CRESC Working Paper Series

Working Paper No. 68

Numbers Regimes: From Censuses to Metrics

Dr Evelyn S. Ruppert

CRESC, Open University

September 2009

For further information: Centre for Research on Socio-Cultural Change (CRESC)
Faculty of Social Sciences, The Open University,
Walton Hall, Milton Keynes, MK7 6AA, UK
Tel: +44 (0)1908 654458 Fax: +44 (0)1908 654488

Email: cresc@manchester.ac.uk *or* cresc@open.ac.uk

Web: www.cresc.ac.uk

Numbers Regimes: From Censuses to Metrics

Dr Evelyn S. Ruppert
Research Associate and Fellow
CRESC, The Open University

Abstract

Identification practices such as population and address registers, identity cards, biometric visas and passports, and the joining up of administrative data are rapidly becoming part of a new regime of government practices in the EU. The paper investigates how these practices are being advanced in the UK for the purposes of constructing and knowing whole populations as objects of management and government. The argument is developed that the knowledge of subjects compiled by identification practices constitutes a new numbers regime called 'population metrics,' made up of different measurements of subjects such as biographical, biometric and transactional data. Population metrics is also used to refer to the practice of combining, matching, assembling and reassembling metrics to identify 'new' populations. While identification practices raise concerns about privacy or surveillance, population metrics bring to the fore their totalizing effects, and raises questions about how we are known and governed. The paper poses some of these questions and suggests that this regime of metrics opens up a new politics of population.

Acknowledgements

I should like to thank colleagues in the Culture, Governance and Citizenship theme of CRESC for their comments on an earlier draft of this paper. I also acknowledge the financial support of the UK Economic and Social Research Council (ESRC) funding of award RES-000-22-3493, 'The Last Census: Governing Britain with Metrics.'

Keywords

population, metrics, biopolitics, census, identification

Numbers Regimes: From Censuses to Metrics

Introduction

Displaying or swiping a card that is linked to a database is a common everyday practice in states such as the UK. The card, rather than our words, mediates the verification of our identity and access to services, entrance to buildings, on-line purchases, commercial transactions, and so on. In many cases, the card has an embedded machine-readable chip that contains data such as a unique personal identifier, name, address, gender, and age and increasingly biometric data (as in the case of driver licenses, identity (ID) cards, visas and passports). Each time we use the card, we add another digital trace about ourselves, another transaction to a database. But cards are not the only mediator and link to databases. Unique personal identifiers such as national health (NHS) and insurance (NIN) numbers also connect us to databases, are used to verify our identities, mediate our access to services and track and record our transactions.

Each of the identification practices mentioned above is thus not only about verifying or authenticating identities but also the recording of transactions and movements in databases. New information and communication technologies (ICTs) have in part advanced these practices as they enable the storing, maintenance, searching and linking of large volumes of personal identification data. New technologies also make it possible to join up and share data regularly collected by separate government agencies and to incorporate new techniques such as biometrics and machine-readable microchips. However, these technologies have not been the principal drivers of these developments (Bellamy, 6, and Raab 2005).

The security imperatives of combating terrorism, organised crime, identity and benefits fraud and illegal migration have fuelled the advancement of identification techniques during the past few decades and their surveillance and control effects have been the object of much debate and analysis by major academic research projects. What these projects attest to is that identification techniques that have been mobilised to meet security objectives are also circulating and being taken up, modified and deployed at numerous governing sites and increasingly part of citizens' everyday lives. Indeed, the take up of these techniques in numerous government domains has led some critics to declare this as the age of the 'database state' or 'surveillance society.' These concerns are reflected in a large and growing literature in the fields of surveillance and security studies where identification techniques are analysed and critiqued in relation to their consequences for individual privacy, liberty, mobility, policing, data protection, discrimination and social sorting (for example, Ball 2005; Bigo et al. 2007; Graham and Wood 2003; Gutwirth 2007; Hier and Greenberg 2007; Lyon 2003; Marx 2007; Mordini and Petrini 2007). The ubiquitous nature of these techniques has led some researchers to declare that the centralized Foucauldian-Benthamite panoptic model no longer holds and instead current practices constitute a more decentralized Deleuzian assemblage or network model consisting of digital flows of data (Haggerty and Ericson 2000; Hier 2003; Lyon 2003; Mathiesen 1997). Whilst such interpretations of Foucault's panopticon resemble more of an Orwellian version of surveillance, they do point to a shift in identification techniques. In short, technological advances have enabled a shift away from the direct observation, supervision and containment techniques analysed by Foucault to the dispersed monitoring, recording and digital techniques that Gary Marx (2007) has named the 'new social surveillance' and Haggerty and Ericson (2000) a 'surveillant assemblage.'

The same techniques are being used and proposed to serve different governing objectives and producing many effects beyond surveillance and securitisation. For example, the introduction of identity cards and the building of centralised and joined-up government databases containing basic personal data are advocated to achieve policy objectives such as economies, efficiencies and service delivery improvements in the public sector. Indeed, the current

government's policy of service integration and evidence-based policies depend on joined-up information flows (6, Raab, and Bellamy 2005). The techniques also serve the objectives of freeing up of transactions and mobilities, simplifying access, and verifying and proving eligibility and entitlement to public services and benefits (Cabinet Office 2005, 2008).

Surveillance and the monitoring behaviour is thus one objective that is perhaps more central to techniques such as CCTV, ATM, Internet and mobile phones rather than identity cards and government administrative databases. That is, the techniques are not only means of surveillance nor do they necessarily produce negative effects such as the invasion of privacy. Like most governing techniques they can be deployed to achieve progressive governing outcomes such as ensuring equity and addressing gaps in public service delivery. Identification practices are also part of the struggle to both claim rights and make claims to rights. The rights to reside, cross borders, receive benefits and vote are all acquired on the basis of state identification practices. Constituting all of these practices as surveillance places the focus on information and privacy rights, rather than the plethora of other rights that they confirm and authorise. As Caplan and Torpey (2001) argue accounts that focus on the oppressiveness of state bureaucratic practices represent only one side of the 'bureaucratic identity equation.' Individual identification has been enabling as well as subordinating, has created rights as well as police powers. Higgs (2004) advances a similar argument, that the conceptualisation of identification techniques as methods of social control miss the ways in which data collected by states has historically served many social and legal functions. Additionally, there are many examples of how identification techniques have been taken up and used in creative ways from forgeries and frauds to the creation of new identities, the affirmation of existing ones, and as a way of writing oneself into history (Caplan and Torpey 2001, 7).

But there is another problem related to the constitution of identification techniques as surveillance and the consequent focus on privacy rights. People are not governed in relation to their individuality or identity but as members of populations. The embodied individual is of interest to governments insofar as the individual can be identified and recognized as a member of a population. As Foucault (1997) argued, the general problematic of governing is to know the nature and then govern and regulate the forces of the collective body, that is, the population. Population is the referent object of biopolitics, a form of power/knowledge concerned with managing, regulating and maximising the potential of a population, dealing with rates, profiles, patterns, and probabilities about a population and its ever changing, flowing and contingent nature (Dillon and Lubo-Guerrero 2008). Indeed, it is first through the identification of populations—of illegal migrants, terrorists or homeowners—that governing interventions are defined. How this referent object is constituted and the kind of knowledge that makes it up is the concern of this paper.

This is not to deny the importance of concerns about privacy or surveillance but to suggest that they obscure the politics of population, the totalizing effects of identification techniques and the consequences for how we are known and governed. The politics of population is not about us as individuals but how we are constituted as members of governable populations. As such it is a politics that brings into question the populations that methods conjure up and legitimise.

In government what we are witnessing is not the emergence of a 'surveillant assemblage' but a regime of technologically enhanced identification techniques that construct a new numbers regime and kind of knowledge of populations that can be analysed and deployed in myriad ways (e.g., surveillance) and produce specific power effects (e.g., inclusion/exclusion). The techniques constitute inscription devices—techniques that translate, simplify and make visible objects for governing (Latour 1986)—and a particular kind of knowledge of subjects that can be called 'population metrics,' a kind of knowledge that is leading to the discovery of 'new populations' and a new biopolitics of population. The construction of national population

statistics is one governmental use of population metrics that is explored in this paper. Censuses and more recently sample surveys have been the main methods that national statistical organisations have used to construct knowledge of populations. However, government officials and academics have long critiqued both methods as sources of population statistics (Judson 2007; Martin 2006). In the effort to create better population statistics, censuses are thus supplemented and potentially being supplanted by the various identification techniques discussed above such as joined up government administrative databases, population and address registers, e-Borders migration databases, and identity cards. For example, a proposal advanced by the UK's Office for National Statistics (ONS) calls for a 'integrated population statistical system' that would standardise and join-up many identification techniques and databases by making them interoperable (Office for National Statistics 2003). If implemented such a population statistical system could make the 2011 Census the last census in the UK as proposed by recent parliamentary inquiries and government studies (Office for National Statistics 2003, 2005b; Treasury Committee 2002, 2008). While the examples outlined in this paper pertain to techniques in the UK they are being adopted by many European states and also advanced and promoted by EU agencies (e.g., ID cards, e-Borders). Thus the theoretical arguments are generally applicable to other EU contexts.

The first section describes recent developments in identification techniques in the UK that make up the new numbers regime. Following this the argument that these techniques constitute a kind of knowledge called 'population metrics' is developed. The next section examines the kind of inscription device metrics are and the differences between knowing populations through metrics compared to censuses. Through this comparison the conclusion is drawn that population metrics are based on a conception of populations as modulating correlations of biographies and conduct. The next section then is a description of how these metrics discover 'new' populations. The concluding remarks suggest that population metrics are prescriptive rather than descriptive and constitute a new politics of population.

Population metrics

Historically, numerous practices have been involved in the constitution and classification of identities as legal and bureaucratic categories, which have been fundamental to the multiple operations of the state (Caplan and Torpey 2001). It is through a variety of practices of identification that governing authorities know a population and create a 'legible people' (Scott 1998). Improving the coverage and accuracy of identification have thus been key pursuits of states. During the past few decades, information and communication technologies have facilitated a move from the traditional paper-based documents to a variety of new identification techniques.

In the UK, the main innovations include a number of components that are currently proposed or being implemented as part of the UK National Identity Scheme (Home Office 2008). A cornerstone is the National Identity Register (NIR). If implemented, it will contain information similar to that stored on the passport database today, which covers 80 per cent of the UK population. It will include biographical data (such as name, address, date and place of birth, gender), biometric data (such as facial image and fingerprints), and administrative data (related to the issue and use of the identity card such as national insurance number, a log of who has looked at the record, and every transaction that the card has been used for). The data will be stored separately, on the Department for Work and Pensions Customer Information System (CIS), and existing Identity and Passport Service systems (Home Affairs Committee 2008). The Nordic countries have adopted population registers as a source of statistics since in the 1970s (Denmark, Finland, Sweden, Norway). In 1981, Denmark was the first country in the world to conduct a totally register-based census, Finland followed in 1990, from 1980, the censuses in Norway and Sweden have been partly register-based and these countries are

planning for their first register-based censuses in 2011 (United Nations Economic Commission for Europe 2007).

Connected to the NIR is the National Identity Card (ID Card). The ID card will be the size of a credit card and will show the person's name and photograph, and Identity Registration Number (IRN). It will also contain a machine-readable chip with the same biographical and biometric data stored on the NIR as well as a Personal Identification Number (PIN), which the cardholder can set and use. Beginning in 2008 the first identity cards were issued in the form of biometric immigration documents to foreign (non European Economic Area (EEA)) nationals.[1] Implementation plans currently involve issuing cards in the second half of 2009 to British and foreign nationals (including EEA citizens) working in sensitive roles or locations, starting with airport workers. From 2010 identity cards are to be issued on a voluntary basis to young people and from 2011/12 British citizens will start to be enrolled. If fully implemented, everyone residing in the UK could eventually be required to have one. Most EU states currently have some form of ID card in either non-electronic or digital form (Bennett and Lyon 2008).

The future of both the NIR and ID cards is uncertain at the moment of writing and may not proceed beyond the registration of foreign nationals. However, in the event that neither proceeds in their intended form, alternative systems are being developed and implemented. The Identity and Passport Service (IPS) together with the Department for Work and Pensions (DWP) and the Driver and Vehicle Licensing Agency (DVLA) is developing a minimum set of trusted identity data consisting of biometrics and a unique identifier (Home Office 2009). Everyone over the age of 16 applying for a passport will have these details added to the register beginning in 2011. Over time this will be expanded to include additional layers of data: one with data established at birth, which does not change, such as name and nationality at birth, and another that reflects changes since birth such as current name and nationality. Thus independent of the ID card, the data will be collected and a unique identifier assigned via the passport service. The point is that there are many techniques to fulfil the objective of assigning a unique identifier to every person to achieve 'identity management,' which is required to confirm identities and join up databases across government sites and functions.[2]

Identity management has been enhanced most notably through the addition of biometric identifiers to passports, visas, and immigration documents. Many EU states have incorporated biometric identifiers into some of their techniques of identification (Bennett and Lyon 2008). These enhancements will facilitate the e-Borders programme and will 'ensure people can be identified securely and effectively' and enable the e-Borders programme to collect and analyse information on 'everyone who travels to or from the United Kingdom by air, sea or rail' (Treasury Committee 2008). The Home Office coordinates the programme in partnership with the UK Border Agency, which is responsible for delivering the programme, and with the support of the police and HM Revenue & Customs. The ONS has been participating with the Home Office to use the e-Borders data as part of the tracking and estimating of international migration.

The ONS is also studying and evaluating the creation of an Address Register, a single database of addresses covering all properties in England and Wales to improve the collection of population statistics. It could include information on all properties including communal establishments and non-residential properties. In addition to basic information on the characteristics of each property, such a register could hold key characteristics of the population associated with it to support statistical needs, including turnover rate and multi-occupancy (Office for National Statistics 2003, 2005a). The registration of addresses has been most notably developed as either part of or separate from the population registers adopted by the Nordic countries (United Nations Economic Commission for Europe 2007).

In addition to these, joined up and interoperable administrative databases are being investigated as additional sources of identifying and constructing knowledge of populations. The possible kinds of data include the registration of life events (birth, death, marriage), licensing (driving, business), service use (education, benefits, health, pensions), employment (national insurance, taxation), and citizenship (identity cards, passports, visas, electoral registration). Through all of these activities (e.g., taxes paid, licenses obtained, benefits received) data are compiled on whole populations. Objectives of joined-up government have advanced the sharing or joining up of data within particular fields of public policy (e.g., health and social care) and across the public sector (e.g., community care case planning), though the former have been more successfully implemented than the latter (Bellamy, 6, and Raab 2005). The potential of linked administrative data for generating population statistics is also being advanced. While linking data from independent administrative sources is 'far from simple' and would require radical legislative reform (Jones and Elias 2006) there are many proposals, studies and initiatives underway seeking to increase the potential of this as a source of population statistics. For example, the ONS is investigating the linking of individual records pertaining to National Insurance Numbers, the Worker Registration System (WRS) and NHS Patient Registration data to track and measure population change (Treasury Committee 2008). Several countries have joined up administrative databases with central population registers such as Finland, Denmark and Sweden. Since 1996, The Netherlands has operated a social statistics database that links individual records from a variety of administrative sources (e.g. benefits data and employee insurance) with a population register (Jones and Elias 2006).

There are many technical, methodological, legal, data security and privacy issues associated with each of these techniques, which are at various stages of development and implementation. When and whether any or all will be fully and successfully implemented is uncertain. While each is proposed and can be deployed for different governing purposes, the focus in this paper is how they could be used to replace censuses and surveys as key sources of population statistics. Consequently, the discussion and analysis that ensues is in part speculative. Be that as it may, immense legal, bureaucratic and economic resources are being allocated towards the development of these techniques and their potential as sources of population statistics. Yet, analyses of these techniques have largely focused on technical and legal issues with little scrutiny of their epistemological and ontological consequences.[3]

What kind of numbers regime and knowledge of subjects and populations do these inscriptions construct? I argue that the knowledge constitutes 'population metrics.' In business, government and academia metrics or quantitative measurements are increasingly being adopted to evaluate and compare the performance and progress of people, groups, and things. For example, in education league tables and scores evaluate schools, in universities bibliometrics measure academic performance and in health care standards such as wait times evaluate service delivery. The same logic arguably applies to government identification techniques. Each of the identification techniques in question assembles different measurements of subjects that can be compared, combined and reassembled in myriad ways. That is, identification techniques are not the unit of knowledge but rather the specific measurements or metrics of bodies that they contain. For example, the ID card is the technique and its contents or categories are the metric.

The measurements are categories or classes of equivalence through which individuals pass from their singularity to a generality. Categories are 'conventions of equivalence, encoding, and classification, [that] precede statistical objectification' and are the 'bonds that make the whole of things and people hold together' (Desrosières 1998, 236). Generalising the individual into the population involves classifying and identifying her difference and resemblance to numerous categories (male, female, married, single, etc.) in relation to a pre-formatted classification grid (sex, marital status, racial origins etc.). A population is thus an entity divided and differentiated into numerous categories (Ruppert 2008). The work of making population thus involves establishing similarities and differences between and sorting

individuals into categories. For the state it is a totalising and objectifying technique such that when all categories are assembled (genders, origins, occupations, incomes, etc.) the entity called a population comes into being.

With population metrics then it is through the identification, measuring, recording, registering and monitoring of different *categories* of subjects (identified and identifiable natural persons) at myriad sites that populations and their performance can be and are being constructed, evaluated and governed. That is, categories of equivalence make up populations and individuals are governed as members of those populations. Numerous populations can be created by states for the purposes of governing and defined on the basis of different categories such as location (e.g., housing estate or neighbourhood population) or demographics (e.g., low income population). It is through their constitution as members of populations that people then become subjects of governing.

What then are the categories of subjects that make-up this numbers regime? The categories can be understood according to three types of classification of identification—biographical, biometric, and transactional. Each identification technique uses various combinations of these to construct what is referred to as 'data doubles' of subjects. Biographical data are the basic identifiers and locators of subjects and include classifications such as name, date and place of birth, gender, and address. It is the kind of data usually included on passports and will be used in the UK ID cards and National Identity Register (Home Office 2008). Biometric data are additional identifiers based on measurements of the physical attributes of bodies. The data consists of digital representations and measurements of physical or bodily characteristics such as fingerprints, eye retinas and irises, voice patterns, facial patterns and hand measurements. Such data pertaining to bodily characteristics can be and is already collected for identification purposes by EU states in areas such as immigration and naturalisation services, border controls, public aid programs, security systems, and medical services (Lodge 2007).

Biographical and biometric data are the core of what is sometimes referred to as 'identity management,' which is required to join up databases across government sites and functions. Different administrative systems currently define and constitute their target populations and subjects of government differently: from the general registrar (birth, death, marriage), revenue (NIN), health (NHS), passport and immigration (Home Office) and so on. These differences in part are due to the different governing objectives of each agency (health, border control, etc.). Identity management involves standardizing these classification systems so that they are comparable and can be joined up (Gandy 2007). This is part of the logic of the development of the NIR, which is key to achieving the inter-agency coordination necessary to join up administrative data. It is also the logic of EU initiatives, which seek to also standardize identification across member states to render them 'interoperable.' For example, an EC funded consortium of academics, technology firms and government—The Future of Identity in the Information Society (FIDIS)—has developed an identity classification system based on the logic of library schemes, which sequence knowledge systematically, and enable locating and accessing items via a catalogue (FIDIS Consortium 2007). The objective of the scheme is to provide a common classification system for government and commercial users towards creating interoperable e-government, e-health, and e-commerce systems.

While biographical and biometric data are principally used for the purposes of identification standardisation, verification and validation, when checked against and linked to other government databases they facilitate tracking the movement and conduct of subjects. In the commercial sector, transactional data are increasingly used to know populations in ways that are making the sample survey and in depth interview of social scientists outdated (Savage and Burrows 2007). Commercial users have increasingly turned to digital data generated routinely as a by-product of commercial transactions to provide comprehensive or total counts of whole populations (sales data, mailing lists, subscription data). Savage and Burrows (2007) suggest that the traditional market survey is being challenged by transactional data, which has the

potential to provide continuous and current knowledge of whole customer populations. Like commercial transactions, people regularly transact with government throughout their lifetime and collectively these produce administrative databases. Each of these databases records not only a subject's biographical identification but also conduct in relation to government: their registration of life events, income earned and taxes paid, licenses obtained, cars purchased, borders crossed, benefits received, visits made to hospitals, and so on.

Over the past decade or so, policy makers and data users have increasingly investigated transactional data as an alternative source of population statistics. For example, a recent audit of government data related to education, labour market, health, business and demographics conducted for the UK National Data Strategy concluded that the scale of available administrative data resources is extensive and insufficiently explored as a source of knowledge (Jones and Elias 2006). Of course, transactional data has long been used for population estimates between censuses. The population base of censuses is regularly updated by birth and death registrations (from aggregated transactional data), and modified by migration data (sometimes from source such as the UK National Health Service Registry data, or from migration surveys) (Judson 2007). Similarly, transactional data has been used to construct knowledge of particular populations (workers, migrants, students, and so on). However, the standardisation of biographical and the joining up of transactional categories across governing sites makes it possible to construct populations on the basis of various combinations of conduct. This is the innovation underpinning population metrics, which can be teased out by comparing the differences between population knowledge constructed by censuses and that of population metrics. The comparison is somewhat exaggerated in an effort to identify key characteristics of population metrics. But the comparison is apt since metrics are proposed as an alternative to censuses and are increasingly being used as a source of population statistics in a number of government domains discussed below.

From periodic to ongoing measurements: Historically, censuses have been taken every five or ten years and have constructed fixed 'snapshots' of populations at particular points in time. The time lag between censuses means that the data is always considered out-of-date and this has been a key criticism of censuses as a data source. However, population metrics offer the possibility of providing current and ongoing measurements of conduct and transactions with government as well as verifying subjects on a regularised basis. The data compiled constitute on-going and dynamic measurements of the movements and transactions of people in relation to government. Rather than stable or relatively fixed, populations are constituted as modulations, continuously changing and requiring constant monitoring, updating and assessing.

From varying to fixed biographies and biometrics: Censuses involve subjects periodically identifying themselves with classification systems such as gender, income, occupation and ethnicity. These classifications are subject to variation due to changes in how individuals report from census to census, changes in the questions that are asked and in the way the questions are posed. However, within a regime of population metrics subjective identifications are centrally coordinated, standardised and stabilised. A National Identity Register seeks to stabilise these core biographical identifiers, with many presumed fixed throughout a subject's lifetime.

From partial to whole populations: Undercounting has always plagued census taking and more recently response rates have been in decline especially in relation to 'hard-to-count' populations such as transients, migrants, tenants, and youth (Office for National Statistics 2004). The identification techniques that construct metrics cover almost 100 percent of target populations with some of the best coverage noted in education, benefits, and hospital episode data (Jones and Elias 2006). Population metrics can also be tracked over time (and back to previous time periods) thereby enabling longitudinal analyses with less problem of attrition or the 'loss' of individuals that can occur with surveys or census data linkage.

From engaging subjects to recording conduct: Biographies and biometrics are stabilised classifications that are centrally administered and coordinated. Whilst censuses principally measure biographical changes in a population, population metrics also measure transactions, a dynamic classification whereby the population to which one belongs can be defined based on their transactions with government agencies: offender, patient, welfare recipient, pensioner, migrant, and so on rather than Asian, married, and elderly. Because transactional data is recorded through established means of data collection (e.g., tax, national insurance or school records), the collection procedure is deemed less intrusive or burdensome for subjects (unlike surveys or censuses) (Jones and Elias 2006). Correlations with biographical data are of course still relevant but the emphasis is on conduct in relation to government. Thus populations can be constructed on the basis of what people do in relation to government (transactions) and less on the basis of what they say they do and who they say they are. Descriptions of conduct thus become more important than subjective identifications.

From probable to actual conduct: Census data are used to establish *causal* links between subjective identifications (biographical attributes and variables) and probable conduct. Whilst censuses also include some measurements of behaviour in relation to variables such as employment and housing, these are based on subjects' reporting and recall in response to questions (e.g., type of housing, change of address, employment status) and suffer from the same criticisms as that noted for the reporting of biographical data. The analysis of census data involves variable-centred linear modelling to produce generalised explanations of how variables might work together such as between biography and employment (Uprichard, Burrows, and Byrne 2008). In comparison, population metrics can be used to make predictions using data on individual cases of *actual* conduct (activities, movements, transactions).

From separate to joined-up databases: Censuses construct individual level datasets that are separate and distinct from other population datasets compiled by government agencies. The population is then the compilation of biographical classifications recorded by the census. However, the logic of metrics is that the revealing classification of subjects is conduct in relation to government. By joining up data on this classification across governing sites and by identifying patterns and correlations in the conduct of subjects across several domains, security threats, risks, fraud, inefficiencies and service gaps and problems can be detected and addressed. It is through joining-up of different bits of data about subjects that correlations between the subject's biography and their transactions and that of others can be identified and evaluated. In this regard, joined up data is considered to be a more comprehensive evidence base for policy-making (Jones and Elias 2006).

In sum, population metrics are based on a conception of populations as *modulating correlations of biographies and conduct*. The recording of conduct has of course always been the basis of government administrative systems. The difference population metrics make is in the standardisation of biographical and the joining up of transactional categories across government sites and functions such that what varies or counts is conduct (movement, activities). Whilst all techniques of identifying and knowing populations construct data doubles of subjects (e.g., census, NHS or NIN doubles), the logic of metrics involves assembling categories in novel and myriad ways to produce 'new' data doubles and populations. That is another reason why the relevant unit of knowledge is the categories of subjects rather than the identification techniques themselves. Populations are made by combining and matching different categories of subjects collected by various identification techniques. The next section considers the techniques that make possible the discovery of new populations.

New populations

What happens when the referent object of biopolitics—the population—undergoes a change and is created on a different basis? This is a question Dillon and Lubo-Guerrero (2008) pose in relation to the life sciences and the impact of the molecularisation of biology on the understanding of population. The same question can be posed in relation to population metrics: what happens when populations are constituted by metrics and understood as modulations of correlated biographies and conduct? One hypothesis is that populations become entities that cannot simply be enumerated but must be discovered. This can be explored through examples of how metrics are being and can be used to discover 'new' populations.

Current practices of joined-up administrative data illustrate how population metrics enable the discovery of new populations. From identifying children 'at-risk' to inequities in the allocation of services and resources to particular populations there are many data sharing arrangements that have been implemented or are being developed in the UK. For example, the UK's Social Exclusion Action Plan (Cabinet Office 2006) recommends data sharing between agencies as key to identifying **'**people experiencing, or at risk of, severe social exclusion.' The plan proposes extending existing practices such as the joining up of multi agency data related to children's services and youth justice to identify children and youth at risk.[4] Bellamy et al. (2005) describe how data sharing and crime audits undertaken by multi-agency crime and disorder partnerships (CDRPs) are being used to identify risky populations such as adult offenders while record matching between social security, national insurance, tax credit and personal taxation databases are being used to identify 'benefit thieves.' Data sharing arrangements also include commercial transactional data: 'The Department of Works and Pensions provide a data-matching service for local authorities: it routinely matches housing benefit records with data in social security, national insurance and tax systems. Customer data from gas, electricity and telephone companies are used, too, to identify properties that may be the subject of fraudulent claims' (Bellamy, 6, and Raab 2005, 401).

These examples of data sharing involve matching data doubles (or individual level data records) across different government agencies to produce new data doubles. Patterns, inconsistencies, or contradictions in transactional categories of the new data doubles can then be identified (e.g., matching data in housing benefit and student award claims). That is, relevant or expected correlations between certain transactional categories are defined a priori and used to identify subjects who are 'benefit thieves.' The main objective of data matching is thus the detection of the individual fraudster. However, out of this analysis a population profile can be constructed based on additional correlations discovered amongst the individual data doubles (e.g., age, number of dependents and multiple claims). The profile only comes into being through the matching of transactional categories in joined up databases (e.g., multiple benefits claims such as housing, welfare and student awards). The profile of a population called 'benefit thief' can then be used to identify new members based on their fit with the profile.

In this example of 'benefit thieves' data matching is largely conducted 'offline' due to the lack of integration between existing computer systems (Bellamy, 6, and Raab 2005). However, it effectively illustrates the logic underpinning joined up transactional databases: identifying correlations in transactional categories and constructing population profiles. Instead of the individual case-based matching of data doubles organised in relation to specific policy objectives (e.g., benefits fraud), standardised biographical and joined up transactional categories can be analysed to 'discover' correlations hitherto unknown or anticipated. That is, individuals are first sorted into categories and then categories are sorted into different population profiles. Using new software analytics such as data mining—'a procedure by which large databases are mined by means of algorithms for patterns of correlations between data, without establishing causes or reasons' (Hildebrandt 2008, 18)—different joined up

datasets can be analysed to reveal patterns. Populations are not known a priori but instead emerge from the process of data mining. For this reason specialists refer to data mining as Knowledge Discovery in Databases (Gandy 2007).

The patterns and correlations between categories discovered by data mining can be used to make up population profiles (Hildebrandt 2008).[5] Members of a population so identified do not necessarily share all of the correlated categories of a profile and thus Hildebrandt argues that a profile is 'non-distributive.' This means that profiles are probabilistic and cannot be applied to all members of a population without qualification:

> …they basically describe the chance that a certain correlation will occur in the future, on the basis of its occurrence in the past. … the correlation does not imply a causal or motivational relationship between the correlated data, they merely indicate the fact that the occurrence of one will probably coincide with the occurrence of the other
>
> (Hildebrandt 2008, 21-22)

Therefore, profiles are predictions based on past behaviour and data mining is an inductive knowledge: 'the correlations stand for a probability that things will turn out the same in the future' (Hildebrandt 2008, 18). One of the most prominent and criticised applications of profiling and data mining are the so-called intelligence-led approaches to combat terrorism. Law enforcement and intelligence agencies 'connect the dots' in government and commercial databases using techniques such as profiling, data mining, social network analysis, risk analysis and other predictive technologies (Amoore 2006). For example, the U.S. Department of Homeland Security's US VISIT programme identifies suspect populations or 'risky groups' through technologies that categorise people into degrees of riskiness. It does this by integrating existing databases, from police authorities, to health, financial and travel records. Of course, the discovery of correlations and their use to predict behaviour is not new. However, analyses have typically been confined to separate and discrete databases, based on hypothesised causal relationships between variables (Savage 2009), and have involved analyzing primary survey data and large data sets based on fairly routine statistical procedures (Uprichard, Burrows, and Byrne 2008). The difference that population metrics make is the application of methodologies based on data mining, business intelligence, web analytics, online analytical processing and text mining procedures. Uprichard et al. argue that a main driver of this transformation has been the increased digitization of data, which has transformed the most widely known and used statistical software package used by sociologists—the Statistical Package for the Social Sciences (SPSS). In the late 1990s through various acquisitions SPSS changed from a tool for empirical social research to a corporate brand primarily concerned with 'predictive analytics.' This is the new logic or 'new face' of quantitative sociological research and Uprichard et al. (2008) argue that the inscription devices of SPSS are key catalysts shaping how quantitative sociological knowledge about populations is being constructed. Whilst quantitative sociology is still dominated by traditional methodologies, the new reality of large volumes of digitized transactional data in both the commercial and government sector is thus presenting a significant challenge to those methods and the future of empirical sociology (Savage and Burrows 2007).[6]

However, inscription devices such as predictive analytics are not simply driven by the availability of large volumes of digital data and the computational power of technologies but related to a complex set of social relations and styles of thought (Osborne and Rose 2008). They respond to a particular problem-space or problematisations of population. Arguably all of the studies, inquiries, tests, proposals and practices related to the development of population metrics reflect a transformation in what Osborne and Rose (Osborne and Rose 2008) call a 'collectivity of thought' about population. Collectivities of thought 'give a kind of purposiveness or focus to any inscriptional set-up, together with some agreement as to the problem space towards which these are to be directed' (553-554) and make coherent the

plethora of practices that make up a numbers regime. What then is the problem to which metrics are the solution?

Several authors have argued that the problem-space and collectivity of thought mobilizing identification techniques that make up metrics as neoliberal rationalities of governing. The increasing preoccupation in contemporary social policy with risk management and the prevention of harm is said to be driving data sharing in social welfare and the identification of 'benefits thieves' (Bellamy, 6, and Raab 2005). In criminology it is the 'new penology' (Feely and Simon 1992), 'actuarial justice' or 'risk-based penology' (O'Malley 1998). These rationalities are concerned with reducing risk through a range of risk assessment techniques, where 'targeting populations of offenders and potential offenders according to the threat they pose therefore comes to take precedence over traditional values of due process and individual rehabilitation' (Bellamy, 6, and Raab 2005, 396). In social welfare it means a shift from relieving individual need and fulfilling entitlements to reducing moral hazard and risk (Dean 1999).[7] In sum, 'risk-based approaches, using actuarial-type assessments of individuals, families and neighbourhoods, are now used in child protection, mental health, public protection against high-risk offenders and many other fields' (6, Raab, and Bellamy 2005, 117). Across government more generally the rationality is identified in an orientation to objectives of efficiency and economy and customer-oriented service delivery systems, whereby customers have changing and multiple needs that need to be tracked, joined up and assessed (Cabinet Office 2005). Similar rationalities underpin securitisation practices such as the profiling and screening of airline travellers to predict and identify 'risky bodies' and divide legitimate (business, leisure, etc.) from 'illegitimate' mobilities (terrorist, immigrant) (Amoore 2006; Lyon 2003). Using data mining techniques profiles of 'risky' travellers are created and used for targeted searching. Such applications have particularly advanced since the 2001 attacks on New York and Washington as a way to identify, track and predict future terrorist activity (Gandy 2007).

All of these are certainly important and there are probably more examples of neoliberal rationalities that could be identified. However, rationalities are responses to particular problematisations. The question still remains—what is the problem of population to which this regime is a solution? This problem-space is provisionally identified in the concluding section.

The politics of population metrics

The answer can be found in how population metrics model population. The comparison with censuses concluded that while censuses modelled populations as relatively fixed assemblages of biographies that changed slowly and could be captured in subjective identifications, metrics model populations as modulations of correlated biographies and conduct. This is the problem-space to which metrics are oriented: a nature of population understood as a complex assemblage of transactions, movements and conduct that modulates, changes and transmutes and must be tracked and measured on an ongoing basis. It is a conception of population that gives focus to the various identification techniques—from biometric ID cards and passports and population registers to joined-up transactional databases—and the different metrics that they contain.

The foregoing suggests that population metrics bear much resemblance also to what Savage (2009) has called 'descriptive' methods, which are becoming dominant in practical fields such as security, marketing, finance, medicine and government. These methods are not anchored in establishing causal relationships but in producing 'useful' information such as indicators and recordings of behaviour. The interest is in 'surfaces rather than depths' and of performance indicators, measurements and profiles. However, as Savage also notes, knowledge is only organized on the basis of predefined categories. Certainly this applies to all of these methods

as well as population metrics, which are not merely information, but pre-formatted categories into which biographies, bodies and transactions are sorted and organized. Biometric categories are not merely descriptions of the body, or just another type of personal information. Rather, identification is mediated by technologies that 'read' the body in a particular way (van der Ploeg 2003). All classifications of population metrics—biographical, biometric and transactional—can be understood this way, as mediated inscription devices where, for example, biographical categories are constructed through negotiations involving numerous actors and actants (Bowker and Star 1999; Ruppert 2008). Clearly, drawing a division between the descriptive and explanatory is problematic. In this light, population metrics are more aptly *prescriptions*—authoritative assignments of identification that seek not only to know but also make up populations.

In this regard, metrics can be understood as enacting a particular version of population that is becoming authoritative, will be deployed in various government programs, and used to support particular political projects. As such, metrics constitute a new politics of population. Rather than calling for yet more laws and regulations to 'protect' data, the knowledge and totalizing effects of identification techniques and the consequences for how we are known and governed need to be critically interrogated. We need to research and investigate how this referent object of biopolitics is being reconstituted and the epistemological and ontological consequences of making up populations with metrics.

---

[1] By September 2009 some 50,000 cards had been issued.

[2] The IPS refers to this as 'identity standards', the processes by which an individual or an organisation uses to establish or prove identity.

[3] For example, see the review of administrative data as a source of population statistics by Jones and Elias (2006). The review identifies legal, technological, resource, access and privacy as the main issues and barriers to greater use for research purposes.

[4] Another example is the Cabinet Office's (2005) recognition of data sharing as key to the evaluation of the progress towards performance-related service delivery targets or the compliance with equal opportunity targets.

[5] The relation between populations and social groups is a large topic that I do not address in this paper.

[6] The ESRC has identified the sharing and linking of administrative records as key to evidence-based social science research (Economic and Social Research Council 2008) and as part of its national strategy for data resources.

[7] The foregoing examples are summarized from (Bellamy, 6, and Raab 2005).

References

6, P., C. Raab, and C. Bellamy (2005) 'Joined-up Government and Privacy in the United Kingdom: Managing Tensions between Data Protection and Social Policy. Part I', *Public Administration*, Vol. 83, No. 1, pp. 111-133.

Amoore, L. (2006) 'Biometric Borders: Governing Mobilities in the War on Terror', *Political Geography*, Vol. 25, No., pp. 336-351.

Ball, K. (2005) 'Organisation, Surveillance and the Body: Towards a Politics of Resistance', *Organisation*, Vol. 12, No. 1, pp. 89-108.

Bellamy, C., P. 6, and C. Raab (2005) 'Joined-up Government and Privacy in the United Kingdom: Managing Tensions between Data Protection and Social Policy. Part I', *Public Administration*, Vol. 83, No. 2, pp. 393-415.

Bennett, C. J., and D. Lyon, eds. (2008) *Playing the Identity Card: Surveillance, Security and Identification in Global Perspective*. London and New York, Routledge.

Bigo, D., S. Carrera, E. Guild, and R. B. J. Walker (2007) *The Changing Landscape of European Liberty and Security: Mid-Term Report on the Results of the Challenge Project*, An Integrated Project Financed by the Sixth EU Framework Programme.

Bowker, G. C., and S. L. Star (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge, Massachusetts, The MIT Press.

Cabinet Office (2005) *Transformational Government: Enabled by Technology*. London, The Stationary Office Limited.

Cabinet Office (2006) *Reaching Out: An Action Plan for Social Exclusion*. London, The Stationary Office Ltd.

Cabinet Office (2008) *Data Handling Procedures in Government*. London: The Stationary Office Ltd.

Caplan, J., and J. Torpey (2001) 'Introduction', in J. Caplan and J. Torpey (editors) *Documenting Individual Identity: The Development of State Practices in the Modern World,* Princeton, NJ, Princeton University Press.

Dean, M. (1999) *Governmentality: Power and Rule in Modern Society*. London, Sage.

Desrosières, A. (1998) *The Politics of Large Numbers: A History of Statistical Reasoning*, Translated by C. Naish. Cambridge, Massachusetts/London, England, Harvard University Press.

Dillon, M., and L. Lubo-Guerrero (2008) 'Biopolitics of Security in the 21st Century: An Introduction', *Review of International Studies*, Vol. 34, No., pp. 265-292.

Economic and Social Research Council (2008) *ESRC Response to MOJ Data Sharing Consultation*.

Feeley, M., and J. Simon (1992) 'The New Penology: Notes on the Emerging Strategy of Corrections and Its Implications', *Criminology*, Vol. 30, No., pp. 449-474.

FIDIS Consortium (2007) *Review and Classification for a FIDIS Identity Management Model*, edited by J. Backhouse and B. Dyer, Future of Identity in the Information Society (FIDIS), An EU Framework Six Project.

Foucault, M. (1997) 'The Birth of Biopolitics', in P. Rabinow (editors) *Ethics: Subjectivity and Truth,* London, Penguin.

Gandy, O. H. (2007) 'Data Mining and Surveillance in the Post-9/11 Environment', in S. Hier and J. Greenberg (editors) *The Surveillance Studies Reader,* Maidenhead (UK), Open University Press.

Graham, S., and D. Wood (2003) 'Digitizing Surveillance: Categorization, Space, Inequality', *Critical Social Policy*, Vol. 23, No. 2, pp. 227-248.

Gutwirth, S. (2007) 'Biometrics between Opacity and Transparency', *Ann Ist Super Sanita*, Vol. 43, No. 1, pp. 61-65.

Haggerty, K. D., and R. V. Ericson (2000) 'The Surveillant Assemblage', *British Journal of Sociology*, Vol. 51, No. 4, pp.

Hier, S. P. (2003) 'Probing the Surveillant Assemblage: On the Dialectics of Surveillance Practices as Processes of Social Control', *Surveillance and Society*, Vol. 1, No. 3, pp. 399-411.

Hier, S. P., and J. Greenberg, eds. (2007) *The Surveillance Studies Reader*. Maidenhead (UK), Open University Press.

Higgs, E. (2004) *The Information State in England: The Central Collection of Information on Citizens since 1500*. Hampshire, England, Palgrave MacMillan.

Hildebrandt, M. (2008) 'Defining Profiling: A New Type of Knowledge?', in M. Hildebrandt and S. Gutwirth (editors) *Profiling the European Citizen: Cross-Disciplinary Perspectives,* Belgium, Springer Science.

Home Affairs Committee (2008) *A Surveillance Society? Fifth Report of Session 2007-08*. London, The Stationary Office Ltd.

Home Office (2008) *National Identity Scheme Delivery Plan 2008*. London, The Stationary Office Limited.

Home Office (2009) *Safeguarding Identity*. London, Identity and Passport Service.

Jones, P., and P. Elias (2006) *Administrative Data as a Research Resource: A Selected Audit.* London, Economic and Social Research Council.

Judson, D. H. (2007) 'Information Integration for Constructing Social Statistics: History, Theory and Ideas Towards a Research Programme', *Journal of the Royal Statistical Society*, Vol. 170, No. 2, pp. 483-501.

Latour, B. (1986) 'Visualization and Cognition: Thinking with Eyes and Hands', *Knowledge and society: studies in the sociology of science past and present*, Vol. 6, No. 1, pp.

Lodge, J. (2007) *Are You Who You Say You Are? The EU and Biometric Borders*. Oisterwijk, Wolf Legal Publishers.

Lyon, D. (2003) *Surveillance as Social Sorting: Privacy, Risk and Digital Documentation*. London and New York, Routledge.

Martin, D. (2006) 'Last of the Censuses? The Future of Small Area Population Data', *Transactions of the Institute of British Geographers*, Vol. 31, No. 1, pp. 6-18.

Marx, G. T. (2007) 'Hey Buddy Can You Spare a DNA? New Surveillance Technologies and the Growth of Mandatory Volunteerism in Collecting Personal Information', *Ann Ist Super Sanita*, Vol. 43, No. 1, pp. 12-19.

Mathiesen, T. (1997) 'The Viewer Society: Michel Foucault's 'Panopticon' Revisited', *Theoretical Criminology*, Vol. 1, No. 2, pp. 215-234.

Mordini, E., and C. Petrini (2007) 'Ethical and Social Implications of Biometric Identification Technology', *Ann Ist Super Sanita*, Vol. 43, No. 1, pp. 5-11.

O'Malley, P. (1998) 'Introduction', in P. O'Malley (editors) *Crime and the Risk Society,* Aldershot, England, Ashgate.

Office for National Statistics (2003) *Proposals for an Integrated Population Statistics System*.

Office for National Statistics (2004) *The 2011 Census: A Design for England and Wales*.

Office for National Statistics (2005a) *Address Register Development*, edited by Advisory Group Paper Clip (05) 03.

Office for National Statistics (2005b) *Citizen Information Project: Better Sharing of Citizen Contact Information across the Public Sector*. London, HMSO.

Osborne, T., and N. Rose (2008) 'Populating Sociology: Carr-Saunders and the Problem of Population', *Sociological Review*, Vol. 56, No. 4, pp. 552-578.

Ruppert, E. (2008) "I Is; Therefore I Am': The Census as Narrative and Practice of Double Identification', *Sociological Research Online*, Vol. 13, No. 4, pp.

Savage, M. (2009) 'Contemporary Sociology and the Challenge of Descriptive Assemblage', *European Journal of Social Theory*, Vol. 12, No. 1, pp. 155-174.

Savage, M., and R. Burrows (2007) 'The Coming Crisis of Empirical Sociology', *Sociology*, Vol. 41, No. 5, pp. 885-899.

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

Treasury Committee (2002) *The 2001 Census of England and Wales: First Report of Session 2001-02*. London, The Stationary Office Limited.

Treasury Committee (2008) *Counting the Population: Eleventh Report of Session 2007-08*. London, The Stationary Office Limited.

United Nations Economic Commission for Europe (2007) *Register-Based Statistics in the Nordic Countries: Review of Best Practices on Population and Social Statistics*, United Nations, New York and Geneva.

Uprichard, E., R. Burrows, and D. Byrne (2008) 'SPSS as an 'Inscription Device': From Causality to Description?', *Sociological Review*, Vol. 56, No. 4, pp. 606-622.

Van Der Ploeg, I. (2003) 'Biometrics and the Body as Information', in D. Lyon (editors) *Surveillance as Social Sorting: Privacy, Risk and Digital Documentation,* London and New York, Routledge.