as sociotechnical and contingent in relation to how they configure, enable, or constrain how subjects might act. 'Calibrating responses' examines some of the data practices involved in digital censuses and how they seek to maximise and guide the responses of subjects. 'Sieving tweets' focuses on data practices involved in experiments with Twitter for generating 'live' data about the dynamics of student internal migration. In both cases, we examine how classifying and encoding subjects, as defined by Desrosières (1998), involve different forces of subjectivation that seek to maximise the obedience and submission of subjects. The conclusion reflects on these forces to consider the consequences of data practices for the possibilities of subjects to act as 'data citizens' (Ruppert, 2018) in how they are categorised and encoded as part of a population.

Calibrating Responses

In 2011, following years of design and development, Estonia tested and conducted its first e-census. Reporting on the outcomes, Estonian statisticians declared that the country 'reached international premiere league' in that 'all people could fill out their personal questionnaire online' with the result that the country 'set the world record' with 66 per cent of respondents using the e-census (Tiit, 2013, 2015). This evaluation of success reflects the relation of the e-census to similar NSI method experiments with digital, online, or e-censuses (generally referred to as digital censuses) over the past decade. As one solution to the problems of paper questionnaires, these experiments are at various stages of design and implementation and

circulate in reports, international presentations, and demonstrations within and beyond EU NSIs. Rather than inventions of individual NSIs, problems and solutions are being identified, developed, repeated, referenced, debated, and contested and travel and circulate in and through the transnational field of statistics (Scheel, Grommé, and Ruppert, 2016). As such, the field includes states that make up the EU as well as those that form part of the UNECE. The examples analysed here are understood to be part of this field and through which national statisticians introduce, demonstrate, and defend new data practices as well as compete to set 'world records.'

Returning to the report on the Estonian e-census, statisticians noted that achieving a high online response rate involved an 'information and motivation campaign' that explained how a tachometer would track the volume of active respondents completing the census. One report described how the use of online enumeration rose to unexpected levels, despite the tachometer warning that the platform was experiencing a high volume of activity and that respondents might best do their submission later. Because of high volumes, the time required for responding was doubled, which further exacerbated online congestion. Customer support was subsequently unable to answer all incoming questions and internet services were interrupted at one point for about half an hour. Measures were taken to improve the situation on the following day and no further major technical setbacks were experienced. After this intense start-up, when approximately 50,000 people completed the online questionnaire in one day, levels dropped to 20,000 over the final two weeks (Statistics Estonia, 2012).

This account highlights some valuations and considerations related to NSI method experiments with digital censuses, which are more generally positioned as part of a broader move to 'digital government.' For example, Estonian statisticians

described the e-census as 'essentially, a grand IT project' (Statistics Estonia, 2012) that is part of what the government refers to as e-Estonia:

> Estonian people are used to thinking that Estonia is an e-country. We have an e-state and a wide range of e-services. Sometimes we worry whether other countries are overtaking us in the e-race. It is, of course, difficult to measure a country's e-capability, as there are no uniform indicators in this area. However, the census reinforced the notion of e-Estonia, which is positive. Not only because we are proud to be e-Estonia, but also because the active participation in the e-census will probably help us to conduct the next census with lower costs and greater efficiency (Oopkaup and Servinski 2013, 17).

Reflecting on the case of e-Estonia, a UK report described this as transforming government through technology and 'the relationship between citizens and the State – putting more power in the hands of citizens and being more responsive to their needs' (UK, 2017: 21). While oriented to numerous objectives, such as lower cost and efficiency, accounts of digital government, and more specifically of the Estonian e-census, proclaim the possibilities of digital technologies to establish a new relation between subjects and the state. However, the data practices that make up digital censuses configure this relation in particular ways that enable, constrain, and configure the forces of subjectivation and how subjects are categorised and become part of a population. Rather than simply tools, technologies such as the live tracking of responses and tachometers are part of an array of sociotechnical actors that make up these forces.

Such an array of forces is exemplified in Australia's design of a digital census. According to a statistician in a presentation made at a UK international conference in 2014, rather than an

online census, a digital census does not simply use digital technologies such as the internet to collect data and disseminate results.[7] It means to do all aspects of the census digitally. Their presentation reflected on the Australian Bureau of Statistics' (ABS) plans for its first 'digital census' in 2016, which they said would involve a 'transformation' rather than simply 'translation' of a paper questionnaire into digital format. It would involve a move from digital publishing to digital transacting and interacting with subjects at all stages of enumeration, and a responsive approach that would make data collection adjustments in near 'real-time' based on field intelligence and response rates (Australian Bureau of Statistics, 2015). A central management centre would achieve this by digitally monitoring a range of management information, including online response rates, paper form requests and returns, and social media. For example, when the response rate of an area lagged others, then a variation to the enumeration approach would be designed, reviewed, and actioned.

The statistician's presentation conveyed how relations between a digital census and subjects are understood. They are relations that can be interpreted as involving entangled human and technological relations that emerge through a dynamic call-and-response between subjects and technologies. While no method can direct subjects to one and only one way of acting, the data practices that make up the digital census are arranged to manage and guide how subjects act. In other words, they anticipate how a subject might act and identify, and seek to manage, direct, and channel those possibilities. It is in this way that a digital census anticipates subjects. As other researchers have elaborated, anticipatory logics underpin both governing and technical practices and are speculative regimes and forces (Adams, Murphy, and Clarke, 2009; Ratner, 2019). Anticipatory and pre-emptive logics, for example, have

been explored in relation to security and surveillance (Aradau and Blanke 2018). However, these studies address anticipatory logics involved in the analysis of data rather than the practices that configure relations to subjects. As developed below, the data practices that make up digital censuses anticipate how subjects might act and do so dynamically through what we describe as calibration.

For example, the ABS statistician, in their presentation to the UK international conference, described how putting a questionnaire online does not merely change the relation to subjects but transforms it into an interaction that is 'easy, responsive, fun.' The proposed design would do this by providing more information through pop-up windows to guide correct responses; drag and drop techniques to facilitate the ease of completing questions; assistance prompts to guide experience such as supplementary questions; and images and summary compilations that visualise responses so that they can be verified by subjects. The Estonia e-census also included help texts and 'soft and strict logical controls' to 'prevent or highlight the majority of logically impossible responses' (Statistics Estonia 2012, 3).

For the Australian digital census, the management of relations also extended to a 'field force' of workers who would use digital technologies to better capture and monitor subjects. By digitally monitoring progress through handheld devices, constant feedback on operational progress and instructions would be fed back to workers to optimise their activity and highlight problem areas in response rates. Social media platforms such as Twitter would also be used by workers to communicate experiences to each other so that problematic subjects and areas could be better targeted. Similarly, Estonia's e-census included 'The Survey Fieldwork Information System (VVIS)', which created work lists for enumeration areas, managed the roles of census team members, and monitored interviews amongst other things (Statistics Estonia 2012, 3).

All of these features were implemented in Australia's 2016 digital census and Estonia's 2011 e-census. Through numerous data practices, subjectivation was transformed into an interactive and live process of calibrating the responses of subjects by prompting and guiding them and making the process fun and easy and thereby maximise their submission to the census. 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, in contrast to paper questionnaires, this was conceived of as happening in 'real time', rather than through long processes of testing, piloting, and field worker feedback. With digital censuses then, relations between digital technologies, central management, and field workers that make up the method are organised by data practices that are dynamic, recursive, and responsive.

At the same time, the humans and technologies that participate in digital censuses extend to multiple other data practices such as those comprising administrative registers, self-completed paper questionnaires, and interviews conducted by enumerators using digital questionnaires on laptops. For example, in Estonia, registers were used in various ways such as to pre-fill some answers on questionnaires and supplement results when data was missing (Statistics Estonia, 2012).[8] In these ways, digital censuses are part of broader method assemblages that consist of data practices involving numerous technologies, rules, things, concepts, and people.

Producing New Problematic Subjects

At the 2015 annual meeting of the UNECE Group of Experts on Population and Housing Censuses, a statistician from the UK ONS noted that his office had learned much from international colleagues and their census practices. He noted that

international practices had influenced the UK's decision to introduce a major change in what he referred to as the '2021 Census Transformation Programme': that censuses would be conducted 'online first' and supplemented by multimode follow-up methods to capture non-responding households.[9] The statistician noted that the online census would also go beyond the simple translation of a paper questionnaire to incorporate many of the elements adopted in the Australian digital census 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 to fit smaller screens so that subjects could respond using handheld devices (ONS, 2015a).

Over time, this initial conception of the ONS Census Programme lead to the design of an online census that was promoted as a 'digital-first approach' and which would be 'easy to complete, and rewarding for respondents, so 70% provide data without follow-up' such that '75% of responses [are] provided online, and assistance provided to those who need it, to make this the most inclusive census ever' (HM Government 2018, 3). It would adopt smart type-in options and 'search-as-you-type' capabilities and functions such as routing, validation, and guidance. Additionally, through multi-channel and multi-lingual communications, community engagement, and the advice and help of field force and contact centre staff, the design would 'ensure people can tell us how they wish to identify themselves' (10). These and other sociotechnical arrangements would make up the many 'interactions with the census respondent'.

The validation and smart type-in features of digital censuses referred to above are made possible by the generation of paradata, which is a type of metadata.[10] Rather than being descriptive of the practices through which data has been

generated as in traditional metadata, paradata constitutes 'process' data on a subject's digital actions.[11] It is sometimes referred to as big data because it is generated in 'real-time', and in large volumes that require processing by algorithms. It includes data on devices being used; timestamps; 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 because of one of these design elements. In these ways, paradata involves tracking the relation between the digital census and the subject through metrics about data collection and are part of a 'data driven approach', which informs strategies for increasing response rates and the submission of subjects. It is a by-product of digital technologies that can be put in the service of better calibrating responses.

Using 'smart' technologies such as autocomplete, the data practices of digital censuses thus operate like commercial digital platforms. Indeed, one justification for digital censuses is that subjects regularly engage with digital platforms for both public and commercial purposes and thus have the familiarity and skills necessary. At the same time, digital censuses adopt many of the elements of the user interfaces that make up these other platforms – especially those of Google, Facebook and Amazon – and which are criticised for channelling choices and directing queries (König and Rasch, 2014; van Dijck, Poell, and De Waal, 2018). While user interfaces such as Google's query function appear neutral, autocomplete suggestions anticipate and predict what users want to know and direct 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. In these ways, digital censuses

incorporate practices innovated and designed by private technology companies. As such they also adopt similar logics, especially those advanced by data science, which seek to tame, control, and guide the actions of subjects through a new science of societies that challenges existing forms of data and knowledge such as that generated by traditional methods and practices of national statisticians (Grommé, Ruppert, and Cakici, 2018).

While all data practices variously channel and direct answers of subjects through techniques such as tick boxes on questionnaires, digital technologies do this in ways that are less evident and work in the background to increase submission by reducing the possibilities of intervening and subverting. Like internet platforms that espouse process data as working in the service of a better and faster customer service, so too is paradata mobilised in the service of better and faster responses to digital censuses. Through both the identification and subsequent capture of evasive, hard-to-count subjects, calibrating aims to normalise them through techniques that entice responses through fun elements and gamification and that discipline by anticipating and preventing illogical or unrecognised responses. In this way forces of subjectivation configure capacities and possibilities for acting.

However, while an online census was promoted by ONS for its capacity to ensure correct responses from subjects, it also produced new problematic subjects. Four groups of problematic subjects were anticipated based on their expected access to and/or willingness to use the internet to digitally engage with government via the internet (Figure 7.1). Problematic subjects – like hard-to-count subjects discussed in Chapter 4 – were differentiated according to several criteria. For each group, their related sociodemographic characteristics were identified (age, location, etc.) as well as reasons for

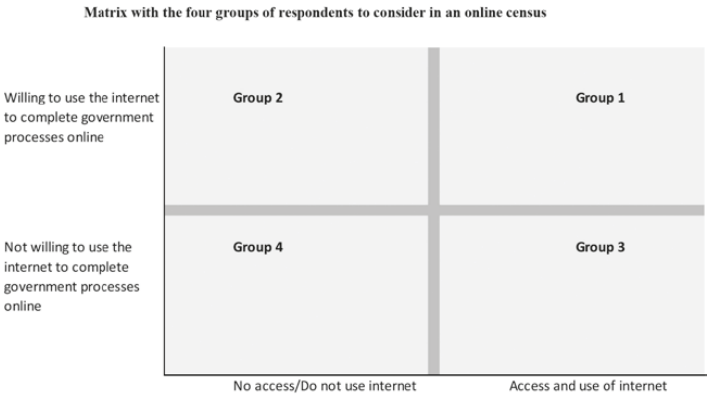

Figure 7.1 Categorisation of Respondents[a]
[a]Source: ONS, 2015a

being unwilling to digitally engage (lack of trust, internet security, etc.). In this conception, a digital divide was conceived not simply between who does or does not have access to the internet, but as divisions that occur 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 when targets were not being met so as to increase the number of subjects who submit to the census.

Such management involved offline modes as demonstrated in the ONS's test of its online census in 2017. The test 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 completing an online questionnaire. The decision on the design of the 2021 census included this service, which involved 'trusted suppliers who have the staff, premises and technology' to help respondents as well as the organisation of 'completion events' to stimulate response rates (HM Government 2018, 5).

Subjectivation thus involves data practices that anticipate how subjects might act and then calibrate how they do act through the ongoing process of digital management and directing. That is, a 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. However, as illustrated above, the design of a digital census is generative of new problematic subjects and calls forth management solutions in the form of new actors (librarians, enumerators), sites (libraries and computer terminals), and data (paradata), which all participate in subjectivation. All of these participate in the forces of subjectivation and inventive of data subjects who do not pre-exist but come into being through data practices that configure the relations, interactions, and dynamics between human and technological actors.

Yet, management is not only necessary to direct subjects, but also to address the instability and vulnerabilities of digital technologies. While this can take many forms, such as a change in an operating system as noted in the next section, 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 led to a major inquiry into cybersecurity and the

close partnership between ABS and IBM.[12] Loss of public trust and confidence were widely noted as a major consequence but what the incident points to are the contingencies of digital technologies. Not only are they subject to operational failures, but other forms of subversion because of the introduction of new technological and human actors that reconfigure those possibilities. Additionally, such contingencies reduce the submission of subjects to the digital census and, in turn, desired response rates.

In response to a recommendation in that report, ABS established an Independent Assurance Panel (the Panel) to secure trust in census operations and the quality of data generated. Rather than an assessment of individual features of the digital census, their assessment was that the 2016 census produced data of comparable quality to previous censuses and 'is useful and useable' (Census Independent Assurance Panel to the Australian Statistician 2017, iii). The relevance of the digital mattered only in relation to the DDoS rather than all the other proclaimed benefits and operational features detailed above. While internal reviews may well focus on this, the public response concerned the security of the digital census and confidence in its quality, and the degree to which subjects submit to and act in ways anticipated. As we explore in the next section on method experiments with Twitter data, the dynamics of sociotechnical relations, and the contingencies of data practices that configure subjectivation, can lead to other unexpected outcomes.

Sieving tweets

In this section, we explore the dynamics of subjectivation in relation to one experiment, an ONS pilot project that sought to use Twitter data to investigate how populations move within

the UK (ONS, 2015b). In 2014 and 2015, ONS statisticians experimented with a method to identify patterns in when and where users create Twitter posts based on aggregated data collected from publicly available Twitter profiles.[13] Their driving 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 move from one place to another. They argued that this would be an improvement over subjects declaring their mobility patterns on questionnaires as it would avoid false reporting and underreporting (i.e., where respondents either provide a wrong address, or provide only one address when they occupy several). The statisticians believed that it could also provide more timely statistics about how people move between addresses throughout the year.

This section explores how this method experiment involved sieving as a data practice and force of subjectivation. Like the previous example, the experiment was offered as a potential solution to problematised subjects, in this instance that of higher education students. They are deemed hard-to-count because of their irregular movements between universities and multiple residences within the academic year, which makes it difficult to encode them to a usual residence (on the problematisation of mobile people as 'hard-to-count' see Chapter 3). As elaborated below, sieving involves repurposing tweets to filter and sort subjects and then infer and enact the category of migrating students. In distinction to calibrating, which iteratively incites, disciplines, and interacts with subjects to participate in their categorisation, sieving is a force of subjectivation that does not engage with subjects but categorises them based on repurposing big data about their conduct. That is, rather than guiding subjects, sieving eliminates the possibilities of subjects to act in – or even know – how they were categorised and the possibilities of their intervention.

The experiment more generally held the promise of providing more timely statistics that reflect lived experiences and which do not rely on elicited (and unreliable) responses from subjects. By repurposing the data traces of Twitter users, the pilot followed method 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 (Liu and Ruths, 2013; Mislove et al., 2011; 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 engender new problematisations.

The method experiment involved several stages beginning with investigation of the free-text location field included in Twitter profiles. After a brief study, the statisticians in charge concluded that the text field is an unreliable data source as users seemed to use it in different ways, sometimes leaving it blank, and sometimes subverting the intended use by filling it with fictional places. The free-text field provided the potential for subjects to act in ways that subverted and were 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 include GPS coordinates as these messages, also known as geolocated tweets, provide standardised data about the location from

which a tweet was posted. These were much easier to analyse using existing statistical methods, and less prone to the kinds of uncertainties introduced by users. However, they made up a fraction of the total number of tweets, and many were posted by the same users. Furthermore, 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.

To eliminate tweets that did not include GPS coordinates and thereby focus on a desired subset, the statisticians engaged in the data practice of sieving. Kockelman's (2013) conceptualisation of sieving in algorithmic devices shows how sieves have desires built into them; they retain a set of 'desirable' elements while allowing the 'undesirable' to disperse. This process was evident in the separation of tweets depending on the availability of the GPS coordinates, where the geolocated tweets – constituting a smaller volume – were gathered for further analysis and the rest were discarded. Such procedures were repeated with different sieves, for example one that allowed the removal of Twitter bots (accounts that post exceptionally high numbers of tweets in relation to the rest of users). Another sieve was necessary when the statisticians discovered that two sets of data they used, one purchased from a data reseller and another obtained using the Twitter API,[14] included duplicates because there was an overlap in the dates when the data were collected. While the work of sieving involved separating tweets in both cases, its significance was that it transformed undifferentiated collections into a potential source of data for inferring categories of subjects using existing statistical methods. In so doing, rather than engaging the desires of subjects in categorisation, sieving materialised categories that reflected the preferences and desires of statisticians for reliable and verifiable geolocations.

Although tweets in a chosen subset could now 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 the translation, the statisticians used a clustering algorithm called DBScan, which arranged the stream of tweets for each user into clusters of nearby data points. Next, they used a set of rules about the time of day and frequency of posts to infer whether the assigned locations could be considered the home or the workplace of the posting user (see ONS (2012) for a detailed description of the method). Finally, they compared the positions of the tweet clusters to the borders of local authorities, and they flagged those that appeared in different local authorities from one month to the next as instances of internal migration. Using this analysis, the statisticians quickly detected a 'strong signal' coinciding with the cycle of the academic year. The signal indicated that in local authorities with high proportions of students, the volume of tweets seemed to 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, movements that were not possible to detect using any existing data sources.

The production of dominant tweet clusters is another example of sieving in action. The algorithm (DBScan) converts a larger set of tweets into a much smaller one by allowing closely located tweets to pass and be included while blocking and discarding more dispersed ones. Which tweets are allowed to pass or are discarded are determined individually for each Twitter user, that is, a different sieve is used for each user, but the tweets themselves, and the location data they contain, remain unchanged throughout the process. In other words, the algorithm performs as a sieve by neither changing which tweets it catches, nor which ones it lets through.

While the data practice of sieving tweets led to inferring and in turn enacting the category of migrating students by repurposing existing Twitter user data, it also eliminated the possibilities of subjects to act in – or even know – how they were categorised and the possibilities of their intervention. To demonstrate the effect in action, we can consider the final inference that enacted the student migrant population. As noted previously, higher education students are often problematised subjects because their movements between universities and other residences within the academic year make it difficult to encode them in a usual residence (see discussion in Chapter 3). For example, statisticians have long argued that population counts conducted at different times in the same geographic area can display high variations if the size of the student population is sufficiently large (Duke-Williams, 2009; Mitchell et al., 2002). It is in the context of this problematisation that the statisticians on the pilot project came to recognise and identify a solution: by converting Twitter posts into geographic indicators the mobility of Twitter users could be inferred. That is, it was in relation to a well-known and debated problem that the pilot project invented a solution which could be legible and recognised as useful to produce statistics. It did so through the further stabilisation of the notion of student mobility, where studying involved living away from home while remaining connected to a home that exists in another location. The role of sieving as a data practice in this configuration is that it generated a potential solution to a problem by inferring and enacting the category of the migrating student.

Detecting and inferring student migration was a promising result for the pilot project as it solved the problem of categorising a hard-to-count mobile student population. However, the statistician in charge of the project noticed a significant decrease in the number of data points at a particular date in

the one-year sample of Twitter posts. After a period of investigation, they found out that the date of this decrease coincided with the release date of iOS 8 (an operating system used by Apple devices). Further investigation pointed to a change in the default settings in the operating system for location sharing, meaning that on that date many devices stopped reporting their locations, and thus disappeared from the dataset. This disappearance led the statistician to characterise the dataset as volatile, that is, unreliable and prone to sudden changes, and ultimately unsuitable as a data source for official statistics. In other words, problematic subjects were replaced by problematic, unreliable, and volatile technological actors.

While complications that arose when using GPS data for population data were easier to anticipate and handle for the statisticians, the GPS coordinates were thus also linked to a much larger method assemblage, a hinterland of actors consisting of networks of satellites, sensors, and mobile devices, all of which generated a new set of unanticipated issues. In this instance, the data practices were contingent due to their dependency on this assemblage, where changes in software release schedules or operating system settings of Twitter users, could jeopardise the otherwise stable results.

When the chief statistician described the data source as 'volatile,' the description captured the contingencies of forces of subjectivation. In the pilot, using GPS coordinates to overcome the challenges of determining a location through free-text fields exposed other dependencies beyond the control of the project. At stake was the possibility of being able to anticipate technological actors; that is, even if the sharp decrease in user numbers could be tied to a single event this time, a similar change in the future might be impossible to anticipate, explain, or even to detect. Configuring subjectivation, in other words, was beyond the reach of their method as it was part of a widely

distributed assemblage of infrastructures and temporalities. In these ways, forces of subjectivation involve not only configuring, anticipating, and remediating the acts and actions of human subjects, but also those of technological actors.[15]

The Twitter pilot began as a method of a more 'live' tracking of mobility by sieving geolocated tweets to produce categories from clusters of data points made possible by a highly technical analysis. For us, it demonstrated how subjectivation is differently configured by data practices, 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 data practice chosen by statisticians inferred and enacted the category of migrating students, 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.

So, while the data practice of sieving was a solution to the problem of categorising migrating students, it was generative of a series of new problems. Subjects were problematised for their use of a free-text field, which generated unanticipated categories or interpretations. While GPS coordinates were identified as a solution, this made the method vulnerable to technical forces of operating systems involving actions beyond their control or knowledge. In these ways, while reconfigurations of forces of subjectivation may solve one set of problems, they can also be generative of new ones.

Conclusion

This chapter covered just a few examples of data practices that configure the capacities of subjects to engage and participate in their categorisation and how they become part of

a population. It highlighted that while cost, time, efficiency, and quality are key objectives of method experiments, they also are directed at reconfiguring how people are subjectified to meet desired ends through data practices that are not linear but recursive and dynamic. From the iterative calibrating of responses of digital censuses to the repetitive sieving of tweets, data practices work to minimise the subversive and maximise the submissive actions of subjects.

This objective was exemplified in problematisations of subversive or hard-to-count subjects such as those who answered Jedi in response to the 'no religion' question of the 2011 census of England and Wales. While a digital census was offered as a possible solution, by reconfiguring the forces of subjectivation, new hard-to-count subjects were anticipated and produced due to the introduction of digital technologies. In this regard, solutions are inventive of new possibilities for subjects to act, be excluded, or problematised. This is in part because data practices such as calibrating and sieving introduce new actors, such as the assumptions, 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 data practices differently configure their subjectivation, which can be anticipated and guided but not settled in advance.

Yet, there is another consequence. 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, the method experiments we have analysed 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, some 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.

Data from digital platforms and mobile devices are also potentially transformative of the how European population statistics may be produced in the future. Method experiments such as those with Twitter – or mobile phones (e.g., see Ruppert and Scheel, 2019) – introduce big data that are transnational in their generation and ownership. Given that European population statistics are largely generated by and reliant upon national statistical institutes, big data introduce the prospect of transcending national borders to produce European level statistics. That is, rather than harmonising and assembling national data, European statistics could be based on transnational data. Since this data is owned by multinational corporations, European level governance and negotiation may be necessary to secure access if experiments are to lead to the production of internationally comparable population statistics.[16] Furthermore, if, as proposed in Chapter 1, statistics help to constitute what is the population and who are the people of Europe, then big data could be a key political technology through which the EU could possibly constitute its public and secure its legitimacy. It may offer the possibility of transcending national categories such as usual residence by capturing transnational and mobile modes of living (see Chapter 3). However, and in line with the conception developed in this book, data practices are part of a transnational field of statistics where scales of

the local, the national, and the international overlap and intersect and involve complex relations of power and influence such that what they enact are neither 'national' nor 'European' statistics. This is a point which we return to in Chapter 9.

This reflection is critical as digital technologies become ever more part of social life and at the same time part of new data practices for knowing and governing. What we have focused on in this chapter is what this may mean for relations between subjects and the making of population statistics, which are by no means given or settled. Of critical importance 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 democratic 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 possibilities for such democratic interventions and claims are significant, if, as we have argued, being a citizen is to be both a subject to and subject of power, where obedience, submission, and subversion are always-present potentialities. In relation to official statistics, it means to consider subjects as 'data citizens' with the right to shape how data is made about them and the societies of which they are a part, an issue which we return to in the concluding chapter (Ruppert, 2019). That is, the possibilities and potentials of citizens to act in their subjectivation are as important, if not more, than the promises of digital technologies for more timely, efficient, cheaper, and reliable statistics.

References

Adams V, Murphy M, and Clarke AE (2009) Anticipation: Technoscience, Life, Affect, Temporality. *Subjectivity* 28(1): 246–265.

Aradau C and Blanke T (2018) Governing Others: Anomaly and the Algorithmic Subject of Security. *European Journal of International Security* 3(1): 1–21.

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: UNECE.

Balibar E (1991) Citizen Subject. In: Cadava E, Connor P, and Nancy J-L (eds) *Who Comes After the Subject?* London: Routledge, pp. 33–57.

Bexley S (2017) It's All About Inclusion: How ONS Plans to Support the Digital Have-nots. Available at: https://blog.ons.gov.uk (accessed 23 March 2018).

Bruns A and Burgess J (2015) Easy Data, Hard Data: The Politics and Pragmatics of Twitter Research After the Computational Turn. In: Langlois G, Redden J, and Elmer G (eds) *Compromised Data: From Social Media to Big Data*. New York: Bloomsbury Academic, pp. 93–111.

Cakici B and Ruppert E (2019) Methods as Forces of Subjectivation: Experiments in the Remaking of Official Statistics. *Journal of Cultural Economy*: 1–15.

Census Independent Assurance Panel to the Australian Statistician (2017) *Report on the Quality of 2016 Census Data*. Australia: ABS.

Couper MP and Singer E (2013) Informed Consent for Web Paradata Use. *Survey Research Methods* 7(1): 57–67.

Cremonesi L, Irrera O, Lorenzini D, et al. (eds) (2016) *Foucault and the Making of Subjects; Rethinking Autonomy between Subjection and Subjectivation*. London: Rowman & Littlefield International.

Desrosières A (1998) *The Politics of Large Numbers: A History of Statistical Reasoning* (ed. RD Whitley). Cambridge, MA and London: Harvard University Press.

Duke-Williams O (2009) The Geographies of Student Migration in the UK. *Environment and Planning A: Economy and Space* 41(8): 1826–1848.

Foucault M (1982) The Subject and Power. *Critical Inquiry* 8(4): 777–795.

Grommé F, Ruppert E and Cakici B (2018) Data Scientists: A New Faction of the Transnational Field of Statistics. In: Knox H and Nafus D (eds) *Ethnography for a Data Saturated World*. Manchester: Manchester University Press, pp. 33–61.

Guild E (2019) Data Rights: Claiming Privacy Rights Through International Institutions. In: Bigo D, Isin E, and Ruppert E (eds) *Data Politics: Worlds, Subjects, Rights*. Abingdon and New York: Routledge, pp. 266–283.

Haggerty KD and Ericson RV (2000) The Surveillant Assemblage. *British Journal of Sociology* 51(4): 605–622.

HM Government (2018) *Help Shape Our Future: The 2021 Census of Population and Housing in England and Wales*. White Paper. UK Parliament.

Isin E and Ruppert E (2015) *Being Digital Citizens*. London: Rowman & Littlefield International.

Kockelman P (2013) The Anthropology of an Equation: Sieves, Spam Filters, Agentive Algorithms, and Ontologies of Transformation. *HAU: Journal of Ethnographic Theory* 3(3): 33–61.

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, pp. 1–14.

Liu W and Ruths D (2013) What's in a Name? Using First Names as Features for Gender Inference in Twitter. In: *AAAI Spring Symposium: Analyzing Microtext* (conference paper), pp. 10–16.

MacGibbon A (2016) *Review of the Events Surrounding the 2016 eCensus:* Australian Government, Department of the Prime Minister and Cabinet.

Marres N (2017) *Digital Sociology: The Reinvention of Social Research*. London: Wiley.

Mislove A, Lehmann S, Ahn Y-Y, et al. (2011) Understanding the Demographics of Twitter Users. In: *Fifth International AAAI Conference on Weblogs and Social Media* (conference paper).

Mitchell R, Dorling D, Martin D, 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.

Mocanu D, Baronchelli A, Perra N, et al. (2013) The Twitter of Babel: Mapping World Languages through Microblogging Platforms. *PLOS ONE* 8(4): e61981.

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, Gravel R, Trieschnigg RB, et al. (2013) 'How Old Do You Think I Am?' A Study of Language and Age in Twitter. In: *Seventh International AAAI Conference on Weblogs and Social Media* (conference paper).

ONS (2012) 2011 Census: Key Statistics for local authorities in England and Wales. Available at: http://bit.ly/2ncI70x (accessed 28 March 2017).

ONS (2015a) Research for 2021 Census England and Wales: possible innovations under consideration. Paper 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 2015.

ONS (2015b) *Using Geolocated Twitter Traces to Infer Residence and Mobility*. Available at: http://bit.ly/2mLVFof (accessed 28 March 2017).

Oopkaup A and Servinski M (2013) A Positive View of Demographic Trends. *Eesti Statistika Kvartalikiri [Quarterly Bulletin of Statistics Estonia]* 2013(3): 15–18.

Ratner H (2019) Europeanizing the Danish School through National Testing: Standardized Assessment Scales and the Anticipation of Risky Populations. *Science, Technology, & Human Values*: 1–23.

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.

Ruppert E (2018) Sociotechnical Imaginaries of Different Data Futures: An Experiment in Citizen Data. *3e Van Doornlezing*, Rotterdam, NL, 14 June. Erasmus School of Behavioural and Social Sciences, Erasmus University Rotterdam.

Ruppert E (2019) Different Data Futures: An Experiment in Citizen Data. *Statistical Journal of the International Association for Official Statistics* 35: 633–641.

Ruppert E and Scheel S (2019) The Politics of Method: Taming the New, Making Data Official. *International Political Sociology* 13(3): 233–252.

Scheel S, Grommé F, Ruppert E, et al. (2016) *Transcending Methodological Nationalism through Transversal Methods? On the Stakes and Challenges of Collaboration*. ARITHMUS Working Paper 1. Available at www.arithmus.eu (accessed 12 May 2018).

Scheel S, Ruppert E, and Ustek-Spilda F (2019) Introduction: Enacting Migration Through Data Practices. *Environment and Planning D: Society and Space* 37(4): 579–588.

Sloan L, Morgan J, Burnap P, et al. (2015) Who Tweets? Deriving the Demographic Characteristics of Age, Occupation and Social Class from Twitter User Meta-Data. *PLOS ONE* 10(3): e0115545.

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

Statistics Estonia (2012) *Summary of the Conduct of the Estonian Population and Housing Census (PHC2011)*. Tallinn: Statistics Estonia.

Tiit E-M (2013) Estonian Census 2011. *Papers on Anthropology* (22): 234–246.

Tiit E-M (2015) The First Census Without Enumerators. *EestiPäevaleht*. Available at: http://epl.delfi.ee/archive/article.php?id=72013177.

UK (2017) *Government Transformation Strategy*. Cabinet Office. Available at: https://bityl.co/6HCB (accessed 3 November 2019).

van Dijck J, Poell T and De Waal M (2018) *The Platform Society: Public Values in a Connective World*. Oxford: Oxford University Press.

8

Statistician Subjects: Differentiating and Defending

Francisca Grommé, Baki Cakici, and Ville Takala

Introduction[1]

Who are the professional subjects of data practices? How are their skills, capacities, mindsets, and ethical positions shaped in relation to data practices? While the previous chapter explored how data practices subjectify people and how they are categorised, here we turn to consider how the statistician subject is being shaped, and the profession of national statistician repositioned, through what we refer to as 'professionalising practices'. The chapter develops this understanding by returning to technological changes described in the Preface: how digital technologies such as the internet, mobile devices, and big data are both challenging traditional methods of producing official statistics while at the same time offering possibilities to innovate the production of statistical knowledge about populations. We consider how data scientists are leading the development of such methods, especially those that involve big data.

As this chapter sets out, it is through the valuing and performing of data practices that engage with big data and related analytical techniques (which we will herewith refer to simply as big data) that the statistician subject is being shaped and the profession of national statistician repositioned. As expressed by leading national statisticians at conferences and in policy

papers, to meet the challenges and realise possibilities of big data requires more than simply changing data practices. It also requires that statisticians develop skills and knowledge not typically deployed in the production of official statistics, for example, analytic techniques such as machine learning and predictive modelling. Yet, what they also acknowledge is that skills alone are insufficient. 'Cultural change' is also necessary as advocated by proponents of the uptake of big data. As one speaker at a 2015 international seminar about the future of official statistics noted: 'Machine learning is the future. Big data cannot be processed by hand. Therefore, the current culture is a liability.' This speaker specifically referred to routine data practices, where the production of statistics often requires professional human judgement. However, as they observed, large volumes of data require automated forms of data processing which involve less human intervention. In their view, this is one way that the current culture of official statistics is being challenged.

Generally, cultural change has come to refer to a broad, fuzzy set of organisational, practical, and other desired changes in the profession of statistician.[2] This includes their skills and ways of thinking and a direction of change that are in part modelled on various understandings of the 'entrepreneurial mentality' of data scientists working in the technology sector. For instance, in a report on the value of official statistics, a UNECE task force provided an inventory of corporate practices, which included those of Apple, Amazon, and Google. Even though the report notes that statisticians 'have considerable comparative advantages' (UNECE, 2018: 10) to meet the needs of an information age, it also states the following:

> But competing information providers [e.g., Google] have advantages, too. Sometimes, they will have resources available to them which dwarf those available to most NSOs [national statistical

organisations]. They may also have cultures which allow them to take up new technologies and methodologies more quickly than traditionally has been the case in the official statistics community. They may also have cultures, driven by commercial necessity, which make them more responsive to customer needs (UNECE, 2018: 10).

What the quote epitomises is that private sector data providers are emerging competitors of NSIs. Data scientists in the private sector have the expertise and skills to 'take up new technologies and methodologies' required, and which are valued in relation to those of national statisticians.

We approach these questions of skills, expertise, mentalities, and cultures required to take up big data by considering them as objects of valuation and struggle within what we have previously conceived of as the transnational field of statistics (Scheel et al., 2016). As we will elaborate below, it is through struggles that the faction of national statistician is competing with other professions over the relative valuation of cultural capital and habitus required to work with big data. Such competition is occurring mostly in relation to data science and its professional subject, the data scientist. Yet, what constitutes data science or a data scientist is not universally agreed nor stable[3] (as is the case for other scientific disciplines and professions). Contemporary definitions of data science and data scientists are closely associated with big data, a term that became mainstream around 2011. In addition, the relations between official statistics and data science are framed in different ways. Whereas many frame them as competitive, some statisticians publicly speak 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).

Professionalising practices are part of such struggles. In some situations, this involves recognising forms of cultural capital and cultivating a habitus aligned with conceptions of the faction of 'data scientist', while in others it involves defending the faction of national statistician;[4] both situations are the object of the analyses that follow. Our aim is twofold: first, to understand *how* such change is being pursued through professionalising practices. Considering insights from STS and related fields, we do this by analysing how skills, capacities, mindsets, and ethical positions are valued discursively through job interviews, but also performed through material-semiotic practices such as data camps. Our second aim is to consider how such professionalising practices of national statisticians involve a tension between entrepreneurial and public service skills and habitus.

In what follows, we first elaborate our conceptualisation of the transnational field of statistics. Next, we empirically examine the shaping of the statistician subject and the repositioning of the faction of national statistician by analysing three professionalising practices: recruitment job interviews in the UK; a brainstorming workshop and data camp modelled after hackathons at Statistics Netherlands (SN); and presentations by statisticians at Eurostat and UNECE conferences. In the conclusion we highlight how thinking about professionalising practices is important to understand how data practices do not simply involve struggles over methods of producing statistics. They also involve professional struggles over the skills and habitus that are valued and cultivated. That is, to advocate the valuing of a particular data practice also involves recognising the required skills and habitus to perform them and in turn the relative advantages of professionals who possess them. We then suggest that data practices are bound up with professionalising practices and need to be considered together

to investigate the politics of method and the production of official statistics.

Shaping the Professional Subject in Relation to Big Data

As noted, we conceive of professionalising practices as part of struggles over the legitimacy of methods and their related data practices in the production of official statistics.[5] Such struggles are situated in, and help to shape, what can be understood as a transnational field of statistics. It is through specific practices that actors from competing professions attempt to advance or defend their relative positions within a field (Bourdieu, 1989). The field of statistics comprises differently positioned professions such as statisticians, demographers, domain specialists, academics, policy makers, and other users of statistics (cf. Scheel et al., 2016). While statisticians have long occupied a dominant position, data scientists are an emerging faction challenging this dominance within the field. For data scientists, at stake is recognition of big data and related analytical methods as legitimate and authoritative and, in turn, the cultural and symbolic capital that this will confer. For statisticians, their stakes are to protect and advance their authority and position in relation to each other and this faction. Through this understanding of the field we conceive of these stakes as a politics of method, as it provides a way to analyse 'the emergence of new kinds of practices' (Bigo, 2011: 240–241). In brief, the transnational field of statistics involves struggles over data and methodological innovations and authority in the production of official statistics. We understand this as 'a messy, competitive context [in which] the roles of different kinds of intellectuals, technical experts and social groups are at stake' (Savage, 2010: 237).

This competition does not only involve claims about data and the positioning of the existing methods of official statistics vis-à-vis those of data science; it also involves establishing the statistician subject as a trustworthy and competent professional. Here, we consider how the statistician subject is shaped by their socialisation within the field of statistics. This dynamic works in two directions: professional subjects are both shaped by the field and come to shape it through their practices. The factions of national statistician and data scientist are distinguished from each other by the valuation and appropriation of certain forms of cultural capital over others. Cultural capital includes skills, but also the knowledge of what to value and what professional ethics to support. This positioning is not entirely a matter of conscious choice. Instead, orienting towards a position in a field 'functions below the level of consciousness and language and beyond the scrutiny or control of the will' (Bourdieu, 1984: 466). Statistician subjects thus take up positions both as a result of the valuations of some skills, normative inclinations, and dispositions above others (for instance, in textbooks or by authority figures), and come to also embody these. The particular combination of skills, habits, normative inclinations, and so on that subjects come to embody constitute what Bourdieu refers to as 'habitus' ('a system of dispositions').[6]

Two examples well illustrate how the introduction of new methods to know populations can affect the formation of the professional subject. In the field of statistics, Savage shows how a shift from the 'gentlemanly social scientist' to a professional with a 'technical orientation' took place in the 20th century through the invention of the sample survey (2010). He argues that for the sample survey to be recognised as legitimate, differently positioned actors within the field needed to be convinced of the trustworthiness and validity of interview. The statistical

technique of random sampling was advanced to support the reliability of data as it reduced selection bias and led to more representative statistics. The consolidation of the interview method in turn supported the rise of social science and technically oriented researchers as part of the state apparatus. The second example is from the field of border security. Bigo examines how data collection technologies were relevant in distinguishing different professional positions within the field.[7] For instance, database analysts formed their professional positions around the authority of 'smart technologies' (such as predictive software). Bigo argues that dispositions within the field were 'activated – or not, as the case may be – by the use of specific technologies, and [dispositions] determine the capacity to restrain the deployment of these technologies, to modulate them' (Bigo, 2014: 210). Consequently, professional dispositions are not determined but can be activated, shaped, and reinforced by technologies deployed to know targe populations.

STS studies have pointed out that professional subjects are shaped, and positions valued and inhabited, through not only discursive but also material-semiotic practices. In their study of the rise of the experimental technique during the English Restoration, Shapin and Schaffer (1985) show how the social technology of prescribing 'modesty' was a key attribute of the formation of the experimental scientist. As Haraway (1997) later pointed out, such technologies also positioned scientist subjects as essentially male. Following these and other studies (see, for instance, Latour, 1993), Ruppert and Scheel demonstrated how the dynamics that arise when new methods are introduced within the field of statistics cannot be reduced to discursive claims:

> material-semiotic practices like demonstrations that seek to legitimize innovations in methods and data as official. In this way, we

underscore that the politics of method are not reducible to a competition between human actors who can put forward the best argument in the most compelling manner. Rather, the politics of method requires a symmetrical analysis that accounts for how different kinds of digital devices are mobilized in struggles over methodological innovations in the production and legitimation of official statistics (Ruppert and Scheel, 2019: 3–4).

In this chapter we adopt this focus on the relevance of both discursive and material-semiotic practices. We suggest that professionalising practices 'make explicit' the cultural capital and habitus involved in the formation of professional subjects. As Muniesa and Linhardt explain, 'making explicit' involves 'the actualization of the virtual' and 'about expressing something, provoking it in variable, conflicting, unanticipated manners, putting it to the test of becoming an actual configuration, an actual event' (2011: 546). Making things explicit does not unfold without problems, hesitations, or tensions. Rather, sensibilities are made visible and can then be put up for consideration, debate, or negotiation whether they arise in job interviews, brainstorming workshops, or conference presentations.

In the empirical sections below, we examine how professionalising practices make explicit tensions and congruities between the cultivation of entrepreneurial and public service skills and habitus. Our analysis of an entrepreneurial habitus draws on two ethnographic studies. The first concerns Irani's study of professional designers in India (2019). Characterising the tech sector, Irani shows that an entrepreneurial disposition includes a strong belief in technological innovation as the prime locus for societal change (instead of, for instance, poverty alleviation policies). Further aspects of an entrepreneurial disposition are a sense of optimism and urgency to

accomplish innovation and a strong belief in collaboration as the key to solving complex issues and problems. In addition, it includes the practice of experimentalism, in the sense that work is not always aimed at producing immediate tangible results. This notion of experimentalism is not only embraced to support learning through trial and error, it is also embraced because it allows for suggesting or hinting at future potential and value:

> But it is not tangible productivity, but what anthropologist Kaushik Sunder Rajan characterises as the 'felt possibility of future productivity or profit' (2006, 18). They produce and respond to vision, hope, and hype as they pursue speculative capital and investment; they promise not only financial value but also social value and legitimation for socially responsible funders and investors (Friedner 2015) (Irani, 2019: 16).

Even though statisticians are less affected by technology hype cycles and the pressures of external investors, this aspect of an entrepreneurial disposition may be relevant as statisticians do need to attract internal and external support and funding for new ideas.

Mackenzie's (2013) study of the practices of data scientists connects this disposition to data practices that involve the use of machine learning. The adoption of predictive analytics in these practices, he demonstrates, is part of a habitus that embraces probabilistic outcomes, likelihoods, and the optimisation of models, rather than their verification (as in 'traditional' statistics). This logic of optimisation and prediction can be applied to problems in a wide range of social domains, which is a key feature of entrepreneurialism. Finally, Mackenzie shows that an entrepreneurial ethos is further internalised by data scientists through competitions and

hackathons that involve a rhetoric of addressing them as 'wonderful people' (2013, 394): a highly desirable group equipped with a unique combination of skills to address the social challenges of our times.

What these studies offer is that an entrepreneurial habitus can be at work in several connected ways: in the development of skills and sensitivities to identify potential 'social problems' (and thereby potential markets), as well as in the appreciation and internalisation of a particular set of methods and related sensitivities. We will explore how valuations of future potential are relevant for the shaping of the statistician subject but are in tension with those of public service, a tension that is made explicit in professional practices.

Recruiting Data Scientists

Looking for Data Scientists

The first professionalising practice that we explore is government recruitment interviews for data scientists in the UK. Through the analysis of job descriptions and interviews we consider how valuations of 'data scientists' and the professional skills necessary for working with big data are made explicit. Specifically, we highlight how data scientists are differentiated from national statisticians and valued in relation to their 'future potential'. At the same time, we show how national statisticians are differentiated and valued in relation to their public service skills and dispositions.

In 2015, a recruitment committee interviewed applicants for data scientist posts distributed across several government departments. The committee included a statistician of the NSI, and 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, was situated in a small room, and involved a question-and-answer exchange between the recruitment committee and the applicant. The applicants were expected to demonstrate how their previous experience and knowledge were compatible with the role of the data scientist. Meanwhile, the committee needed to reach a consensus on whether the applicants' responses fulfilled the requirements for becoming a data scientist. The applicants were also required to take a multiple-choice test in an adjacent room following their interviews, which included questions on basic statistics knowledge such as the definition of terms, probability calculations, and so on.

The job description document that advertised the position presented an ideal type of data scientist: someone with a collection of skills in programming, computing, data, and statistics. The interview committee was asked to formulate questions in relation to this description to assess the candidate's competency in different skills. They were also provided with a 'marking matrix', a document listing the categories and the grades they should use to assess the performance of the applicants during the interview. This matrix outlined the data scientist profession across two categories of questions, 'job specific' and 'competency', each with four subcategories. Job-specific categories referred to the technical skills of data scientists: 'computing' focused on programming languages; 'scripting' emphasised experience in using statistical tools such as R, SAS, SPSS; 'software' referred to big data analytics tools such NoSQL, Hadoop, Spark, and so on; and 'statistical skills' as the knowledge of traditional statistical methods, such as how to determine if a sample is representative. The competency category included references to broader skills that are applicable to all civil service positions and define a common core of skills and dispositions that civil servants are

expected to possess: collaboration, personal improvement, meeting deadlines, leadership, and communication with an emphasis on the ability to explain technical issues to non-technical audiences. Under this category, the job interviews defined the position of a data scientist but also differentiated it in relation to the national statistician by introducing public service skills and dispositions.

Of note is that the data scientists sought in the interviews were not being hired for a specific government task or practice. They could be placed in different government departments, but still expected to contribute their own skills independent of the domain. In other words, the cultural capital of the data scientist was conceived as highly convertible, allowing them to work in different domains with the same set of skills (cf. Mackenzie, 2013). However, all were expected to perform as civil servants in ways listed under the competencies category of the marking matrix.

The question-and-answer session made explicit many of the skills and values at stake in defining data scientists, but also for advancing and valuing the skills of national statisticians. To prove their potential as government data scientists, the candidates were expected to demonstrate their statistical expertise by answering questions 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 their application during the assessment round, asking him to 'please look at the statistical techniques required [for the position]'. Statistical tools were also discussed, as most candidates brought up Matlab, SAS, SPSS, and R when asked about their experience with software. R, short for the R Project for Statistical Computing, was often emphasised as the ideal tool due to its status as open-source

software, but also because it was 'less clunky than SPSS', in the words of one candidate. The interviewers also queried the applicants' familiarity with big data through questions such as: 'What did you learn from your experiences working with big data projects?' to which one candidate replied 'Use fewer programming languages', which displayed their familiarity with a shared perception within data science of the proliferation of tools and languages. Through their answers, the candidates implied that some new technologies were used in a project for the sake of having used them, and that such uses did not belong in 'proper' data science. Consequently, not only knowledge of skills and tools were tested and demonstrated, but also preferences and subtle distinctions.

Following each interview, the committee members were required to individually assign different scores to the eight subcategories in the marking matrix based on 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. The committee filled in their forms individually, and then discussed their answers to reach consensus on the final assessment of a candidate, which did not prove very difficult as their assessment of most categories were either the same, or very similar. During one such discussion, a committee member stated that given sufficient background such as a quantitative PhD, or prior experience in statistical programming, the applicants would be able to pick up some of the necessary skills even if they did not seem to possess them at the time of interview. In other words, they evaluated the applicants' potential to become data scientists. As explained next, some of this potential was articulated by referring not to data science skills, but to a set of skills that differentiate them from those of national statisticians.

Data Scientist As …

Recruitment processes are framed prior to interviews in application documents such as the job description, guidance for candidates, sample multiple-choice tests, and other supporting texts, as well as those submitted by applicants in the form of CVs and test answers. These documents describe the profession, and list expected skills, but the recruitment process is far from an exercise of fitting people into predefined boxes; the situated performing of the job interview also refines what it means to belong to a profession.

Who, then, are the data scientists as enacted by the job interview? They can 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 also 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 data practices.

In short, the job interview enacted the data scientist as possessing a set of skills and dispositions. Candidates were expected to possess cultural capital in the form of particular accumulated technical skills such as statistical analysis and programming that could be converted to advantage in the ongoing struggle to define the profession of data scientist (Halford and Savage, 2010). The candidates needed to possess

certain cultural capital such as statistical expertise and related technical skills to succeed in the recruitment process, but as the interviewers also acknowledged, the interview included an evaluation of their potential to become data scientists. That is, being a data scientist involved a process that built on cultural capital that a candidate already possessed but through the recruitment interview they needed to also perform the capacity to learn and acquire yet unknown skills. In this need to build on something, we identify their relation to the faction of national statistician. To become a data scientist involves a process of accumulating cultural capital beyond that possessed by statisticians, such as new programming languages, or familiarity with new data analysis tools as technologies change and evolve. The situated performance of the recruitment interview is where such future potential is assessed.

While recruitment interviews for data scientists valorised new skills in the data practices of government, skill alone was not sufficient. It needed to be bundled with other forms of cultural capital such as statistical knowledge as a foundation, as well as the habitus of a civil servant. However, in this specific bundle, technical skills counted for more when granting legitimacy to the performance of the data scientist candidate. When applicants argued for why different skills should be considered part of the bundle, they built on those of the profession of national statistician as a foundation while also differentiating theirs as more than just those of a national statistician. Some skills, for example familiarity with database management, a task once relegated to IT-specialists, played a much more prominent role, defining the data scientist and differentiating it from that of national statistician.

In these ways the situated performance of the recruitment interview made explicit the differentiation between the two factions in the field of statistics. What the valuations of

'future potential' and the competencies of public service skills and dispositions show is the forms of cultural capital that are stakes in the struggles for recognition in the field of statistics.

Innovation Events: Brainstorming Workshop and the Data Camp

From the Potential of a Job Candidate to the Potential of Big Data

To experiment with big data, NSIs need statisticians not only with data science skills but also 'big data sensibilities', as stated by a senior national statistician. Such sensibilities can be understood as making up a data scientist habitus: embodied cultural capital that includes tastes, habits, normative inclinations, and other knowledges and sensibilities that are not normally made explicit. Just as the skills that make up a data scientist and how they differ from or resemble those of national statisticians emerges through interviews, what constitutes the habitus of data scientists emerges through specific material-semiotic professionalising practices. We develop this by discussing our observations of two professionalising practices focused on innovation and organised by Statistics Netherlands (SN): a brainstorming workshop and a data camp. The events took place in the context of a wider debate within SN on the uptake of big data. Several introductory sessions and presentations took place before both events during which some statisticians regularly expressed their scepticism towards big data. For instance, a frequent objection to using social media data was that it is not representative of a national population, and that relevant background characteristics (age, gender) cannot be verified. In addition, using social media data would imply diverting from the international definition of a statistical

population, that is, usual residents (for instance, a Twitter population can also include tourists). As one statistician phrased it: 'this would be very dangerous.' But the events also involved a small group of statisticians within the NSI who already had an active interest in adopting big data and thus were committed to engaging in experimentation; this was the group participating in the two professionalising practices discussed in this section.

We first discuss the data camp to highlight how it fostered an entrepreneurial disposition necessary to develop future-oriented, 'risky' projects (Irani, 2019). Next, we highlight how this disposition included the capacity to work with techniques and visualisations that can demonstrate the potential of big data and data science. Rather than performing the future potential of job candidates to be data scientists, as elaborated in the previous section, the brainstorming workshop and data camp involved performing a disposition necessary to demonstrate the potential of big data. Furthermore, this disposition was not only performed and cultivated discursively, but also through material-semiotic data practices.

From Skills to Sensibilities

The aim of the brainstorming workshop was to develop ideas for public sector innovation by combining different types of data. It was organised by the NSI's innovation lab for the purpose of developing submissions to a competition organised by the Ministry of Economic Affairs. Eight people from different backgrounds and positions took part in a two-hour session, led by an NSI innovation expert. We first take a closer look at the fostering of a set of sensibilities or dispositions focused on the development of new projects.

From the outset, it was clear that to brainstorm is not only a cognitive process but also involves particular dispositions.

This included working at a fast pace and generating ideas quickly. For example, as participants (including one of the authors) gathered around a flip-chart they were advised to 'stay active' by standing up (lunch would be a stand-up lunch) and walking around. Moments for 'inward' individual reflection were limited in favour of fast and collaborative idea generation. All ideas counted; inhibitions and concerns were cast aside for the duration of the session.

Relevantly, the ideas to be generated were not randomly determined but targeted to particular goals; they needed to be future oriented, as the session leader explained: 'we need to be anticipatory, so that when it becomes relevant, we have the data … We want to know what the relevant issues will be in two years.' Following these instructions, the participants came up with a list of topics that included 'robotisation', 'clean drinking water', and 'what Google knows about its users'. The group leader made a point of explaining the difference. Statistics would not just be 'user-oriented', that is, tailored to the needs of policy makers, journalists and other user groups. Rather, the focus was on future 'social problems' to be mitigated by data analysis. Examples of such problems, as the group leader explained, are clean drinking water, whereas needs refer to the statistics that users state that they require.

But how then are social problems to be identified in advance? Or, as some participants phrased it, how can we 'get the signal from society'? New analytical techniques and media were discussed at length as possible solutions: 'We could vlog [produce YouTube reports], start an online focus group or become data journalists. As long as we can get the signal.' To apply these techniques, someone else said: 'We need to be able to take risks, to experiment, for instance by monitoring social media.' Referring to Google's Project X (a secretive and high-risk research facility funded by Google), she proposed to

form a 'risk taking group' with everyone present in the brainstorming workshop and other interested employees. To this another participant responded: 'the civil servant rebels! They [the "risk taking group"] are here to sacrifice sacred cows … they experiment with data and possibilities!' The brainstorming workshop thus helped develop a particular disposition required to generate ideas and projects that anticipate future social problems.

Whereas the brainstorming workshop was a short and focused event, the data camp was more immersive. It was attended by national statisticians, students, and researchers with PhDs in computer science and related disciplines. The format loosely imitated a hackathon, and included skills training, lectures, presentations, and group work. Twenty participants and seven mentors from SN and a university stayed on a university campus for a week. The mixed NSI-university teams worked until late at night on topics, not even stopping work during the 'data dinners'. Among the projects initiated by the teams were the analysis of Twitter data to learn whether gender can be derived from profiles or statements; the use of Twitter statements to predict tourist behaviour or crowded events; the use of road sensor data to predict economic growth; and the use of citizen science data to model the development of the blooming phase of flowers over space and time.

The data camp demonstrated how the ability to articulate the potential of big data necessitated acquiring sensibilities about particular analytical techniques and their aesthetics. Three sensibilities taught at the data camp illustrated this. The first was an 'appreciation of algorithms'. In the plenary sessions following the group work, and in reports about the group project, participants mentioned the relevance of algorithms, by which they referred 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. Correspondingly, in their evaluations and reports, participants emphasised the relevance of algorithms to process data and to get insight into data sets.

But such statements of relevance amounted to more than simply acquiring skills. Some participants stated that their work required an 'appreciation' of algorithms. For instance, one of the reported outcomes of an evening evaluation session was that 'algorithms love statistics' (see Figure 8.1). As automated correction and processing work also happens at NSIs, using algorithms was not new for statisticians. Yet in this instance statisticians referred to an intimacy between algorithms and statistics that helped them not only understand but also to realise the potential of large data sets to clean data so it can be analysed early on in the production process. However, participants emphasised in the plenary sessions that algorithms did not necessarily make data processing and analysis quick and simple tasks – they required patience. A data science habitus thus included an appreciation for algorithms paired with the virtue of patience to realise the potential of data (see Figure 8.1).

The second sensibility was a preference for a particular visual aesthetic; two lectures about visualisation during the camp are especially instructive. The first was given by the CEO of an NGO working according to the principle of what he referred to as 'objects of concern' (drawing on the work of Bruno Latour).[8] The speaker argued that this can require 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

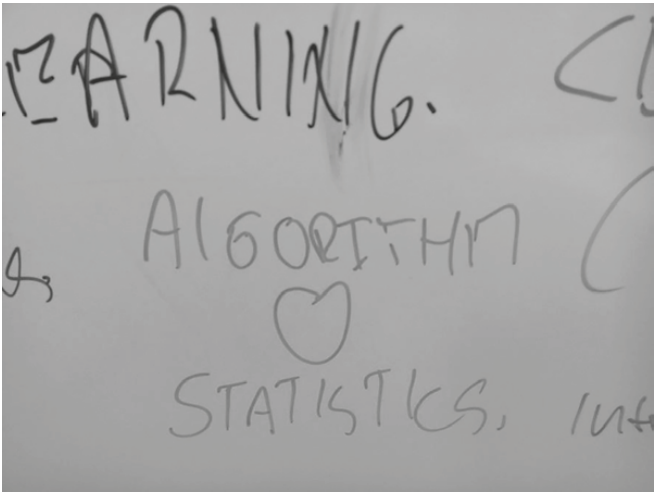

Figure 8.1 The Love of Algorithms
Source: Photo of the Data Camp's Whiteboard after Group Evaluation

making a choice as well'. The second lecture by a SN statistician contrasted this assertion. When the presenter was asked whether their visualisations had an explicit political viewpoint, they responded that they left the politics to the public, 'so you [analysts and statisticians] don't have to make choices'. Much like in the brainstorm workshop, what the first lecture introduced was an orientation to social problems, and in this case, the appreciation of an aesthetic to represent and bring attention to these problems. Furthermore, the NGO visualisations were presented as aesthetically more pleasing than those of the NSI. The NGOs were detailed, interactive, applied subtle colour schemes, and were easy to grasp because they were based on geographic maps. The NSI visualisations, although innovative, were clunky, less concerned with continuous and

cohesive colour schemes and, while understandable to statisticians, less easy for publics to read. The data camp mentors encouraged attractive visualisations, 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.

But visualisations were not only encouraged because they could draw in publics; they were also discussed and used as analytical techniques for interpreting large volumes of data that are not easily analysed using traditional techniques such as graphs. That is, the aesthetics of visualisations not only make them 'attractive' but through their use of contrasts, colours, and animations also facilitate analysis. So, while maps, graphs, and diagrams have always been part of statistical analyses, the difference here is the appreciation of the analytic possibilities of advanced aesthetics. Much like algorithms, they help to demonstrate the potential of big data.

A final sensibility was introduced by the NSI mentors in the context of preparing for the closing presentations: experiments and other risk-prone formats as instruments for developing business cases to support innovation projects. Participants were encouraged from the start to not only think in terms of results-oriented projects for specified groups of users, but also to be inquisitive and to take risks. This valuation of the importance of experimentation was underscored by the NSI's Director General in their presentation at the end of the final day of the data camp. As the data camp demonstrated, this included learning to manage the tensions in doing trial-and-error work that may not always lead to the desired results or quality in a short time span. Statements such as 'there is a lot in the data' helped resolve such tensions, as well as suggesting the future potential of a project.

The brainstorming workshop and the data camp are professionalising practices that made explicit dispositions of

embodied forms of cultural capital: a feel for the business case and users; the aesthetics of visualisations; experimentation; patience; and an affinity for and appreciation of algorithms. Rather than all-encompassing or constituting a universal data science habitus, these are some of the acknowledged sensibilities that make up an entrepreneurial habitus required to recognise the future potential of big data. While embodied by data scientists working in the technology sector, the sensibilities are valued for their capacity to solve social problems through the uptake of big data such as that generated by social media. It is through such valuations that it can be said that big data and the entrepreneurial habitus of data scientists are at once in tension and compatible with a public service habitus, which entails a commitment to working for the common good. In the following section we explore how this tension plays out in the professionalising practice of conferences where the entrepreneurial skills and habitus of data scientists were both valued and opposed by differentiating and defending them from those of public service.

Conferences: Defending by Differentiating

As in other fields, the profession of national statistician is shaped and defined through complex interactions and exchanges, from small meetings and official documents to those of international task forces and conferences. Amidst calls to embrace novel working practices, such as the data camp discussed above, statisticians also regularly convene at international meetings to discuss changes and challenges facing their profession. In this section we focus on Eurostat and UNECE conferences where big data and innovation were part of the agenda. The meetings followed a very traditional bureaucratic format of presentations and discussions typically centred on

PowerPoint presentations from authoritative figures in the field. Mundane arguments about the novelty of big data, data science, and so on were often repeated in an uncritical manner. Such repetitions highlight that the skills and habitus of a data scientist emerge not only through material-semiotic practices, but also through discursive ones. In what follows, we examine conference statements and debates as professionalising practices that involve differentiating the skills and habitus of national statisticians not only in line, but also opposition to that of data scientists.

As elaborated previously, data scientists are being defined not only in relation to particular sets of technical and analytical skills (or cultural capital) needed to manipulate large data sets, but also particular embodied sensibilities. For example, at a 2016 UNECE conference, a statistician criticised a lengthy presentation about the impact of big data on official statistics by pointing out that the change needed from NSIs goes much beyond the acquisition of new skills and toolsets:

> The [previous] presentation was very much tool oriented. We are very familiar with all these tools and the thing that was missing from the presentation was an acknowledgment of the fact that what is actually changing at the moment is the paradigm around how we conduct research. With big data you have the data first and then you ask the questions. The issue is therefore not what tools to use but what questions to ask. That's the crux of the matter, and that is where the skills come in.

NSIs are no longer 'the farmers', the presenter continued, but 'foragers of data'. As such, the key concern was 'what questions to pose and how to draw inference' and 'how to produce the best possible estimates to meet user needs from multiple data sources'. For other speakers, it was urgent that statistical

agencies shift their focus from producing statistics to a 'more service-oriented attitude … to connect, aggregate and tailor' statistical information based on user needs and to do so increasingly. 'Service orientedness' was an often-repeated term, which is defined in a number of ways. At this particular conference a consensus seemed to exist that 'service orientedness' refers to 'value added' activities such as analysing and interpreting data, rather than a narrow conception of the NSI role as data collector (UNECE, 2015: 4). In sum, big data is seen to disrupt not just established methods and techniques, but an entire paradigm of producing statistics, which also requires new sensibilities – for example, what questions to ask – and new skills – for example, what valued added activities to deploy.

For some statisticians, the appropriate response to what they conceive of as the challenge of big data is that NSIs need to become more like their private sector competitors. For example, at a meeting organised by Eurostat in 2016, a senior manager explained that not only do private companies now accumulate vast amounts of big data, they have the 'mindset of a big data company':

> The big advantage they [Facebook and Google] have is that they have the big data to accomplish a maximum effect. They also have the mindset of a big data company, which the statistical community does not. When we started using administrative data at [our NSI] statisticians were violently opposed to them with fundamental principle reasons. The same thing is happening with big data.' 'This is not statistics, this is not quality', they say. The first thing to do, therefore, is to get the mindset right.

The move from a product to service orientation was identified as involving a cultural change at NSIs, one that must begin at the very top level of managers. At a practical level, the shift in

mindset referred to in the above quote was conceived of as involving a willingness to accept different definitions of quality, since the sources from which data are derived are becoming increasingly varied. That NSIs look to the private sector for examples of adopting a more service-oriented approach is perhaps unsurprising. Amidst increasing competition, the 'modernisation' of statistics often refers to the adoption of a private sector mindset or entrepreneurial habitus.

However, the appropriation of an entrepreneurial habitus was not the only response of statisticians to the challenges of big data. Nearly as regularly, the future of the profession was also defined in contrast to values held in the private sector by reinforcing and defending long-held public service values in the production of official statistics. Indeed, while big data raised questions about the skills and competencies of national statisticians, existing values that they command such as trustworthiness, public accountability, civil service and democratic legitimacy were also defended. As in the case of job interviews, conference presentations stressed a public service habitus that values ethics and quality standards involved in the everyday production of official statistics. Such valuations occur, for instance, when some statisticians ethically objected to the use of corporate data sources because they cannot verify their quality according to formal standards (Struijs, Braaksma, and Daas, 2014).

These values constitute another repetition often asserted at international conferences: 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. As stated in a paper presented at a UNECE conference in 2013, official statistics have a 'trademark' based on quality criteria that need to be protected:

> 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: 2).

Statisticians, in other words, asserted their authority to establish, but also to evaluate adherence to, quality criteria in the production of official statistics. Thus, while the effects of big data are considered disruptive, it affords the opportunity to defend the relative advantages of the official statistics and the skills and habitus of national statisticians. Data scientists were not 'taking over' or replacing statisticians but were differentiated from national statisticians. In other words, while requiring new skills, big data is also (potentially) reinforcing established values and norms.

Yet again, like the different positions taken on the challenges of big data, counter arguments were also advanced about the extent to which NSIs can hold on to such traditional values in the midst of increasing competition between data producers. At a 2016 UNECE conference this came up in relation to discussions of data ethics. Responding to a presentation about the numerous potential ethical issues concerning NSIs using big data, a statistician made the point that even total abstinence would not free NSIs from ethical concerns. For them, this would only result in big data being left solely in the hands of actors who care less about ethical considerations than statisticians:

I am concerned about finding the right balance. In your assignment, you have explored all potential objections to using big data in official statistics. But there is also an ethical concern with us not engaging with the data, because even if we did not use them, others still would.

> For example, we have been experimenting using Twitter data, and our legal experts have been complaining to us about it. But individual social data is already on the market. Individual psychological profiles can be purchased from social media companies. This is the reality, and in this reality we cannot be too strict about ethics.

In other words, increasing competition from different private sector data producers raised a concern whether NSIs can hold on to their long-held principles such as those related to data ethics in the context of a 'new reality'.

This, as in the other professionalising practices, makes explicit tensions between entrepreneurial and public service skills and habitus. Whether discursive or material-semiotic, such tensions are manifest in multiple ways and how they play out and their consequences for the statistician subject and profession of national statistician are by no means settled or certain. Rather, they are objects of struggle over recognised forms of cultural capital (skills) and habitus (embodied dispositions) that are valued and recognised in the transnational field of statistics. It is to that point that we turn in the conclusion.

Conclusion

This chapter shifted attention to professionalising practices to understand how data practices not only involve struggles over methods of producing statistics. To advocate particular data practices also involves valuing the skills and habitus required to perform them and in turn the relative advantages that may be conferred to professionals who possess them. We highlighted how such valuing happens through both discursive and material professionalising practices. Acknowledging that there are numerous professionalising practices (e.g., training programmes, university curricula etc.), our aim is to exemplify

one aspect: how they involve a tension between the cultivation of entrepreneurial and public service skills and habitus.

This tension was evident in the professionalising practices that we analysed. Whereas the workshop and data camp blurred the boundaries between addressing social problems through technological innovation (an entrepreneurial positioning) and working for the common good (a public sector positioning), the conferences engaged in boundary making around values. With regard to the role of official statistics in the changing landscape of data production, this is likely to be an enduring tension. For example, business and political concerns were publicly raised about SN's use of private sector data, its increasing presence as a market competitor for work commissioned by businesses, and its uptake of predictive methods. In response, new regulations were adopted in 2020 that stipulated that SN would primarily produce statistics for the public sector (Brasser, 2019; Minister van Economische Zaken en Klimaat, 2020). What this exemplifies is that entrepreneurialism, innovation, and values of public service will likely continue to be objects of struggle within a political economy of data production.

However, following the statement from SN's Acting Director General that 'we'll still contribute to major social issues such as energy transition, sustainability, poverty and debt problems', among others through 'data-driven working and innovation', entrepreneurial and public service dispositions remain closely aligned for this NSI (Statistics Netherlands, 2020). Moreover, it demonstrates the relevance of professionalising practices: they play a role in cultivating the priorities and values relevant for how NSIs are positioning themselves in a changing landscape of data producers and shape the practices they adopt to produce official population statistics.

References

Bigo D (2011) Pierre Bourdieu and International Relations: Power of Practices, Practices of Power. *International Political Sociology* 5(3): 225–258.

Bigo D (2014) The (In)Securitization Practices of the Three Universes of EU Border Control: Military/Navy – Border Guards/Police – Database Analysts. *Security Dialogue* 45(3): 209–225.

Bourdieu P (1984) *Distinction: A Social Critique of the Judgement of Taste*. Cambridge, MA: Harvard University Press.

Bourdieu P (1989) Social Space and Symbolic Power. *Sociological Theory* 7: 14–25.

Burrows R and Savage M (2014) After the Crisis? Big Data and the Methodological Challenges of Empirical Sociology. *Big Data & Society* 1(1): 1–6.

Brasser P (2019) Hoe commercieel mag het CBS opereren? *Het Financieel Dagblad*. 13 December.

Gehl RW (2015) Sharing, Knowledge Management and Big Data: A Partial Genealogy of the Data Scientist. *European Journal of Cultural Studies* 18(4–5): 413–428.

Grommé F, Ruppert E, and Cakici B (2018) Data Scientists: A New Faction of the Transnational Field of Statistics. In: Knox H and Nafus D (eds) *Ethnography for a Data Saturated World*. Manchester: Manchester University Press, pp. 33–61.

Halavais A (2015) Bigger Sociological Imaginations: Framing Big Social Data Theory and Methods. *Information, Communication & Society* 18(5): 583–594.

Halford S and Savage M (2010) Reconceptualizing Digital Social Inequality. *Information, Communication and Society* 13(7): 937–955.

Haraway DJ (1997) *Modest Witness@Second Millennium. FemaleMan© Meets OncoMouse™*. Feminism and Technoscience. New York and London: Routledge.

Irani L (2019) *Chasing Innovation: Making Entrepreneurial Citizens in Modern India*. Princeton, NJ: Princeton University Press.

Kitchin R (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.

Lahire B (2012) *Monde Pluriel: Penser l'unité Des Sciences Sociales*. Paris: Seuil.

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

Latour B (2004) Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern. *Critical Inquiry* 30(2): 225–248.

Mackenzie A (2013) Programming Subjects in the Regime of Anticipation: Software Studies and Subjectivity. *Subjectivity* 6(4): 391–405.

Meulman J (2016) When It Comes to Data, Size Isn't Everything. *STAtOR* 17(2): 37–38.

Minister van Economische Zaken en Klimaat (2020) Beleidsregel taakuitoefening CBS. WJZ/19207028, Den Haag: Ministerie van Economische Zaken en Klimaat.

Muniesa F and Linhardt D (2011) Trials of Explicitness in the Implementation of Public Management Reform. *Critical Perspectives on Accounting* 22(6): 550–566.

Naur P (1974) *Concise Survey of Computer Methods*. Sweden: Petrocelli Books.

Pasquale F (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Ruppert E and Scheel S (2019) The Politics of Method: Taming the New, Making Data Official. *International Political Sociology* 13(3): 233–252.

Ruppert E, Law J and Savage M (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.

Savage M (2010) *Identities and Social Change in Britain since 1940: The Politics of Method*. Oxford: Oxford University Press.

Scheel S, Grommé F, Ruppert E, et al. (2016) *Transcending Methodological Nationalism through Transversal Methods? On the Stakes and Challenges of Collaboration*. ARITHMUS Working Paper 1. Available at www.arithmus.eu (accessed 1 April 2020).

Shapin S and Schaffer S (1985) *Leviathan and the Air Pump: Hobbes, Boyle and Experimental Life*. Princeton, NJ: Princeton University Press.

Statistics Netherlands (2020) New regulations direct the role of CBS. Statistics Netherlands Corporate. Avail. At www.cbs.nl/en-gb/corporate/2020/27/new-regulations-direct-the-role-of-cbs (accessed 17 April 2021).

Steinmetz G (2016) Social Fields, Subfields and Social Spaces at the Scale of Empires: Explaining the Colonial State and Colonial Sociology. *The Sociological Review* 64(2). SAGE Publications Ltd: 98–123.

Struijs P, Braaksma B and Daas PJ (2014) Official statistics and Big Data. *Big Data & Society* 1(1): 1–6.

UNECE (2013) *What Does 'Big Data' Mean for Official Statistics?* 18 March. Geneva: United Nations.

UNECE (2015) *Innovative Products and Services for European Statistics: Challenges and Opportunities*. ECE/CES/2015/42. Geneva: Conference of European Statisticians.

UNECE (2018) *Recommendations for Promoting, Measuring and Communicating the Value of Official Statistics*. New York and Geneva: United Nations.

9

Conclusion: The Politics of Data Practices

Evelyn Ruppert and Stephan Scheel

The chapters of this book have investigated data practices that statisticians and other practitioners in the transnational field of statistics mobilise to count and account for the people of Europe at a moment of major methodological changes. The move away from traditional questionnaire-based methods to administrative registers and the reuse of data produced via digital technologies such as internet platforms are not only stimulating methodological innovation but also diversifying methods for producing official population statistics. What is significant for this book is how these changes and diversification have consequences for the European Union's ambition to harmonise enumeration methods and data across member states. For us, the question has been: what then do these changes mean for making up what is the European population and in turn who are the people of Europe? Drawing on scholarship on the performativity of knowledge practices, we have responded to this question by adopting the understanding that statistical methods and related data practices do not just measure, mirror, quantify, or represent already existing populations. Rather they enact – or make up – Europeans as both an intelligible object of government (a population) and as a distinct peoplehood (a people). We have argued that doing so involves technologies of enumeration that states have historically deployed to make the people within their territories legible and under their control (Scott, 1998). But, bound up

with this statecraft and requirement of governing, population statistics – intentionally or otherwise – also help to enact a distinct form of peoplehood, a 'transnational European public' whose interests can be represented and championed by supranational bodies (Shore 2000, 19). That understanding is expressed in statements that population statistics 'make it easier for people acting at national or even regional level to see their situation as part of the larger European picture' (Eurostat, 2009: 1) and to know 'who are we' (Eurostat, 2015).[1]

However, as the chapters in this book have argued, making up the population of Europe requires data practices that are invented, recognised, legitimised, and circulate within the transnational field of statistics. Thus, to analyse data practices requires following them across myriad sites not only of NSIs but also government administrative departments, international organisations (e.g., Eurostat, UNECE, IOM, EGRIS), and private companies. To put this another way, it is not possible to speak of data practices as European or international or national for that matter. Moreover, making up a population requires classifying and encoding individuals into categories – such as usual residents, refugees, homeless people, and migrants. As these categories suggest, many constitute 'Others' and mobile subjects enacted in relation to a persistent and dominant conception that the population and the people of Europe are sedentary and reside within national borders as expressed in the category of usual residence.

Our approach was informed by two political considerations. On the one hand, the EU promotes European citizenship as strongly intertwined with freedom of movement and the right to live and work in an EU member state of one's choosing. Hence, mobile people – those who move between member states and who prior to the Maastricht Treaty were categorised as foreigners and migrants – are defined and encoded

as mobile European citizens. Accounting for and knowing this category is an important political objective of the European project, which seeks to facilitate and account for Europeans who exercise their mobility rights. Some of the book's chapters have analysed how particular national methods and data practices that produce population statistics may contribute to or counteract this political objective.

At the same time, migration acts as a foil for debates on contested political questions about Europe, such as the relation between religious institutions and the state, gender roles, or the practical meaning of freedom of expression. Hence, we follow de Genova (2016: 76) to understand contemporary debates on the migration question, and regularly reoccurring invocations of a migration crisis, first and foremost, as debates about competing notions of Europe and Europeanness. This not only concerns debates about migrant 'Others' from outside Europe, but also people who are EU citizens, as illustrated in discussions about health and welfare tourism (Mantu and Minderhoud, 2016), or the curtailment of the freedom of movement of Sinti and Roma (e.g., Plajas, M'charek, and van Baar, 2019; van Baar, 2018). Thus, how and through what kinds of data practices migrants, refugees, and other mobile subjects are enacted into being through population statistics have a direct bearing on these contested political questions.

It is such political questions to which we return in this concluding chapter. Hence, we do not reiterate in detail the analyses, conceptual moves, and findings of individual chapters. Instead, we provide a brief overview of our conception of data practices and how this was taken up in each chapter before we discuss five key political issues related to making up the population and people of Europe that emerge across chapters. In the final section, we suggest what these issues mean for official statistics, academic research, and citizen data rights.

In Chapter 2, we noted that while the terminology of data practices is widely used in a growing body of literature on datafication, an explicit theory or conception of data practices is yet to be articulated. Hence, we drew on contributions to practice theory to develop a conception of data practices as both empirical objects and conceptual register for analysing the data activities of statisticians. Empirically, we conceived of data practices as activities such as defining, collecting, generating, managing, organising, analysing, reporting, and circulating *data*. Conceptually, we adopted five theoretical commitments and related analytical sensitivities. To recall, data practices (1) are *sociotechnical* in that they involve relations between humans, materials, infrastructures, and technologies; (2) are *situated* in and produced by sets of relations; (3) are *performed* by actors who mobilise them as stakes in struggles over authority and power within professional fields of practice; (4) are *contingent* in that they are not determinate but rather involve continuous adaptations and practical adjustments; and (5) contribute to the *enactment* of the very objects and subjects that they seek to represent. Each chapter variously took up this conception through empirical analyses of specific data practices involved in making up categories of people with a focus on at least two theoretical commitments summarised in Table 9.1.

Chapters 3 to 6 analysed two specific data practices involved in classifying and encoding individuals into categories to establish equivalences between them and enact kinds of people. It is no coincidence that they highlighted the performativity of data practices since that is a central argument of the book. Each also focused on one or two of the other theoretical commitments. That also applies to Chapters 7 and 8, which stepped back from how data practices enact kinds of people that constitute Europe to consider two subject positions that

Table 9.1 Conceptual Overview of Book Chapters

Chapter	Data Practices	Category	Theoretical Commitments
3	Defining & deriving	Usual Residents	Sociotechnical Enactment
4	Coordinating & narrating	Refugees and Homeless People	Situated Enactment
5	Omitting & recalibrating	Migrants	Contingent Enactment
6	Inferring & assigning	Foreigners	Sociotechnical Enactment
7	Subjectivating practices: calibrating & sieving	Data subjects	Sociotechnical Contingent
8	Professionalising practices: differentiating & defending	Statistician Subjects	Performed Sociotechnical

data practices also produce and require: the data subject and the statistician subject. To account for the enactment of these subject positions, the chapters showed how data practices also involve subjectivation (Chapter 7) and are related to professionalising practices that reconfigure the recognised skills and habitus of statisticians (Chapter 8).

In addition to developing this conception of data practices, our empirical analyses accounted for normative and political struggles, stakes and choices implicated in them. As such, the chapters contribute to scholarship and debates on what can broadly be described as the politics of data.

That includes work on data feminism (D'Ignacio and Klein, 2020; Fotopoulou, 2020; Leurs, 2017), data justice (Dencik, Hintz, and Cable, 2019), data colonialism (Couldry and Meijas, 2019; Isin and Ruppert, 2019; Madianou, 2019) and data politics (Beraldo and Milan, 2019; Ruppert, Isin, and Bigo, 2017), amongst others. However, our conception of data practices is both theoretically distinct and empirically different. Rather than analysing the politics of strategies and technologies of governing policies and programmes, we analysed how politics are situated *in* and performed *through* specific data practices, that is, the activities statisticians engage in when they produce, share, analyse, and exchange various types of data. What our analyses show is that governing strategies almost never materialise as desired, imagined, or dreamed as there are always limits to how they play out in and through specific practices. This is not simply due to questions of technique, but the human and technological relations (sociotechnical) through which they must be realised, which include normative and political struggles, stakes, and choices (performed), and the unanticipated and collateral effects that they produce (contingent and situated). To put it differently, the theoretical and empirical approach that we have developed unravels grand strategies of power and rationalities of governing by attending – through situated analyses of the details, specifics, and contingencies of data practices – to how rationalities of power and governing get taken up and adjusted and play out across multiple sites and practices. It is also an approach to analysing the politics and effects of data practices that can be taken up in other inquiries in fields such as finance, health and social care, border and mobility management, education, and security, to cite a few.

In what follows, we highlight five political issues that cut across the chapters to emphasise how politics happen in and through data practices which are thus irreducibly political. In

sum, the issues concern (1) the sedentary bias of population statistics; (2) the double edge of enumeration; (3) the production of non-knowledge and the performativity of what is absent; (4) the politics of knowledge and the performativity of what is present in categories; and (5) the politics of method in and of data practices. In the final section we consider what these political issues mean for the future of official population statistics, academic research, and citizen data rights in making up the population and people of Europe.

The Sedentary Bias of Population Statistics

The sedentary bias of population statistics and the related problematisation of mobile people as special cases were explored in Chapters 3 and 4. The assumption that people are sedentary and normally have a place of usual residence located in one, and only one, bounded nation-state has long served as a basis of population statistics. More recently, the assumption underpins the 12-month rule recommended by UNECE and adopted in EU regulations to ensure the international comparability of population statistics.[2] Not only does the rule apply to census statistics, but also population statistics on refugees (Chapter 4) and international migrants (Chapter 5). The rule reinforces a sedentary bias, which is often at odds with transnational and mobile modes of living and thus raises methodological challenges such as the growing issue of 'double-counting' (Chapter 5). Consequently, it gives rise to data practices that attempt to account for movements whose durations are shorter and longer than 12 months. As Chapter 3 argues, this involves data practices that define special cases (e.g., posted workers), exceptions (e.g., higher education students), and exceptions to exceptions (e.g., cross-border workers). Hence, statisticians engage in numerous data practices to implement and

sustain the category of usual residence and in turn safeguard its explanatory power and legitimacy as the epistemic cornerstone of international population statistics. Be it by upholding the 12-month rule by defining special cases, exceptions, and exceptions to exceptions; or by deriving who is a usual resident from administrative registers (Chapter 3); or by problematising refugees and homeless people without a permanent address as 'hard-to-count' and then introducing exceptional practices to produce 'good-enough' numbers about them (Chapter 4); or by producing non-knowledge about the contingency, uncertainty, and unreliability of migration statistics through data practices like recalibrating (Chapter 5), a great deal of effort is invested in accomplishing and sustaining the rule.

The sedentary bias of population statistics also surfaces in the move to origin-based categories to which people are allocated in the shift to register-based methods through data practices like inferring or assigning, rather than practices of self-identification as in questionnaire-based methods (Chapter 7). Origin-based categories reify the dominant (and thus often implicit) assumption that people are – and normally should be – sedentary and that they belong to a particular stable ethnic group whose shared culture and identity are rooted in their intergenerational occupancy of a bounded territory. This territorialisation of culture and (national) identity (Malkki, 1992) is the assumption which underpins the data practice of inferring a person's origin from the place of birth of their parents and grandparents. It confirms Isin's (2018: 116) observation that '[t]he concept "people" itself already signifies an immobile, sedentary, and enclosed body politic bounded within a territory'.

The resulting enactment of citizens who were often born and brought up in their country of usual residence as 'foreign' is one moment in which the political implications of the sedentary

bias of population statistics comes to the fore in a stark and imminent way, with potentially serious consequences for the subjects concerned. Such consequences include the elevation of everyday racism and related discriminations to official policy and the enactment of 'second-class' citizens who are considered as 'foreign' and kept in a state of 'perpetual arrival' (Boersma, 2020). People facing such consequences can be confronted with an ever longer list of integration requirements, or even the denial of citizenship resulting in legal limbo and statelessness, which is, for instance, still faced by people belonging to Russian-speaking minorities in the Baltic states (e.g., Poleshchuk, 2013). What these examples highlight is that statisticians must engage in immense efforts to implement the category of usual residence, but in doing so contribute to sustaining its consequences. While the sedentary bias of population statistics is politically charged in its historical connection to the national and colonial order of things, it also has collateral political effects such as the problematisation of mobile people. These problematisations – and the impossibility of accounting for increasingly mobile modes of living – suggest that the solution is not to be found in evermore elaborate rules and definitions. Rather, the foundations of determining who makes up the population base of Europe requires fundamental reconsideration, especially for a European project that trumpets freedom of movement and the right to live and work within any member state.

A logical, initial response could reside in the suggestion to take inspiration from the 'mobility turn' (Sheller and Urry, 2006; Urry, 2000) in the social sciences and start from the assumption that mobility, and not sedentarism, constitutes the norm. In the field of statistics such a reconsideration is evident in experiments with new methods, mostly based on various big data, to account for increasingly mobile modes of living. Data scientists experimenting with mobile

positioning data generated through the use of mobile phones have, for example, developed a 'continuity model', which assigns numerous activity places and anchor points to individuals instead of allocating them to one place of usual residence and one place of work (Ahas et al., 2010). By tracing periods of movement and sojourn in particular places with geolocation data transmitted by a person's mobile phone, data scientists can create a continuous record of the places a person frequents and their mobilities between them.[3] While this mobility-oriented method potentially challenges the category of usual residence, it has also been deployed to sustain and stabilise it. Statistics Estonia (SE) statisticians have, for instance, experimented with mobile positioning data to identify and rectify incorrect information about people's place of usual residence in Estonia's population register (SE, 2017). According to statisticians, many people declare incorrect addresses to access certain benefits, such as prestigious schools or public transport (which is free only for residents of Tallinn) or to avoid taxation for their summer house in the countryside.[4]

The point here is that transcending the sedentary bias of population statistics through mobile methods is not a solution as the bias is not reducible to a methodological problem. Rather, it calls for acknowledging that it is, first and foremost, an epistemic and political bias that is deeply entrenched in the core concepts and operational logics of population statistics, which are essentially the 'science of the state' (Schmidt, 2005: 15). However, while a central practice of statecraft is that of 'sedentarisation' (Scott, 1998), modes of living are diversifying in ways that do not accord with this logic, including those enabled by EU law. Mobile people thus continue to be 'a thorn in the side of states' (Scott, 1998: 1), and a problem for population statistics especially in relation to the European

project which seeks to promote such movement in policy and account for it in statistics.

However, the Estonian example suggests that big data sources may offer a methodological solution to this 'thorn' through their potential to enact populations as mobile. What the 'continuity model' captures is the impossibility of allocating people to a single location and the possibility of tracing mobility between multiple locations.[5] The same can be said about the practice of sieving tweets to determine internal student migration in the UK explored in Chapter 7. So, while previous chapters identified data practices required to allocate people to a national territory to establish who is usually resident such as catch-recatch in the Netherlands (Chapter 3) or residency index in Estonia (Chapter 5), the continuity model seeks to establish mobility as a phenomenon *within* a national territory.

That people are variously mobile (e.g., weekly commuting, multilocal living, seasonal migration, and so on noted in Chapter 3) is hardly news. What these examples highlight – and which follows from this book's conception – is the mutual constitution between technologies, data practices and the version of population enacted. That is, big data and digital technologies offer the possibility of measuring and categorising mobility as a norm, and thereby enact populations as mobile, fluid, and modulating (Chapter 7). Moreover, while potentially challenging a sedentary bias, it is a possibility that is also driven by the will to know for the purposes of governing.[6] Enacting populations as mobile is not about satisfying curiosities but about attempting to innovate practices of statecraft by capturing human mobilities. For some subjects, that may lead to consequences such as taxation but for others much more serious threats such as deportation. That is a tension addressed in a second political issue that cuts across the chapters of this book.

The Double Edge of Enumeration

The second issue is the 'double edge of enumeration' (Chapter 4): being counted in official statistics is a precondition of political recognition and calculations of government social supports. Yet, at the same time, being counted can make people susceptible to intrusive and potentially harmful government interventions such as deportation. This double edge is characteristic of all knowledge practices that states rely on to create a legible population, such as practices of registration and documentation (Breckenridge and Szreter, 2012; Caplan and Torpey, 2001). While all people are potentially affected by the surveillant, intrusive and intervention effects of the double edge of enumeration, not everyone is equally so. Rather, the risk of being targeted by harmful or violent governing interventions is far greater for vulnerable groups, in particular people who have a precarious legal status, such as asylum seekers or illegalised migrants. This potential is well illustrated by the controversy around the proposal of US President Donald Trump's administration to include a question on citizenship status in the census questionnaire discussed in Chapter 1. Statisticians argued that introducing such a question would likely reduce the response rate of households with migrants who might fear that data could be shared with authorities to enforce deportations.

The double edge of enumeration was palpable in the analysis of refugees and homeless people (Chapter 4). Data on the number and characteristics of homeless people is advocated for the purposes of providing adequate shelter and improving social supports. However, the data could also be used to identify non-citizens among the homeless as a justification for conducting raids and potentially deporting them. It is for this reason that aid organisations argue that homeless people from

other EU member states were underrepresented in a recent count of the homeless population in Berlin. Homeless people may have evaded enumeration due to fears of being stripped of their right to freedom of movement as EU citizens and being deported, a state practice that has gained momentum in recent years (Memarnia, 2020). The exercise of the right to freedom of movement came to an end when EU rough sleepers in Britain were categorised as foreign by immigration laws introduced at the end of the Brexit transition period (Grierson, 2020). Under immigration rules that came into force on 1 January 2021, rough sleeping became grounds for refusal or cancellation of permission to be in the UK. That such fears of being enumerated are not unfounded was also illustrated in the Calais camp (Chapter 4) where France's statistical institute, INSEE, conducted a census with the help of local authorities a few months before its residents were evicted and the camp destroyed by the French police. In this case, data on the size and composition of the refugee population of the camp may well have been used for humanitarian purposes but given the enumeration preceded its destruction, it may also have contributed to the calculation of the tactics and equipment needed to destroy it.

These examples illustrate that the political answerability of statisticians and other practitioners involved in the production of population statistics not only concerns the statistics they produce, but also the potential uses to which they may be put. Given the performativity of data practices, this also concerns decisions about who is included and how they are categorised. On this point, the sociotechnical relations of data practices do not absolve statisticians or other practitioners from answerability. As scholars of material-semiotic approaches have stressed, human actors not only assemble but are responsible for the sociotechnical arrangements they constitute, engage, and participate in and the effects they

produce (Barad, 2007; Haraway, 2016; Law, 1992; Puig de la Bellacasa, 2017; Suchman, 2007).

A second and related point concerns the interpretation of and reaction to data subjects' reluctance, or even active resistance, to practices of enumeration and datafication. What the examples above highlight is that data subjects, in particular from marginalised or vulnerable groups, often have very good reasons to evade or subvert enumeration practices. This may include, for instance, providing incorrect or incomplete information, or not submitting answers at all. However, statisticians tend to register the refusal of asylum seekers, homeless people, and other data subjects more generally to provide correct, comprehensive information about themselves as problems of noise and 'dirty data' (Steyerl, 2019) that have to be identified and cleaned from statistical outputs.

Indeed, the self-eliciting subject has long been problematised as an unreliable source of data about whom statisticians increasingly seek solutions (Chapter 7). This includes experiments with methods that draw on and repurpose data such as that from government administrative registers or sources of big data mostly held by private companies. In both cases, information about individual subjects must often be inferred without relying on their consent or direct participation. Examples include deriving who is a usual resident in population registers (Chapter 3) or the use of register data to infer people's residency index (Chapter 5) or inferring a person's place of residence from electricity data produced by smart meters or mobile positioning data (SE, 2017). These data practices effectively bypass the data subject and largely limit their capacity to shape how much and what is known about them and for what kinds of purposes these data may be put (Chapter 7). In other words, data practices such as inferring are one way that statisticians respond to data subjects' practices of evasion or refusal whereby some try

to subvert or escape potentially harmful forms of datafication and government intervention. However, such a response constitutes a 'politics of debilitation' (Puar, 2017) that aims at minimising, bypassing, or erasing the data subject's capacity to act.

The Production of Non-Knowledge and the Performativity of What Is Absent

That data practices involve the production and circulation of various types of non-knowledge is another political issue addressed in this book. This was explored in Chapter 5, which analysed how data practices do not just produce data and knowledge, they also create, circulate, and perpetuate non-knowledge. They do so not only in the form of missing, incorrect, or unreliable data, but also through what is absent, that is, data that are never produced or if produced are not circulated. Just as in other fields of practice, a will to non-knowledge operates in the field of statistics alongside the will to know populations and to render people legible. Importantly, the will not to know and the production of various types of non-knowledge, such as doubt, ignorance, everyday secrecy, uncertainty or 'undone science' (cf. Aradau, 2017; Hess, 2015; McGoey, 2012b; Proctor, 2008; Walters, 2020), also create power effects that are distinct to those of the will to know elaborated by Foucault (1979).

That said, the production of knowledge and non-knowledge are entangled and intertwined in complex and multifarious ways. Furthermore, like the production and circulation of knowledge, non-knowledge is dispersed and cannot be attributed to single, identifiable actors producing 'strategic unknowns' (McGoey, 2012a) to further political interests or institutional agendas. Rather, non-knowledge may also be created through a non-transfer or mistranslation of knowledge from one field of

practice to another. It may also be the result of a field-effect in the sense that the production of non-knowledge is necessary for sustaining or satisfying certain doxa of a field. The enactment of migration as a reality that can be managed through the circulation of seemingly precise numerical facts about stocks and flows of migrants in the field of migration management, hinges, for instance, on the production of non-knowledge about the known limits of attempts to quantify migration (Chapter 5).

The production of non-knowledge was also apparent in other chapters. The performativity of statistical identity categories was found to reside not only in tacit assumptions, political agendas, and historical narratives concerning an imagined community of belonging (Chapter 6). It was also found to reside in gaps and absences that equally shape how populations are enacted and reified by identity categories used in official population statistics. Likewise, the definition of refugees and homeless people as hard-to-count populations justified the use of exceptional methods to produce data that is 'good enough' while at the same time acknowledging uncertainties, gaps, and inconsistencies in data (Chapter 4).

In sum, what the analyses in these chapters illustrate is that non-knowledge is as productive and generative as knowledge: it helps to enact the populations and people to which it refers in particular ways. As Renan (1996) argues, the constitution and reproduction of nations as imagined communities is as much based on what people actively forget as on what they remember in nationalist storytellings of a supposedly shared past (cf. Anderson, 2006). While we did not pursue this line of inquiry, as argued in Chapter 1, censuses and the statistical identity categories that make them up, are part of myriad nation-building and colonial practices such as official history textbooks, museums, statues, memorials, and other sites of memory politics, which also play a role in active forgetting.

The Production of Knowledge and the Performativity of Categories

The chapters of this book have also attended to the politics of knowledge and specifically how they play out in relation to the categories that make up a population and people. As noted in Chapter 2, statistics involve establishing 'categories of equivalence' that transcend the singularities of individual situations and thereby 'make a priori separate things hold together' (Desrosières, 1998: 236). The chapters variously examined how this works through data practices that classify and encode people into categories. In doing so, each chapter highlighted the performativity of categories which is especially pronounced when the power of naming intersects – as in official statistics – with the authority of numbers.

In Chapter 6 the performativity of categories was located in taken-for-granted premises, institutional interests, and political agendas that are ingrained in and carried by categories. This takes the form of mostly tacit assumptions which operate as self-fulfilling prophecies about the kinds of people and populations to which they refer. The category of the 'third generation' enacts, for instance, what Alba (2005) calls bright boundaries between the 'native' and the 'foreign', that is, hard boundaries which are virtually impossible to cross as they are anchored in ancestry, or more precisely, the place of birth of grandparents. These are exclusionary politics of belonging, which can have serious implications for integration policies and access to citizenship for migrants and ethnic minorities. The latter are construed – often with the help of statistics – as deficient subjects in need of more and better integration, a conclusion that is often mobilised to explain the structural disadvantages faced by minorities and to justify their subjection to a (potentially infinite) list of integration requirements. That such

categories are not given but representations that are situated and carry nationalist and colonial legacies is best illustrated when comparing them across states. While international conventions and efforts to harmonise categories abound, there remain myriad differences across national contexts that attest to the different possibilities of naming and categorising data subjects (Chapter 6).

Similarly, there are myriad differences in the methods and data practices through which data subjects are encoded into categories. These differences often stem from and carry national and colonial legacies. This is well demonstrated in states where registers play a significant role in the production of population statistics. Chapter 6, for example, analysed how the data practices of inferring and assigning involve repurposing data that cannot be influenced, changed, or contested by data subjects. A subject is encoded 'foreign' or 'native' depending on the place of birth of their grandparents (the 'third generation' category in Estonia), or a country code is assigned to them based on their place of birth (the 'Caribbean Netherlands' category). In these ways, data practices involved in processes of encoding operate as forces of subjectivation which configure the data subject's capacity to act and their ability to influence and shape the data of official statistics (Chapter 7). Indeed, such data practices are arguably becoming more prominent with the move to register-based methods as well as the use of digital technologies and big data.

The examples highlight how methodological debates and decisions are not reducible to technical and administrative matters but entangled with the politics of classifying and encoding. Of note is that the shift to origin-based categories, which carry essentialised and colonial notions of nativeness and foreignness rooted in ancestry and territory, are made possible through the reuse of administrative data stored in

government registers for operational purposes. It is by incorporating the sociotechnical arrangements of registers into the method assemblages that make up population statistics that these categories are not only possible but also how they are *done* in practice. In this way, the politics of categories are inextricably entwined with the politics of method.

The Politics of Method in and Through Data Practices

Methodological changes in the field of statistics have consequences for the strategies and interventions of government in many policy fields, ranging from transport planning to social policy, family planning, migration policy, and so forth (Hansen and Mühlen-Schulte, 2012; Schultz, 2018). This is one politic of method addressed in the chapters of this book in addition to three others that we summarise below related to the role of data practices in making up subjects – those who perform (statisticians) data practices as well as those who are subjected to them (data subjects).

The first concerns the impact of methodological changes on the composition of and power dynamics within the transnational field of statistics. Statistical methods and related data practices are *performed* by actors and function as stakes in competitive struggles over authority, influence, and resources within the field. Chapter 8 took up this issue to analyse how methodological changes and related data practices also involve professional struggles over the skills and habitus that are valued and cultivated in the field. That is, the valuing of a particular method also involves recognising the required skills and habitus to perform the data practices that they require and in turn the relative advantages of professionals who possess them. Significantly, through analyses of professionalising

practices the chapter shows how the skills, capacities, mindsets, and ethical orientation of the profession of national statistician are being repositioned in relation to a new faction, that of data scientists. That is, repositioning of both the skills and habitus of the profession of statistician is happening relationally: through the valuing and adopting of entrepreneurial skills and dispositions of data scientists and by defending and differentiating the public service skills and dispositions of statisticians. In this way, the chapter demonstrates how data practices are bound up with professionalising practices that reconfigure the skills and habitus of certain factions within the field of statistics and its power dynamics. Hence, the field of statistics emerges as an arena for the politics of method, which surfaces in struggles over methodological innovations and authority performed through both data and professionalising practices in the production of official statistics.

A second politic concerns how methods configure relations between the state and data subjects, where the shift to register-based, big data, and digital technologies diminish the agential capacities of data subjects. A critical political question posed in Chapter 7 is: what then are the possibilities for subjects to intervene and make democratic demands and claims about the authority and legitimacy of methods deployed to assert that this *is* the European population and people? To this we can add the question of who produces, configures, owns, and controls the register or big data through which the population and people are enacted and known. These are questions we return to in the final section below.

The third politic is that statistics help to enact – or make up – the population and who are the people of Europe, which is a fundamental starting premise of this book. From this it follows that methodological changes not only configure the agential capacities of data subjects but also the very object of

population is transformed, including who is rendered present or absent. This involves not only a politics *of* numbers which largely occur through the uses to which numbers are put after they have been produced. It also involves a politics *in* numbers that happen in and through the data practices that produce and circulate them (cf. Scheel, 2021). It is a politic closely linked to the objectives of harmonising data across NSIs to provide a singular account of the European population (Eurostat, 2019) through data practices. It begins with practices of defining and deriving who is a usual resident, which result in smoothing out differences between mobile lives and ignoring transborder relations (Chapter 3). It extends to practices that narrate homeless people within generic categories that render them a ghostly presence in data (Chapter 4). Moreover, it is a politic evidenced in broader struggles within the transnational field of statistics to make data on refugees internationally comparable by condensing and bracketing their myriad life situations and legal struggles into a statistical category (Chapter 4). Finally, the flattening effects of making data internationally comparable further happens in the transnational field of migration management where the non-transfer of knowledge about the uncertainties and gaps in data generated in the field of statistics is required to enact migration as a precisely knowable reality (Chapter 6).

Together such data practices suggest that to know the European population requires reducing the complexity of lives and flattening differences especially those that exist and persist in national categories and methods. While all statistics involve 'abstracting away individuality' (Porter, 1986) by 'establishing categories of equivalence' (Desrosières, 1998), arguably abstraction increases as granularity and specificity are reduced in the service of making data comparable across EU and international scales. But it is a comparability achieved

by harmonising the 'final statistical' product, the output, rather than methods of data production, the inputs (Baldacci, Japec, and Stoop, 2016; see discussion in Chapter 4). While there are good practical reasons for this, at the core is the persistence and insistence of the national order of things.[7]

Different Futures for Official Statistics, Academic Research, and Citizen Data Rights

If what and who we know as the population and people of Europe depend on and are enacted by methods and their related data practices, then such knowledge is not given or inevitable. That is one conclusion offered by this book where such possibilities can be identified by engaging with its conception of data practices. The conception offered is that knowledge (and non-knowledge) is an object of political struggle over the power and authority to name and enumerate and in which both humans and non-humans – that is, material and technological forms – are implicated. While much can be learned from analyses of grand strategies and political programmes, how they get taken up and adjusted and play out across practices so understood cannot be anticipated or reduced to their aims. Those aims may, for example, include commitments to include so-called hard-to-count individuals or mobile people through innovative practices as many government programmes promote, such as the United Nations' 'no one left behind'[8] initiative and efforts to produce statistics on homelessness across the UK to 'build a better understanding of this critical social problem'.[9] Many of these responses problematise, bypass, and replace self-identifying people through data practices that subjectify them by inferring who they are, what they think, and what they do, such as sentiment analyses of social media data or travel behaviour analyses

based on mobile positioning data. However, these responses and many others covered in the preceding chapters are problematic for two key reasons. First, they do not interrogate the built-in biases and assumptions that have rendered people 'left behind' in official population statistics because their social existence exceeds the dominant norms of contemporary societies. Importantly, these norms often reach back to (and thus highlight) the colonial and nationalist origins of statistics. Second, their adoption of data sources and digital technologies to include and incorporate people as part of a population simultaneously exclude possibilities for them to participate as data citizens in how they are classified and encoded as argued in Chapter 7. That is, whether homeless people, refugees, or migrants, for subjects to perform as data citizens requires possibilities for them to make claims and intervene in their subjectivation and shape how data is made about them.

Another way of putting this is that responses to the problematisation of methods such as questionnaire-based censuses imagine a future for official population statistics that ignores how methods and the data practices that implement and sustain them are objects of political struggles and contestation rather than technical problems to be overcome through digital technologies. In these closing paragraphs we consider a different possible future for official statistics based on *reimagining* some of its key foundations. As philosophers and political theorists have argued, to know what holds societies together requires understanding the imaginaries of its institutions. This is what is Anderson (2006) meant in their definition of a nation as 'an imagined political community' referenced in many chapters of this book. As Anderson elaborates, it is through shared imaginaries of technologies such as the census, the map, and the museum that colonial states came to govern their subjects and territories. Recognising that breaking

from such dominant imaginaries is a formidable challenge, it is at moments of innovation and experimentation with digital technologies and novel data sources that different imaginaries of futures are perhaps most possible (Ruppert, 2018). We suggest there are possibilities based on democratic processes that recognise the politics of method and that citizens have the greatest stake in how they are classified, encoded, and made into a population and people. Along with reimagining population categories and knowledge, in what follows we suggest that recognising such stakes is fundamental to imagine a different future for official statistics.

Reimagining Categories

A point that has been reiterated in this book is that categories of thought and practice such as usual resident, refugee, migration, and origin carry and perpetuate nationalist and colonial biases and assumptions. Indeed, the history of our present is to be found in the persistence of such legacies. These population categories inhabit not only governmental but also practices of academic research in fields such as development, migration, and demographic studies. As Savage (2010) demonstrates, the post-war social sciences are deeply entwined with political projects such as practices of statecraft through their development of scientific accounts of the state of the nation. In this regard, questioning the categories of thought and practice of official population statistics means to also question those of the academy. Yet, as special cases, exceptions, exceptions to exceptions, and problematisations of hard-to-count people reveal in relation to the usual residence category, much effort is required to accomplish and sustain categories. That includes data practices that engage with digital technologies that offer solutions and in turn uphold the legitimacy and

validity of this and other categories. Reimagining such a taken-for-granted category – which is fundamental to constituting a European population and categories such as origin, migration, and citizenship – is what the foregoing analyses in this book suggest. Accomplishing this is far more complex than a simple replacement of one universal (sedentarism) for another one (mobility) (cf. McNevin, 2019 on this point). Furthermore, as argued in relation to methods that enact populations as mobile and fluid, categories have consequences for not only how populations are known but also for how they may be governed. However, acknowledging and taking on such complexities is essential for an EU political project that is centred on the mobility of its citizens and which seeks to transcend political and methodological nationalism.

But more profoundly, sustaining categories such as usual residence can have major consequences for the exercise of social and political rights including the potential of people being subjected to harmful governing interventions. While much attention is paid to data practices that can classify and encode subjects into categories that correspond to governmental rationalities and interpretations, whether such categories are meaningful or accord with the lives, experiences, and rights claims of subjects is given scant attention. This is evident in how statisticians treat what is assumed to be incorrect or incomplete data as noise, and in how they problematise self-eliciting subjects as unreliable. It is also evident in data practices that reduce the actions of data subjects who seek to engage, evade, reject, or subvert enumeration practices to methodological or technical problems that can be solved by technological innovations and solutions. Instead, as suggested by former Eurostat Director General Walter Radermacher, there is a gap between citizen experiences and official statistics. In saying so, he stressed the need for a more democratic

debate between citizens and data producers and owners to achieve a 'more subjective, differentiated understanding of our world', instead of 'technocrats and politicians sitting together and confronting citizens in the end.'[10]

If categories do not reflect but enact subjects and as such are objects of political struggle over subjective meanings or rights claims, then data practices are necessary that recognise and make contestation possible. That is the form of data justice that we suggest arises from the chapters in this book. As those chapters argue, such possibilities are being reduced by data practices that engage with digital technologies and novel sources of data that seek to sustain the legitimacy and validity of categories. In doing so, they have also made more intrusive, widespread, and consequential uses possible and are arguably feeding distrust amongst subjects about governing intentions. While statistical authorities often justify such practices as necessary to capture hard-to-count subjects and produce better population statistics, their practices signal distrust in subjects and their potential role in the production of data. It is this question, the production of population knowledge that we turn to next.

Reimagining Knowledge and Non-Knowledge

Regimes of both knowledge and non-knowledge affect and shape how categories are enacted and how active forgetting of nationalist and colonial legacies is generated and perpetuated (Chapter 6). The performative effects of non-knowledge participate and sustain such forgetting through data practices that perpetuate a sedentary bias. More generally, data practices participate in making up categories or preventing them from being realised and in turn what come to be known as the population and people of Europe. Regarding the former,

practices such as inferring or deriving (Chapters 4 and 5) seek to allocate all subjects to a usual residence in ways that work for the definition and regardless of whether such an allocation is meaningful to subjects or accord with their rights claims. In other words, data practices can work in the service of sustaining categories by making them possible (Chapter 6 on origin-based categories) and when barriers to their realisation are encountered new practices are invented (Chapter 3 on usual residence category and catch-and-recatch). Critically, by sustaining categories data practices contribute to making them real and result in declarations like 'there are 16.9 million usual residents in the Netherlands'. However, such realities circumvent, and seek to limit the contributions of data citizens in their production. So, while statistical authorities may claim the legitimacy of official population knowledge, such knowledge is often not a product of the informed participation of data citizens. This brings into question not only the legitimacy of population knowledge – and the non-knowledge that it generates and requires – but also how it stands apart from the extractive and manipulative data practices of corporations that treat subjects as products to be exploited. It is on this point that we return to the question of stakes in official statistics.

Reimaging Stakes

A frequent refrain of statistical authorities is the necessity of serving the data needs of stakeholders including policymakers, academic researchers, local authorities, statisticians, non-governmental organisations, media, businesses, and the public. These refrains consider stakeholders as users whose stakes are simply the usefulness of statistics for the purposes to which data might be put. Policymakers have stakes in data for policy, academics in data for research, and businesses in

data for corporate decision-making. What then can we say of the two stakeholders we have considered in this book, statisticians and citizens? We have considered how data practices are enrolled in making them up – those who perform (Chapter 8 on statistician subjects) data practices as well as those who are subjectified by them (Chapter 7 on data subjects). Regarding the former, the politics of method involve competitive struggles where the stakes are the relative recognition and accumulation of cultural capital that this can confer. Those struggles include new data producers such as platform owners and data scientists who are ever more influencing the production of official population statistics. Such influence extends beyond that of individual actors; it involves adopting, for example, practices such as algorithms to determine usual residents (Chapter 3), platform logics and 'smart' technologies to format online censuses (Chapter 7), and entrepreneurial skills for innovating statistics (Chapter 8). It also includes new dependencies due to entanglements with the sociotechnical arrangements that make up big data produced by privately owned platforms or administrative data produced by government departments. The former involves regulated or monetised access to data, but also entanglements with the hidden assumptions, different objectives, and biases of platforms (Bruns and Burgess, 2015). The same can be said of administrative data, which is generated by-and-large to serve operational purposes of different government departments and are based on different definitions, standards, and practices (Chapter 4).

What then are the stakes for citizens? As we have suggested above, the struggle is for statistics that are meaningful or accord with their lives, experiences, rights claims and how they are governed. On this point, data practices need not be simply deployed for capturing subjects and steering them so that they submit in ways that work for official statistics as

conceived by governing authorities. They could also engage subjects in the co-production of data about themselves and the populations and people of which they are a part.

While the possibilities of digital technologies have been largely confined to cheaper, more efficient, more granular, and timely data extraction, they can also be enablers of interaction and co-production and a move from 'data driven' to 'democratically driven' data for making up the population and people of Europe (Ruppert, 2019). Rather than doing away with the struggles that this would entail, such an approach would recognise what is more generally understood as the politics of method which are a 'messy, competitive context [in which] the roles of different kinds of intellectuals, technical experts and social groups are at stake' (Savage, 2010: 237). But, for this book, to paraphrase Mol's (2002) conception of ontological politics, the stake they share in common are the normative and political values that make up one version of what is the population and who are the people of Europe, which in turn marginalises or precludes others. While that common stake is fought through debates and pronouncements of political strategies and statistical programmes, this book has sought to pay attention to the contribution of data practices in how it is both fought and won.

That reflection returns us to Chapter 1 and the relation between data practices in making up the population and people of Europe and broader debates on data. The digital interactions and transactions of people with various government, commercial and social platforms, devices, and apps are proliferating and making it possible for different authorities to measure, monitor, track, and analyse myriad aspects of social lives. For this reason, digital technologies have become political not only because people are increasingly engaging with them but because the data they generate is reconfiguring

knowledge and in turn technologies of governing. Data and politics are thus inseparable as they are enrolled in shaping social relations, preferences, and life chances. So, for instance, when data become population statistics, they can become powerful stakes in the policies and management of migration and pandemics, to name two recent examples. As previously stated, this involves a politics *of* numbers and the uses to which numbers are put. But becoming population statistics, as we have painstakingly detailed, also involves politics that happen *in and through* the data practices that produce and circulate them. Illuminating such politics of data practices is both the overarching aim and contribution that this book seeks to make.

References

Ahas R, Silm S, Järv O, et al. (2010) Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones. *Journal of Urban Technology* 17(1): 3–27.

Alba R (2005) Bright vs. Blurred Boundaries: Second-Generation Assimilation and Exclusion in France, Germany, and the United States. *Ethnic and Racial Studies* 28(1): 20–49.

Anderson B (2006) *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.

Aradau C (2017) Assembling (Non)Knowledge: Security, Law, and Surveillance in a Digital World. *International Political Sociology* 11(4): 327–342.

Baldacci E, Japec L and Stoop I (2016) Improving the Comparability Dimension of European Statistics by Minimizing Unnecessary Variation. In: *European Conference on Quality in Official Statistics (Q2016)* (conference paper), Madrid, 31 June 2016.

Barad K (2007) *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham, NC and London: Duke University Press.

Beraldo D and Milan S (2019) From Data Politics to the Contentious Politics of Data. *Big Data & Society* 6(2): 1–11.

Boersma S (2020) Narrating Society: Enacting 'Immigrant' Characters through Negotiating, Naturalization, and Forgetting. *Global Perspectives* 1(1): 1–11.

Breckenridge K and Szreter S (2012) *Registration and Recognition. Documenting the Person in World History*. Oxford: Oxford University Press.

Bruns A and Burgess J (2015) Easy Data, Hard Data: The Politics and Pragmatics of Twitter Research After the Computational Turn. In: Langlois G, Redden J, and Elmer G (eds) *Compromised Data: From Social Media to Big Data*. New York: Bloomsbury Academic, pp. 93–111.

Caplan J and Torpey J (eds) (2001) *Documenting Individual Identity: The Development of State Practices in the Modern World*. Princeton, NJ and Oxford: Princeton University Press.

Couldry N and Meijas U (2019) Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television & New Media* 20(4): 336–349.

de Genova NP (2016) The European Question. Migration, Race, and Postcoloniality in Europe. *Social Text* 34(3): 75–102.

Dencik L, Hintz A, and Cable J 2019. Towards 'Data Justice': Bridging Anti-surveillance and Social Justice Activism. In: Bigo D, Isin E and Ruppert E (eds) *Data Politics: Worlds, Subjects, Rights*. Routledge Studies in International Political Sociology. London: Routledge, pp. 167–186.

Desrosières A (1998) *The Politics of Large Numbers: A History of Statistical Reasoning* (ed. RD Whitley). Cambridge, MA and London: Harvard University Press.

D'Ignacio C and Klein L (2020) *Data Feminism*. Cambridge and Malden, MA: MIT Press.

Eurostat (2009) *The European Union Dissemination of 2011 Census Results*. Meeting of the Group of Experts on Population and Housing Censuses. 28–30 October. Geneva: UNECE.

Eurostat (2011) *EU legislation on the 2011 Population and Housing Censuses. Explanatory Notes*. Eurostat Methodologies and Working Papers. Luxembourg: Eurostat.

Eurostat (2014) *Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics*. Luxembourg: Eurostat. Available at: https://bityl.co/6TND (accessed 15 April 2019).

Eurostat (2015) *People in the EU: Who Are We and How Do We Live?* Statistical Books. Luxembourg: Eurostat. Available at: https://bityl.co/6NxG (accessed 8 January 2018).

Eurostat (2019) *EU Legislation on the 2021 Population and Housing Censuses – Explanatory Notes*. Available at: https://bityl.co/51ER (accessed 13 January 2020).

Fotopoulou A (2020) Understanding Citizen Data Practices from a Feminist Perspective: Embodiment and the Ethics of Care. In: Stephansen HC and Treré E (eds) *Citizen Media and Practice: Currents, Connections and Challenges*. Oxford and New York: Routledge.

Foucault M (1979) *The History of Sexuality. Volume I: An Introduction*. New York: Pantheon.

Government Statistical Service (2021) About the GSS Homelessness Interactive Tools. Available at: https://bityl.co/6HNa (accessed 6 April 2021).

Grierson J (2020) Foreign Rough Sleepers Face Deportation from UK Post-Brexit. *The Guardian*, 21 October. Available at: https://bit.ly/31HvW2W (accessed 25 October 2020).

Hansen HK and Mühlen-Schulte A (2012) The Power of Numbers in Global Governance. *Journal of International Relations and Development* 15(4): 455–465.

Haraway DJ (2016) *Staying with the Trouble: Making Kin the Chthulucene*. Durham, NC: Duke University Press.

Hess DJ (2015) Undone Science and Social Movements: A Review and Typology. In: McGoey L and Gross M (eds) *Routledge International Handbook of Ignorance Studies*. London and New York: Routledge, pp. 141–154.

Isin E (2018) Mobile Peoples: Transversal Configurations. *Social Inclusion* 6(1): 115–123.

Isin E and Ruppert E (2019) Data's Empire: Postcolonial Data Politics. In: Bigo D, Isin E, and Ruppert E (eds) *Data Politics: Worlds, Subjects, Rights*. Abingdon and New York: Routledge, pp. 207–227.

Law J (1992) Notes on the Theory of the Actor-Network: Ordering, Strategy, and Heterogeneity. *System Practice* 5: 379–393.

Leurs K (2017) Feminist Data Studies: Using Digital Methods for Ethical, Reflexive and Situated Socio-Cultural Research. *Feminist Review* 115(1): 130–154.

Madianou M (2019) Technocolonialism: Digital Innovation and Data Practices in the Humanitarian Response to Refugee Crises. *Social Media & Society* 5(3): 1–13.

Malkki L (1992) National Geographic: The Rooting of Peoples and the Territorialization of National Identity Among Scholars and Refugees. *Cultural Anthropology* 7(1): 24–44.

Mantu S and Minderhoud P (2016) Exploring the Limits of Social Solidarity: Welfare Tourism and Eu Citizenship. *EU Law Journal* 2(2): 4–19.

McGoey L (2012a) Strategic Unknowns: Towards a Sociology of Ignorance. *Economy and Society* 41(1): 1–16.

McGoey L (2012b) The Logic of Strategic Ignorance. *The British Journal of Sociology* 63(6): 553–576.

McNevin A (2019) Mobility and Its Discontents: Seeing Beyond International Space and Progressive Time. *Environment and Planning C: Government and Policy* Online First: 1–19.

Memarnia S (2020) Berlins erste Obdachlosenzählung: Es braucht noch mehr Daten. *Die Tageszeitung*, 9 February. Available at: https://taz.de/!5659420/ (accessed 10 September 2020).

Mol A (2002) *The Body Multiple: Ontology in Medical Practice*. Durham, NC: Duke University Press.

ONS (2019) The Emerging Picture of UK Homelessness and Rough Sleeping. Available at: https://blog.ons.gov.uk (accessed 6 April 2021).

Plajas I, M'charek A and van Baar H (2019) Knowing 'the Roma': Visual Technologies of Sorting Populations and the Policing of Mobility in Europe. *Environment and Planning D: Society and Space* 37(4): 589–605.

Poleshchuk V (2013) *Naturalisation Procedures for Immigrants Estonia*. Florence: EUDO Citizenship Observatory.

Porter TM (1986) *The Rise of Statistical Thinking, 1820–1900*. Princeton, NJ: Princeton University Press.

Proctor RN (2008) Agnotology. A Missing Term to Describe the Cultural Production of Ignorance (and Its Study). In: Proctor RN (ed.) *Agnotology: The Making and Unmaking of Ignorance*. Stanford, CA: Stanford University Press, pp. 1–34.

Puar J (2017) *The Right to Maim. Debility, Capacity, Disability*. Durham, NC and London: Duke University Press.

Puig de la Bellacasa M (2017) *Matters of Care: Speculative Ethics in More than Human Worlds*. Minneapolis: University of Minnesota Press.

Renan E (1996) What is a Nation? In: Eley G and Suny RG (eds) *Becoming National: A Reader*. New York and Oxford: Oxford University Press, pp. 41–55.

Ruppert E (2018) Sociotechnical Imaginaries of Different Data Futures: An Experiment in Citizen Data. In: *3e Van Doornlezing*, Rotterdam, NL, 14 June. Erasmus School of Behavioural and Social Sciences, Erasmus University Rotterdam.

Ruppert E (2019) Different Data Futures: An Experiment in Citizen Data. *Statistical Journal of the International Association for Official Statistics* 35: 633–641.

Ruppert E and Scheel S (2019) The Politics of Method: Taming the New, Making Data Official. *International Political Sociology* 13(3): 233–252.

Ruppert E, Isin E and Bigo D (2017) Data Politics. *Big Data & Society* 4(2): 1–7.

Savage M (2010) *Identities and Social Change in Britain since 1940: The Politics of Method*. Oxford: Oxford University Press.

Scheel S (2020) Biopolitical Bordering: Enacting Populations as Intelligible Objects of Government. *European Journal of Social Theory* 23(4): 571–590.

Scheel S (2021) The Politics of (Non)Knowledge in the (Un)Making of Migration. *Journal of Migration Studies* 1(2): 1–33.

Schmidt D (2005) *Statistik Und Staatlichkeit*. Wiesbaden: Springer VS.

Schultz S (2018) Demographic Futurity: How Statistical Assumption Politics Shape Immigration Policy Rationales in Germany. *Environment and Planning D: Society and Space* (Online First): 1–19.

Scott JC (1998) *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven and London: Yale University Press.

SE (2017) Statistics Estonia Is Looking for Volunteers to Participate in a Mobile Positioning Trial Study. Statistics Estonia, Press Release 14th June. Available at: www.stat.ee/en/uudised/news-release-2017-071 (accessed 10 October 2018).

Sheller M and Urry J (2006) The New Mobilities Paradigm. *Environment and Planning A: Economy and Space* 38(2): 207–226.

Shore C (2000) *Building Europe: The Cultural Politics of European Integration*. London; New York: Routledge.

Steyerl H (2019) A Sea of Data: Pattern Recognition and Corporate Animism (Forked Version). In: Apprich C, Chun W, Cramer F, and Steyerl H (eds) *Pattern Discrimination*. Minneapolis: University of Minnesota Press, pp. 1–22.

Suchman L (2007) *Human-Machine Reconfigurations: Plans and Situated Actions*. 2nd ed. Cambridge: Cambridge University Press.

UNECE (2019) No-one Left Behind: How Do We Know? Available at: https://unece.org/statistics/news (accessed 6 April 2021).

Urry J (2000) *Sociology Beyond Societies: Mobilities for the Twenty-First Century*. London: Routledge.

van Baar H (2018) Contained Mobility and the Racialization of Poverty in Europe: The Roma at the Development–Security Nexus. *Social Identities* 24(4): 442–458.

Walters W (2020) Everyday Secrecy: Oral History and the Social Life of a Top-Secret Weapons Research Establishment During the Cold War. *Security Dialogue* 51(1): 60–76.

Notes

1 Introduction: The Politics of Making Up a European People

1 For example, Eurostat developed a vision document for the post-2021 censuses that seeks to address the growing use of data from administrative sources and user demands for more frequent and more timely data than that which are currently available from a decennial census (Eurostat, 2016).

2 There are many appeals for a 'flexible Europe' whereby different state groupings can coexist such that Europe is made up of 'differentiated integration, closer (or enhanced) co-operation, concentric circles, Europe à la carte and two-speed (or multi-speed) Europe'. See Euroknow: www.euro-know.org/europages/dictionary/v.html.

3 During the period of ARITHMUS research, the EU consisted of 28 Member States. However, the EU's statistical programme also includes participation of countries of the European Free Trade Association participating in the European Economic Area ('the EEA/EFTA countries') and to Switzerland. It is also open to the participation of countries which have applied for membership of the Union and candidate and acceding countries.

4 The Trump administration also ended the 2020 census two weeks early which critics argued risked an undercount of difficult to reach people, particularly immigrants, transients, and the poor. The action was upheld by the Supreme Court. Gawthorpe, a historian of the United States at Leiden University, argued that this is one of

many ways the Trump administration attempted to meddle in the census to 'advance its goal of disenfranchising and immiserating parts of the country which do not vote Republican' (Gawthorpe, 2020).

2 Data Practices

1 One exception is Fotopoulou's (2019) work on citizen data practices that considers how the 'practice paradigm in the social sciences and media studies' can be taken up to study data practices from a feminist perspective (227).
2 The special issues (Cakici et al., 2020; Scheel et al., 2019) were the result of a workshop organised by ARITHMUS held in March 2017 at the Tate Exchange in London in the context of the programme 'Who are we?' For more information on the programme, see www.tate.org.uk/whats-on/tate-modern/tate-exchange/workshop/who-are-we (accessed 18 January 2018).
3 Schatzki adopts the terminology of 'practice approaches' in line with practice theorists who often use the expressions 'practice theory', 'practice thinking', and 'the practice approach' interchangeably as a way to stand apart from understandings that 'theories' can 'deliver general *explanations* of why social life is as it is' (Schatzki, 2001: 13). For similar reasons, we adopt the terminology of theoretical commitments and analytical sensitivities.

3 Usual Residents: Defining and Deriving

1 Sheller and Urry (2006) make this argument in relation to their call for a 'mobilities turn' paradigm. Social sciences, they argue, have 'largely ignored or trivialised the importance of the systematic movements of people for work

and family life, for leisure and pleasure, and for politics and protest. The paradigm challenges the ways in which much social science research has been "a-mobile" ' (209).

2 Member States are free to assess for themselves how to conduct their 2011 censuses and which data sources, method and technology are best in the context of their country (Eurostat, 2011: 9).

3 Prior to the internationally agreed definition in 2010, some NSIs were already using a definition of 'usually resident' as one population base but according to different definitions. The statistical agencies of these three international organisations worked cooperatively on a harmonised definition beginning with the 2000 round of censuses. For the UN and UNECE, the definition was adopted as a recommendation and guideline whereas for the EU it was adopted in an EC regulation.

4 (OED Online, 2018).

5 The definition of usual residence was introduced for the 2010–11 round of censuses and then amended in the UNECE guidelines and EC regulations for the 2020–21 round.

6 We refer to mobile people to decouple mobility from citizenship.

7 Following from an understanding developed by Law, Ruppert and Savage (2011), existing investments in infrastructures and practices are locked-in and would be difficult to change to accommodate or meet the requirements of a new definition.

8 See, for example, Ustek-Spilda's (2019) analysis of the discretionary decision-making and adjustments statisticians exercise when implementing international definitions because of discrepancies and contingencies of their national census methods.

9 These meetings took place between 2013–17 and involved those of the UNECE Experts on Housing and Population Censuses and a Eurostat Task Force. The following analysis draws on fieldnotes and observations at the meetings.

10 Fieldnotes. UNECE Group of Experts on Population and Housing Censuses, UN, Geneva, 30 September–3 October 2013.

11 As stated in the implementing regulation for the 2021 EU census data collection and based on the existing European Parliament and Council Regulation (EC) 763/2008 (EC, 2017).

12 Fieldnotes. UNECE Group of Experts on Population and Housing Censuses, UN, Geneva, 30 September–3 October 2013. This meeting discussed recommended guidelines for the 2020–21 round of census enumerations.

13 As stated in the implementing regulation for the 2021 EU census, data collection is based on the existing European Parliament and Council Regulation (EC) 763/2008 (EC, 2017).

14 Living apart together is a term used to describe people who have an intimate relationship but live at separate addresses for various reasons (Liefbroer, Poortman, and Seltzer, 2015).

15 This is one conclusion in a number of studies that have experimented with mobile phone data to analyse mobility patterns in Estonia (Ahas et al., 2010, 2014; Järv et al., 2014).

16 These statistics apply to the intra-EU movement of people who have citizenship in a member state. Interestingly, mobile citizens is a term that has also been used for colonial citizens who exercise their rights to reside in the metropole of empires as in the case of French Indochina (Pairaudeau, 2016).

17 The data is collected in accordance with the requirements of a regulation that aims to provide harmonised EU labour data (Eurostat, 2020).

18 Figures for total numbers of mobile citizens were from Eurostat Migration Statistics; data on working age (20–64) citizens were from the EU-LFS and came to 12.4 million, up from 11.8 in 2016. The EU-LFS was also the source for data on cross-border workers; data on posted workers was compiled from administrative data on the numbers of A1 Portable Documents issued in 2017.

19 Arguably, many of the special cases could be categorised as mobile citizens such as diplomats, military personnel, or children who alternate between two countries of residence.

20 Fieldnotes. ESS task force meeting, 30 June–2 July 2015. This meeting discussed recommended regulations for the 2021 enumerations.

21 Fieldnotes. Conference of European Statisticians, 64th meeting, OECD, Paris, 27–29 April 2016.

22 The European Migration Network (EMN) is an EU network of migration and asylum experts from Member States. It was established in 2008 to provide comparable information on migration and asylum, with a view to supporting policymaking in the EU. It is coordinated by the EC Directorate-General for Migration and Home Affairs.

23 The UNECE Task Force on Measuring Circular Migration draft final report, 'Defining and Measuring Circular Migration', covered existing concepts and definitions, dimensions and key issues for a statistical definition of circular migration (CES, 2016b).

24 The conference was held at the UN in Geneva, from 17–20 May 2016.

25 The International Passenger Survey (IPS) is conducted by the ONS, which collects information about passengers entering and leaving the UK to produce estimates of overseas travel and tourism.
26 The population of Sweden in 2015 was noted as 9,851,017.
27 Fieldnotes. Conference of European Statisticians, 10 April 2014, UN Geneva.
28 'Freedom of movement and residence for persons in the EU is the cornerstone of Union citizenship, established by the Treaty of Maastricht in 1992. The gradual phasing-out of internal borders under the Schengen agreements was followed by the adoption of Directive 2004/38/EC on the right of EU citizens and their family members to move and reside freely within the EU. Notwithstanding the importance of this right, substantial implementation obstacles persist, 10 years after the deadline for implementation of the Directive' (European Parliament, 2018).
29 Feasibility studies were required by article 8 of the Regulation (EU) No 1260/2013 of the European Parliament and of the Council of 20 November 2013 on European demographic statistics. NSIs were required to submit their feasibility study reports by the end of 2016 and could request financial support from Eurostat in the form of grants. More generally, the studies were to assess the scope for improving the comparability of concepts and definitions, and data quality and comparability.
30 For the 2021 round of census enumerations, 13 of the 31 EU/EEA countries are planning a primarily register-based census, eight a traditional census and ten a combined census generally based on a population register (EC, 2018: 8–9).
31 Dutch population registers are produced by municipalities to serve a variety of administrative purposes and are not kept by Statistics Netherlands. Statistics Netherlands

obtain data from these registers to produce a central population register dataset (PR dataset) for demographic and other population statistics (Statistics Netherlands, 2016: 40). After receiving the municipal population register data, the data are cleaned and several basic variables are imputed in a process called 'statistical production' (for instance, age is derived from date of birth). The result is referred to as the 'PR dataset' (technically referred to as the *Demografisch Deelregister*): the basic register-based dataset ready for the analysis and publication of statistics. Importantly, in the process of statistical production 'no adaptations are made with respect to the number of residents', so the 'size of the population follows directly from the [municipal] population register data (Prins, 2017: 19). One interesting circumstance is that the resistance to conducting a questionnaire-based (door-to-door) census to abolishing national census regulations in 1991. As a result, there is no national law stipulating or regulating the register-based census (other than EU regulations).

32 These are set out in Annex 1 and 2 of the study (Statistics Netherlands, 2016).

33 This is a simplified account of the CRC method in relation to only the police register. The full version of the method is based on applying the same procedure in relation to additional registers: the PR, the Crime Suspect Register (CSR), and the Employment Register.

34 This description also draws on the account in: (Statistics Netherlands, 2016: 13).

35 The procedure was conducted with both the CSR and a third, the Employment Register (ER).

36 The EU 15 includes all member states at the time of enlargement in 1995 when Finland, Sweden and Austria joined.

37 Using a personal identification number they were able to link data across different registers.
38 While this was the case in 2016, Statistics Netherlands started to address the production of data on homeless people in 2019.

4 Refugees and Homeless People: Coordinating and Narrating

1 Fieldnotes. Comments by a statistician at the International Conference on Refugee Statistics, 6–9 October 2015, Antalya, Turkey.
2 Fieldnotes. Reflections on a discussion amongst statisticians at a 2016 ESS task force meeting.
3 As the report on International Recommendations on Refugee Statistics notes (EGRIS, 2018), 'refugees and refugee related populations' include: (1) the population in a country needing international protection; (2) persons with a refugee background; and (3) persons who have returned to their home country after seeking international protection abroad. The UNECE convention of 'population with refugee background' includes foreign citizens who were 'forced migrants' together with their dependents living in the same household at the census reference time, including children born after the forced migration (UNECE, 2015: 136).
4 The French census is unique in that since 2004 it has involved a 'rolling census'. The method consists of 'a cumulative continuous sample survey, covering the whole country over an extended period of time rather than an enumeration carried out simultaneously in all areas relating to a specific reference date' (UNECE, 2015: 20). Municipalities of more than 10,000 inhabitants,

such as Calais, are fully enumerated in years ending in 1 and 6, and that is what triggered the count in 2016. INSEE notes that the municipality of Calais called for the camp to be enumerated because enumeration numbers affect State financial allocations to cover municipal costs such as waste management (INSEE, 2016: 5).

5 Camps are included in guidelines and regulations as a form of collective residence to be enumerated.

6 For example, they can be asylum seekers in the process of applying for refugee status, appealing the rejection of their application, or deemed deportable but their deportation cannot be enforced due to existing international agreements or human rights concerns.

7 Examples of data sources that might be used by a Member State include data from administrative registers for register-based censuses or data from questionnaires for traditional censuses. An example of input harmonisation is the European Social Survey, which is designed and implemented by the EC. It is a standardised survey centrally organised to produce data on social issues in the EU and beyond (Baldacci, Japec, and Stoop, 2016).

8 Their study involved participation research into air pollution sensing with residents concerned about the effects of hydraulic fracturing in Pennsylvania, US.

9 This echoes the argument put forward by Aradau and Huysmans (2018) on how credibility is assembled through transversal practices of knowledge creation, circulation, and accreditation.

10 This practical and pragmatic approach was well demonstrated in the CLANDESTINO project – Undocumented Migration: Counting the Uncountable – which attempted to quantify another group constituted as hard-to-count people, namely, undocumented migrants living in the EU

(Jandl, Vogal, and Iglicka, 2008). The project concluded that existing data was of poor quality because national sources are not comparable and ranged from guesses without foundation to serious attempts. Rather than resolve such differences, CLANDESTINO recommended an 'index of plausibility' to evaluate the quality of different numbers (CLANDESTINO Project, 2009).

11 Refugees are defined and protected by international refugee law and States' responsibilities are regulated under international law and national legislation. However, asylum seeker is not a legal term but a general term for someone who is claiming or applying for protection as a refugee and who has not yet received a final decision on their claim. It can also refer to someone who has not yet applied for refugee status recognition (has not yet formalised the administrative requirements in national law) but may nevertheless be in need of international protection. An internally displaced person (IDP) is someone forced to flee their home but who remains within their country's borders; while often referred to as refugees, they do not fall within legal definitions of a refugee. Following the publication of the EGRIS Handbook and subsequent discussions, a separate handbook on IDP statistics is being produced and thus IDPs will no longer be considered part of refugee-related populations.

12 Fieldnotes. International Conference on Refugee Statistics, 6–9 October 2015, Antalya, Turkey. Other examples are documented in EGRIS (2018).

13 For the EU, refugee statistics are also not part of census regulations but harmonised through a separate regulation on migration and international protection. The regulation sets out standards for concepts, definitions and methods with NSIs using data sources 'according to

their availability in the Member State' including: records of administrative and judicial actions; administrative and population registers; censuses; sample surveys; or other appropriate sources (EC, 2017b).

14 Sources for deriving data on status include that from administrative registers such as on the issuance of residence permits, work permits, applications for asylum, and tax or social security records (EGRIS, 2018: 63).

15 Fieldnotes. Reports of statisticians at the ONS Migration Statistics User Forum, London, 2016.

16 See Endnote 14; the statistician noted that Germany's estimate of 1 million refugees was reduced to 650,000 after applying the Eurostat definition of granted status.

17 Fieldnotes. Comments by a statistician at the International Conference on Refugee Statistics, 6–9 October 2015, Antalya, Turkey.

18 However, the history of the Calais camp did not end with the eviction of its residents: this was just one episode in an over 15 year history of encampments and dispersals of people, which has continued despite the destruction of the camp (Agier, 2018). Furthermore, most residents were relocated to different reception centres within France.

19 Sue Clayton is also Professor of Film and Television, Goldsmiths, University of London. More information on the documentary can be found at www.calais.gebnet.co.uk. For unaccompanied children who had the right to claim sanctuary and be transferred to the UK under the Dubs amendment to the Immigration Act, the Home Office reported that some 750 were transferred (Goodwill, 2017). The applicable law, known as the Dubs amendment to the Immigration Act, was named after Labour peer Lord Dubs, who in early 2016 forced the Cameron government to promise to give sanctuary to some unaccompanied

child refugees in the EU (Gentleman, 2016). One of the children in the documentary (ZS) took the Home Office to court, arguing that selection criteria for allowing unaccompanied minors to enter the UK during the demolition of the Calais camp in September 2016 were unfair and lacked transparency about the reasons for the rejection of applications, thereby making it difficult to launch appeals. An appeal court subsequently ruled that the government had broken the law (Bulman, 2018).

20 The UK government required that individual assessments be conducted to determine the eligibility of transfers (Home Office, 2018).

21 Source of estimate: (Solletty, 2016). A report by an NGO – Help Refugees – reported the number as 9,106 people, including 865 minors (Help Refugees, 2016).

22 Our analysis draws on ethnographic observations of their review at quarterly meetings of the task force held at Eurostat between 2014–17. We also draw on the analysis of some of these discussions in a related article (Ratner and Ruppert, 2019).

23 Launched in December 2014, the ESS Census Hub enables users for the first time to access, query, and download census population data for all EU member states via a single portal. See http://ec.europa.eu/eurostat/web/population-and-housing-census/census-data/2011-census. In the 1990s, the EC established guidelines for standardising national population data (definitions, classifications, categories) so that it would be comparable across EU states. For the 2001 round of enumerations, Eurostat assembled this data into tables and disseminated it in pre-defined cross-tabulations on key population topics (e.g., sex, gender and citizenship). In 2008, the production of data for the 2011 enumerations was for the first

time regulated by the European Parliament so that it could be disseminated according to different combinations of three to eight census topics (e.g., age, sex, nationality).

24 Topics is the convention for what is sometimes also referred to as variables: e.g., age, sex, nationality.

25 Living quarters were further defined in the technical specifications for another topic: 'Type of living quarters'.

26 The UNECE recommendations also only include primary homeless people as a core topic.

27 Fieldnotes. The review of the 2011 population data reported on the Census Hub noted that twelve countries were not able to provide any data on homeless people (ESS Task Force meeting, 31 June 2015).

28 See for example, the work of the UK ONS (Prestwood, 2019) and FEANTSA (European Federation of National Organisations Working with the Homeless) (Serme-Morin, 2017).

29 The metadata quotes are from the EU Census Hub entries on household status. See Endnote 26.

30 There are many different meanings of metadata and data cleaning. We focus on the meaning that is specific to the epistemic community we are studying, the transnational field of statistics.

31 The proposed wording stated, 'metadata shall report the number of all primary homeless persons and the number of all secondary homeless persons as well as provide a description of the methodology and data sources used to produce the data on homeless persons'.

32 The proposed wording stated: ' "data source" means the set of data records for statistical units and/or events related to statistical units which forms a basis for the production of census data about one or more specified topics for a specified target population'.

33 The draft wording was later amended to: ‘ “data source” means the set of data records for statistical units and/or events related to statistical units which *directly* forms a basis for the production of census data about one or more specified topics for a specified target population.’

5 Migrants: Omitting and Recalibrating

1 This chapter is based on, but also further develops, arguments and concepts in a previously published article (Scheel and Ustek-Spilda, 2019).

2 Agnotology refers to the study of non-knowledge. Importantly, non-knowledge is not simply understood as the negative of knowledge, but as intertwined with the former. Moreover, non-knowledge is – just like knowledge – productive and yields certain power effects. And, as in the case of knowledge, scholars of agnotology assume that different types and forms of non-knowledge exist, just as there are various tactics and practices to produce and sustain the former (Proctor and Schiebinger, 2008).

3 As elaborated in Chapter 1, we understand the field of statistics with Bourdieu as a field of practice in which various actors compete over influence, authority, and budgets by using various forms of capital as stakes in these struggles. Drawing on the works of Bigo (2011) and others who have tried to overcome the methodological nationalism of Bourdieu’s conceptual framework, we understand the field of statistics, however, as a transnational field. Likewise, we conceive of the field of migration management as a transnational field of practice also made up of struggles over influence and authority. In the following we only speak of fields, but always have the transnational dimension of these fields in mind.

4 We explain the interface and details of the GMFIA in the second section. While currently deactivated, the GMFIA can still be accessed via the internet archive, which creates copies of webpages at irregular intervals: https://web.archive.org/web/*/https://www.iom.int/world-migration (accessed 11 December 2019).

5 See for instance the annual *Risk Analyses* of FRONTEX, the European border protection agency, which has been active since 2004. Since 2010 the agency publishes several 'risk analyses' per year which are full of graphs and maps visualising seemingly exact figures about 'apprehended migrants', 'illegal border crossings' and so forth. For an overview of these reports see: https://frontex.europa.eu/publications/?category=riskanalysis (accessed 11 February 2020). UNHCR's *Interactive Dataviz* is, in turn, described as 'an archive of interactive data visualisation products created using various different technologies and software' on UNHCR's webpage. These data visualisations provide very precise figures on statistical topics related to forced migration such as number of new asylum applications in a particular region or 'First instance Decision Trends'. To access these visualisations, visit: https://data2.unhcr.org/en/dataviz (Accessed 11 February 2020). Likewise, the IOM's more recent *Flow Monitoring app* provides seemingly exact figures for the number of newly arrived migrants in Europe, disaggregated by year and migration route. Numbers are displayed in boxes that pop up if the user clicks on a particular migration route: https://migration.iom.int/europe?type=arrivals (accessed 11 February 2020).

6 The term doxa denotes what is taken for granted as self-evident in a particular society or field (Bourdieu, 1977: 164). In Bourdieu's later work doxa describes the

shared belief of all actors in the 'game and its stakes' that define a given field and which 'they grant recognition that escapes questioning' (Bourdieu and Wacquant, 1992: 98).

7 This and the following quotations relating to the GMFIA were taken from the homepage of GMFIA: www.iom.int/world-migration (accessed 11 July 2017).

8 Figures taken from Eurostat database: http://ec.europa.eu/eurostat/data/database (accessed 22 November 2017).

9 Fieldnotes, Statistics Norway meeting, April 2017.

10 Fieldnotes, Meeting of the Conference of European Statisticians, April 2014.

11 Figures retrieved from a query to SE's statistical database: http://pub.stat.ee/px-web.2001/Dialog/Saveshow.asp (accessed 29 September 2017).

12 Fieldnotes. Interview SE, December 2015.

13 Fieldnotes. Two interviews SE, March 2016.

14 Fieldnotes. Interview SE March 2016.

15 Fieldnotes. Interview SE, March 2016.

16 See the online *Cambridge Dictionary*: https://dictionary.cambridge.org/dictionary/english/recalibrate (accessed 17 May 2019).

17 Conducting a follow-up survey in order to assess the coverage of the census and identify possible 'coverage errors' is a standard procedure recommended by the UNECE (2015: 73–74).

18 Interview SE, June 2016. The metadata on SE's population statistics describes the model for the calculation of levels of unregistered emigration, in similar terms, with the following five parameters: 'unregistered migration is of the same rank as registered migration; over the years the ratio of registered and unregistered migration has shifted in favour of registered migration; age-specific distribution of unregistered

migration is the same as that of registered migration; share of males is somewhat greater in unregistered migration than in registered migration (ratio 6:4); on county level, the distribution of unregistered migration is the same as the distribution of registered migration' (SE, 2014).

19 Fieldnotes. Interview with a demographer at the University of Tallinn, October 2015. While SE's statisticians did not provide a reason why they did not also change data on emigration in the officially published migration statistics, it is likely that they refrained from doing so for two reasons. The increased negative net migration rate for the intercensal period would have pronounced even more the sudden shift to a positive net migration rate as well as the immense increase in both emigration and immigration rates after the introduction of the new RI-based methodology in 2015, which we have described in the previous section. It should be noted that statisticians also planned to recalibrate migration data for the previous years after the change in methodology in migration statistics in 2015 to even out the sudden increase in both emigration and immigration. Fieldnotes. Three interviews with statisticians at SE, June 2016.

20 Figures retrieved from a query to SE's statistical database: http://pub.stat.ee/px-web.2001/Dialog/Saveshow.asp (accessed 29 September 2017).

21 Fieldnotes. Interview SE, March 2016.

6 Foreigners: Inferring and Assigning

1 This chapter is a revised and updated version of an article published in the journal *Nations and Nationalism* (Grommé and Scheel, 2020).

2 Hence, 'refugee' or 'asylum seeker' do not resemble identity categories for us as they refer, first and foremost, to a particular legal status but not a socio-cultural identity.
3 The experiences of the Second World War, when statistical data on religion, ethnicity and race were used to exclude, discriminate against, and even mass murder minoritarian groups, has led to a discreditation and abandonment of statistics on ethnicity and race in Europe. However, this tacit consensus is increasingly coming under pressure by stakeholders in anti-discrimination policies who argue that a lack of data on groups that are affected by racism, antisemitism, antiziganism and xenophobia would make it very difficult to document discrimination and develop effective counter-measures (cf. Simon, 2012).
4 Fieldnotes. Interview SE, December 2015.
5 This working document reflects the outcomes of discussions of the first three meetings of the working group which took place on 9th May, 9th September, and 29th October 2012. The document is a 'living document' which includes comments, edits and additions by different stakeholders in different colours. Its unfinished status is precisely why the document illustrates very well the contested nature of identity categories. The working document was obtained during fieldwork and has been translated into English by a professional translation service.
6 See Teulieres (2007: 43) who observes 'like a mirror, the figure of the migrant unmasks the collective identities and symbolic boundaries of each community'.
7 Fieldnotes. Interview SE, December 2015.
8 English translation of the Annex to the contract between SE and EMC.
9 Fieldnotes. Interview SE, May 2015.

10 Table derived from SE's statistical database: https://andmed.stat.ee/et/stat (accessed 4 May 2017).

11 Fieldnotes. Interview SE, May 2016. In general, statistical categorisations along ethnic lines can be used for integration monitoring, exclusion and control, and to document and counter discrimination (Loveman, 2014). The absence of affirmative action policies in Estonia suggests however, that the 'third generation' category primarily serves as a monitoring tool for integration policies. These policies are based on a socioeconomic understanding of integration emphasising individual responsibility, in particular by requiring command of Estonian as a measure to improve the economic situation of the Russian-speaking minority (cf. Cianetti, 2015). The introduction of the third-generation category was pushed for by demographers of the University of Tallinn, who called for an unplanned meeting of the Scientific Council after violent clashes between the police and members of the Russian speaking minority in April 2007. The demographers successfully lobbied for the introduction of the third-generation category to produce more fine-grained knowledge about the Russian speaking minority (Fieldnotes. Interview Scientific Council, May 2015; Interview SE, May 2015).

12 The development of lists of officially recognised nationalities and nationality definitions were contested processes in which statisticians, geographers, ethnographers, government officials and lobbyists of ethnic groups were involved. Eventually, the determination of nationality through subjective self-definition was chosen to translate the promise of national self-determination – one of the main factors for the military successes of the Bolsheviks during the Russian Revolution – to the individual level (cf. Hirsch, 1997).

13 Fieldnotes. Interview SE, May 2015.
14 Fieldnotes. Interview SE, December 2015. The complexity of the taxonomy of identity categories used in Estonian population statistics is indicative of the complex politics of belonging at work in Estonia. In Estonian population statistics it is for instance possible to create tabulations that feature members of the 'third generation of the foreign-origin population' who do not hold legal citizenship as they are of 'undetermined citizenship', whose mother tongue is Russian but who nevertheless identify as Estonian when it comes to ethnic nationality. Conversely, there are members of the 'second generation of the foreign-origin population' who do not hold Estonian citizenship, whose mother tongue is Russian and who identify as Russian when it comes to ethnic nationality, and so forth.
15 Fieldnotes. Interview SN, February 2015, emphasis by the authors.
16 The *Caribbean Netherlands* are also referred to as the *BES-islands*. Following the dominant terminology of our SN research participants we use the former. Furthermore, we follow legal and governmental terminology in referring to the self-governing constituent territories of the *Kingdom of the Netherlands* as 'countries', instead of '(nation) states' (Charter of the Kingdom of the Netherlands 1954, 17/11/2011, article 5.1).
17 'Special municipality' (also referred to as 'public body' or *openbaar lichaam*) means that the islands are administrative divisions of the continental Netherlands modelled along the lines of municipalities (Oostindie and Klinkers, 2012).
18 Citizenship is not normally considered in definitions of migrants by statistical institutes. However, practices vary

among EU countries producing statistics about people born in overseas dependencies. France, for instance, does not include the populations of its overseas regions in its migration statistics.

19 The population register (PR) of the Netherlands is kept by SN and serves, since 2001, as the population base for SN's demographic statistics as well as the census. While demographic statistics are published monthly, quarterly or yearly, the census presents a 'snapshot' of the population every ten years. Another difference between SN demographic statistics and the census is that the latter defines the national population according to the Eurostat usual resident notion, whereas in the former registration in municipal population registers is a central criterion (see Chapter 3). Finally, the census includes a range of variables, including socio-economic variables that combine data from the PR with data from other registers (Schulte Nordholt, 2018).

20 The country codes were introduced shortly after the islands changed status in 2010. Bonaire, St Eustatius, and Saba each have a different code, for instance, Saba's country code is 5108 (Basisadministratie Persoonsgegevens en Reisdocumenten, 2016).

21 Even though some political parties in the continental Netherlands were in favour of immigration restrictions, the topic of migration was intentionally kept off the table in the negotiations between the island and continental authorities leading to the 2010 changes because of its political sensitivity (Oostindie and Klinkers, 2012).

22 The three islands are commonly not experienced as a single administrative or social entity by residents and local officials (Van der Pijl and Guadeloupe, 2015).

23 We reviewed demographic reports and articles published between 2010 and 2018, available in the online

SN archive (www.cbs.nl/nl-nl/onze-diensten/archief, accessed 14 August 2018). In this period, all publications specifically about the Netherlands Antilles and the Caribbean Netherlands origin groups concern urban residence, life expectancy, teenage motherhood, and single motherhood. We also checked a broader group of publications about relationships, fertility, and family regardless of origin group. Here we found that not all publications distinguish the Caribbean origin categories, but if they do, they highlight teenage motherhood and single motherhood.

24 Fieldnotes. Interview SN, October 2015.

7 Data Subjects: Calibrating and Sieving

1 This chapter builds on an article on methodological experiments with digital technologies by focusing on the sociotechnical and contingent aspects of the specific data practices that make up methods (Cakici and Ruppert, 2019).

2 Fieldnotes. Economic Commission for Europe, Conference of European Statisticians, Group of Experts on Population and Housing Censuses, Fifteenth Meeting, Geneva, 30 September–3 October 2013.

3 For a critique of the concept of 'data double' see (Scheel et al., 2019).

4 For further examples, see (Cakici and Ruppert, 2019).

5 Fieldnotes. Economic Commission for Europe, Conference of European Statisticians, Group of Experts on Population and Housing Censuses, Fifteenth Meeting, Geneva, 30 September–3 October 2013.

6 Fieldnotes. 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. 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.

7 Fieldnotes. ONS Beyond 2011 Research Conference & International Review Panel, 14 May 2014.

8 Registers were used in various ways such as to pre-fill some fields on questionnaires and supplement results when data was missing (Statistics Estonia, 2012).

9 The decision to conduct a predominantly online census in 2021 includes developing at the same time the use of administrative data from across government to produce 'more timely estimates' (HM Government 2018, 3).

10 Fieldnotes. 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.

11 Fieldnotes. ONS Beyond 2011 Research Conference & International Review Panel, 14 May 2014. Paradata has been used to evaluate and improve the functioning of surveys and understand respondents and how they answer surveys (Couper and Singer, 2013). It has also been used to track, evaluate, and intervene in the work of enumerators as they conduct censuses and surveys using digital devices.

12 The report argued that 'vendor lock-in, coupled with a particularly close and trusting relationship between the ABS

and its long-term supplier IBM, meant that the ABS did not seek sufficient independent verification and oversight of critical aspects of the e-Census' (MacGibbon 2016, 6).

13 Fieldnotes. The section draws on a series of meetings in 2014 and 2015 of a team of ONS statisticians in charge of the experiment and attended as part of ethnographic fieldwork.

14 API (an abbreviation of 'Application Programming Interface') is a method in software development where ready-made commands are provided to ease development or allow for additional functionality by other programmes. The Twitter API, itself a shorthand for several separate APIs, allows software to access data held by Twitter on posts, accounts, messages, and ads, to name a few.

15 Researchers working in the social sciences raise similar concerns about the relative instability and indeterminacy of digital methods because 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, Redden, and Elmer, 2015). They note that the development and deployment of digital 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).

16 Fieldnotes. These reflections are based on a series of interviews with national and international statisticians as well as observations at international meetings in 2015 and 2016.

8 Statistician Subjects: Differentiating and Defending

1 This chapter further develops the theoretical and empirical analyses in a previous publication by some of the authors (Grommé, Ruppert, and Cakici, 2018).
2 Whenever we refer to statistician in this chapter we are doing so in relation to that of national statisticians who are involved in the production of official statistics.
3 A quick look on Google Trends shows that the search term 'data science' started increasing in frequency around 2012. Peter Naur's 'Concise Survey of Computer Methods' is often cited as the source of the term 'data science' defined 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' (1974: 30). For examples of recent literature that refers to data scientists as experts who work with big data see (Burrows and Savage, 2014; Gehl, 2015; Halavais, 2015; Kitchin, 2014; Pasquale, 2015; Ruppert, Law, and Savage, 2013).
4 Also see Steinmetz (2016) on sub-groups and factions. We adopt the term faction to capture different professions that seek to distinguish themselves within the field of statistics.
5 We here summarise several writings by ARITHMUS researchers that have developed an understanding of the transnational field of statistics; see especially: (Grommé, Ruppert, and Cakici, 2018; Ruppert and Scheel, 2019; Scheel et al., 2016).
6 We concur with Bigo's understanding of habitus, which does not presuppose that a system of dispositions remains durable or generally permanent within an

individual but are a product of the strength and time of an actors' socialisation such that there are 'weak and strong systems of dispositions and competences' (Bigo, 2014: 210). Bigo's formulation builds on that developed by Bruno Lahire (2012).

7 Bigo's analysis is based on interviews with border security professionals between 2006–13.

8 After 'matters of concern' (Latour, 2004).

9 Conclusion: The Politics of Data Practices

1 Eurostat also maintains an interactive infographic, 'You in the EU', which enables people to compare their socio-demographic characteristics with those of others in the EU. See: https://ec.europa.eu/eurostat/cache/infographs/youineu/index_en.html (accessed 20 November 2019).

2 See Chapter 3. A usual residence is defined as the place at the census reference time which a person has or intends to live continuously for most of a 12-month period. A 'continuous period of time' means that absences (from the country of usual residence) whose durations are shorter than 12 months do not affect the country of usual residence.

3 Fieldnotes. Interview with a Data Scientist in Estonia, April 2016.

4 Fieldnotes. Two interviews with statisticians at Statistics Estonia, June 2016.

5 For other examples of experiments with mobile positioning data see (Ruppert and Scheel, 2019). Mobile positioning data is also enrolled in experiments to trace cross-border movements in the EU especially in relation to the development of tourism statistics (Eurostat, 2014).

6 Fieldnotes. This is a point that participants at a workshop in June 2018 conducted as part of ARITHMUS raised when discussing the definition of usual residence and the ways it does not accord with their mobile lives.

7 As noted in Chapter 4, countries are 'free to assess for themselves' how to conduct censuses including 'which data sources, method and technology are best in the context of their country' (Eurostat, 2011: 9).

8 This is broadly a United Nations call in relation to the sustainable development goals, which its regional commissions, such as the UNECE, have adopted (UNECE, 2019).

9 The ONS produced the first statistics on people who have died homeless in 2018 (ONS, 2019) and in 2020 the Government Statistical Service produced an interactive tool for exploring UK statistics on homelessness (Government Statistical Service, 2021).

10 Fieldnotes. Presentation by Walter Radermacher at the Eurostat conference 'Towards More Agile Social Statistics', Luxembourg, 28–30 November 2016.

Glossary

The following are the most relevant recurring acronyms (e.g., organisations, legislation) and terminology related to census methods referred to in this book.

Organisations – Acronyms

ABS	Australian Bureau of Statistics
CBS / SN	Centraal Bureau voor Statistiek Statistics Netherlands
CES	Conference of European Statisticians
CSB Latvia	Central Statistical Bureau of Latvia
DGINS	Director Generals of National Statistical Institutes
EGRIS	Expert Group on Refugee and Internally Displaced Population Statistics
ESS	European Statistical System
ESSC	European Statistical System Committee
ESS Census Hub	Launched in December 2014, the ESS Census Hub enables users for the first time to access, query, and download census population data for all EU member states via a single portal.
ESSnet	European Statistical System Network
EC	European Commission
EMN	European Migration Network
EP	European Parliament
EU	European Union

	During the period of ARITHMUS research (to 2020), the EU consisted of 28 Member States. However, the EU's statistical programme also includes participation of the countries of the European Free Trade Association participating in the European Economic Area ('the EEA/EFTA countries') and Switzerland. It is also open to participation of countries which have applied for membership of the Union and candidate and acceding countries.
Eurostat	Statistical Agency of the European Commission
FEANTSA	European Federation of National Organisations Working with the Homeless
Frontex	European Border and Coast Guard Agency
GMFIA	Global Migration Flows Interactive App
GSS	UK Government Statistical Service
HLG	High Level Group of the UNECE
INSEE	National Institute of Statistics and Economic Studies Institut national de la statistique et des études économiques French Statistical Institute
IOM	International Organization for Migration
MNO	Mobile Network Operator
NSI	National Statistical Institute
OECD	Organisation for Economic Co-operation and Development
ONS	UK Office for National Statistics
SE	Statistics Estonia Statistikaamet

SF	Statistics Finland Tilastokeskus
Turkstat	Turkish Statistical Institute Türkiye İstatistik Kurumu
UN	United Nations
UNECE	United Nations Economic Commission for Europe
UNSD	United Nations Statistics Division
UNSC	United Nations Statistical Commission
UNHCR	United Nations High Commissioner for Refugees

Legislation

GDPR	General Data Protection Regulation: (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.
Maastricht Treaty	Freedom of movement and residence for persons in the EU is the cornerstone of Union citizenship and was initially established for 12 Member States by the Treaty of Maastricht in 1992.
Schengen Area	It is comprised of 26 EU countries for which all passport controls have been abolished for their mutual borders. It is named after the 1985 Schengen Agreement signed in Schengen, Luxembourg.

Events

Brexit — The United Kingdom withdrew from the European Union on 31 January 2020 on the basis of a Withdrawal Agreement. The withdrawal followed the referendum vote of a slim majority of UK citizens in June 2016 to leave the EU. As of 1 January 2021, relations between the UK and the EU are governed by the EU-UK Trade and Cooperation Agreement (TCA).

COVID-19 — A novel, highly contagious coronavirus was declared by the World Health Organization in March 2020 as a global pandemic.

Terminology on census methods

12-month rule

Also: Usually resident population

The defined period of time for determining the inclusion of an enumerated person as part of the 'usually resident population' for the purposes of international comparison. It is composed of persons who have their place of usual residence in the country at the census reference time and have lived, or intend to live, there for a continuous period of time of at least 12 months. A 'continuous period of time' means that absences (from the country of usual residence) whose durations are shorter than 12 months do not affect the country of usual residence (CES, 2015: 20).

2020-21 round of census enumerations

Also: 2020-21 census round; 2010-11 round of census enumerations; 2010-11 census round; census enumeration

The defined periodicity of national censuses for international comparison is once every ten years. The actual enumeration date varies across two years with the most recent rounds conducted in 2010/11 and 2020/21. For the European Union, however, the census date must fall within the same year, for example, 2011 and 2021.

Administrative data

Also: Administrative Register, Population Register

Administrative data is based on 'records that are collected for the purpose of carrying out various non-statistical programs. This record keeping can be done by institutions belonging to the government sector or by private organisations. For example, administrative records are maintained to regulate the flow of goods and persons across borders, to respond to the legal requirements of registering particular events such as births and deaths, and to administer benefits such as pensions, or obligations such as the taxation of individuals and businesses' (CROS, 2020).

Other examples include national insurance, employment and health as well as population registers of persons who are considered residing in a given country and which also include

information about some of their characteristics (e.g., age, gender) (CROS 2020).

Big data

'Big data is characterized as data sets of increasing volume, velocity and variety; the 3 V's. Big data is often largely unstructured, meaning that it has no pre-defined data model and/or does not fit well into conventional relational databases' (CES 2013: 2).

Combined Census

Also: mixed method census.

A method whereby some information is taken from administrative sources such as a population register while other information is collected through questionnaires as in the traditional census or through sample surveys (CROS 2020).

Digital census

Also: E-census; Online census

The online conduct of a traditional questionnaire-based census (see below).

Population Base

'The '*population base*' is the population used for the compilation of statistical aggregates in a particular tabulation. This may be a sub-set, or the whole, of the 'population to be enumerated'. A country may adopt more than one population base

(for different statistical purposes), but one of these should always be the population base used for international comparisons purposes (more often the 'usually resident' population)' (CES, 2015: 76).

Register-based census

Also: Register-based statistics

A census that is conducted by obtaining data from various government registers and administrative sources (e.g., taxation, social security). Data is integrated normally by making use of a personal identification number that is unique to each individual and included in the various registers (CROS 2020).

Rolling census

A census where information is collected by a continuous cumulative survey covering the whole country over an extended period of time (years) rather than on a particular day or short period of enumeration (France) (CES 2015).

Surveys

The collection of data from a representative sample of a population based on a questionnaire.

Topics

Refers to the subject (e.g., place of birth) for which information is to be sought for each unit enumerated in the census (person, household, dwelling or building) (CES, 2015: iii).

Traditional questionnaire-based census

Also: Traditional census; Questionnaire-based census

'The traditional census is the total process of collecting (by means of a full field enumeration), processing, evaluating, disseminating and analysing demographic, economic and social data pertaining, at a specific time, to all persons and the housing stock in a country or in a well-delimited part of a country. It is taken in a given limited period immediately near to a given reference date (census day). Data are generally recorded on census questionnaires, being either in paper or, increasingly, electronic format, or via a secure online service provision.' (CES, 2015: 13)

References

CES (2015) Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing. ECE/CES/41. Geneva: UNECE.

CES (2013) What Does 'Big Data' Mean for Official Statistics? High-Level Group for the Modernisation of Statistical Production and Services (HLG). Geneva: UNECE.

CROS (Collaboration in Research and Methodology for Official Statistics) (2020). 'ADMIN Cross-linked Glossary.' Brussels: European Commission. Available at: https://ec.europa.eu/eurostat/cros/content/admin-cross-linked-glossary_en. (accessed 21 May 2021).

Index