Big Data Methods for Social Science and Policy
Big data is expanding in its contribution across the social sciences and in public policy. While there are many practical, technical and ethical questions associated with this trend, this workshop focused on the practical application across the social sciences. The workshop was organised by the Cambridge University strategic research initiatives in Big Data and Public Policy and brought together over sixty researchers from a range of departments and disciplines across Cambridge to present and discuss big data applications in social science research and public policy.
Summary

Workshop participants learnt about new data sources (such as mobile sensors; location and behavior data; Twitter feeds; satellite mapping; text mining); how new policy models are being constructed to include new data sources; the challenges of handling new data sources (real time, adaptive, behavioural data, whole population rather than sample data; social data; fine grained data; ‘tall and fat’ data; ‘messy’ data; multi-lingual data) and how best to integrate new data with existing hypotheses and techniques that have been developed to address public policy questions.

Presenting to a diverse research audience allowed for interesting questions and shared strategies to emerge; for example, the challenges of data collection and analysis involving large and ‘new’ data sets; storing, accessing and publishing large data sets; generating new hypotheses and new forms of data; and the new skills that are required of researchers to create new knowledge and applications with big data and to engage with policy makers.

The range of presentations showed the diversity of Cambridge social science. Over twenty different departments or research centres were represented in the presentations and delegates. Researchers came from education, computer science, geography, psychology, engineering, law, linguistics, public health and sociology amongst others.

The types of data that researchers are using includes satellite data on shipping and location; historical census data; on-line media data for language processing; social data such as Twitter feeds, Facebook likes and sensor data in mobile phones themselves; as well as government administrative data sets addressing a range of subjects.

The research questions that were addressed through the day show the huge range of public policy questions that Big data may have relevance to. Do police on the street make a difference? Can shipping data be used to detect piracy and illegal drug imports? Can you identify power theft in developing countries? Can we infer changes in mood from the way that people move? Will workplace technology mean the end of work? What are the levels of corruption in government procurement?

Big data presents challenges for social science researchers, not just in the size of data sets being generated, although this is an issue, but also in the changing nature of data and its responsiveness offering real-time results for policy makers that requires a rapid, adaptive policy in return. Big data is often a rich data, offering refined data points and high quality observations that span different levels of analysis, for example, moving from the individual to the collective. Data is often fragmented, so researchers spend time trying to locate and access diverse data sets. The data requires translation – between languages, and also between disciplines. There may be missing data that causes concern for researchers, or ‘blind spots’ in data where significant sources are ignored in the research because they are not represented in the data sets. And social science researchers that connect with public policy and decision-making are well aware of the problem that measuring is not a passive act. The application of big data represents significant ethical as well as practical challenges, not least the growing predominance of data driven policy and the effect of measurement on behavior in the public sphere.

We also learned about the challenge for researchers in harnessing the potential of data; for example, broadening the scope and scale of research questions and data gathering. Several studies outlined how big data was being used to supplement traditional data sources, enhancing what is already known, offering what one presenter described as “power steering for the mind”. How best to capitalize on what big data has to offer social science and public policy was a recurring theme of the day.
Introduction

Reflecting on the challenges of big data and social science, a recent paper by Gonzalez-Bailon sums up the thinking behind the workshop today:

[The first challenge] is that social scientists can no longer do research on their own: the scale of the data that we can now analyze, and the methods required to analyze them, can only be developed by pooling expertise with colleagues from other disciplines. Social science has the theoretical tradition to build a context for these data, point to the right mechanisms of the dynamics analyzed, and build credible interpretations – which is as important as having access to vast amounts of information and cutting-edge methods. This collaboration requires working on the grounds of a common language, which in turn demands making the analytical toolkit of social scientists compatible with that of other disciplines.

What does big data mean for social science research? Several presentations helped define the nature of big data in social science research. We learned that ‘big’ does not just refer to the size of data sets, although that is a feature when research is being conducted in relation to whole population rather than population samples. Big data also refers to the changing nature of data, providing data at a higher level of detail in refinement and quality of observations. Big data means high spatial and temporal resolution, an overall ‘richer’ data set, spanning a different level of analysis (for example, linking from the individual to the collective) and also data that changes in real time. In his presentation, Dr Liang Wang from the Cambridge Computer Laboratory referred to big data being both ‘tall and fat’; that is, high dimensional data in large volume.

We also learned about the span of social science research at Cambridge. Speakers throughout the day represented the following university departments: Education, Computer Laboratory, Land Economy, Geography, Psychology, Judge Business School, Psychometrics Centre, Engineering, Sociology, Economics, Politics and International Studies, Law, CRASSH and Theoretical and Applied Linguistics. Participants were represented in these and the following departments: MRC Biostatistics Unit, Music Faculty, Sensors Network, Institute of Public Health, Behavioural and Clinical Neuroscience, Pure Maths and Mathematical Statistics, Criminology, Institute for Manufacturing and School of Biological Sciences.

We heard that “social science data is big data”. Social science research has a history of using large scale models, such as global social models to map traffic flows, land use, development and shipping. What is referred to as social data is much bigger than Facebook data – it relates to personal data, population data, spatial data, financial, intellectual, political, psychological, behavioural – and to date, the linkages between these data are largely unknown. Fragmentation within these data sets is a problem for researchers finding out where data is located – in censuses, companies, NGOs and households. Developing global social models that take account of this data for social science research is a significant challenge. But why do we need global, social models?
Many of the most significant policy challenges are global; such as food security, warfare, disease and pandemics and climate change. We can only address these if we expand existing global models to include the social. Most current global models, referred to as ‘earth systems models’ do not include people and their effects on global systems.

Some of the presentations referred to the particular challenges of researching big data. In some research areas (for example, using census data, company data, NGO data, household data) data may be fragmented and locating data is a challenge. In terms of social data (for example, using Facebook and Twitter data) there is disconnect between researcher and data input. Social data also has ‘blind spots’ where certain groups are not represented. This might include countries whose governments restrict access to on-line resources and social media, certain age groups who are not well represented on social media, non-English speaking populations, ‘off-line’ communities, criminal groups and during socially unstable situations. Data sets may themselves contain missing data such as in historical data sets and in missing countrywide administrative data sets.

Other challenges refer to the changing nature of big data. For example, media data such as news reports for mapping disease outbreaks can offer real-time results for policy makers but do policy makers have the policy frameworks to be able to respond? Presentations talked about the need for ‘adaptive policy models’ that respond to new information as it is produced. Here, the challenge for research and policy making is to harness the potential offered by (real-time) data. In other areas, the size of data sets presents a challenge for researchers and policy makers in logistics, storage and access. We heard from the ESRC supported Administrative Data Research Network set up to facilitate access for researchers to restricted administrative data sets.

Participants referred to the challenges of storing and managing large data sets and negotiating the appropriate sharing and release of data sets. There was consensus that data should be made as open as possible.

In a recent study published in Nature and reported in The Economist, researchers sought answers to the question of whether dark-skinned footballers were given more red cards than other players? They gave the same data to several different research groups to analyse. Although overall consensus was achieved in the study, the authors found results varied across research groups. Using different models and making different decisions about which variables to include in the analysis explained why research teams came up with different results.

This problem has bearing on the contribution of big data to social science research and public policy. There are many more factors involved in linking research to policy than simply having access to a wider range of large data sets and new data points. Discussions showed the complex models and research methods that are used to analyze and interpret the data and apply it to public policy problems. Researchers are aware of the responsibility to be able to explain and defend the choice of models and variables that are applied in order to address policy questions. The Economist article reflected that when science is informing government decisions, it would be prudent to ask several different research groups to analyze data and then compare the results. Whilst that may not always be possible – not least because of the cost implications of having different research teams analyzing the same research problem – there are ways in which researchers can enhance the validity and generalizability of their research with big data and these were discussed through the presentations.

2. A further workshop exploring the question of significance testing in relation to big data research is being considered.
3. Centre for Research in Arts, Social Science and Humanities
4. (http://www.adrn.ac.uk/),
5. October 10th 2015 p. 82
People and Place – Location and Geospatial Data

Cecilia Mascolo presented on research combining mobile sensing data with users’ perception of mood in order to better understand mood and behaviour. The research group uses phone sensors in a variety of research projects, often combined with other data sets to better understand a topic such as the effect of the Olympic Games on businesses around the Olympic Park, to psychology studies of mood and emotion. In this project, sensor data tracking behavior is combined with a mobile phone app in which users monitor their own perception of mood. Building on more traditional psychological questionnaires relating to mood, the research is able to create more fine-grained data about mood and behaviour over a period of time. After signing up to the app, users are sent a notification at different times of the day asking them about their mood. Users are asked to input their mood on an ‘emotion grid’, which plots feelings (negative to positive) and activity levels. The data offers therapeutic potential for users in identifying times at which they are most stressed.

Steve Marsh described how geospatial analysis is used to address the question of whether police on the streets make a difference, and where crime hotspots are located. Data sources include incident reports, Twitter feeds, data on police movements and mobile data. Data handling issues in this research include scaling, throughput (involving changing data) and responsiveness (when real-time results are offered). Working with big data means comparing live data with historical data. Steve outlined the potential of geospatial and geo-temporal analysis; for example using shipping data to detect piracy and preventing large-scale illegal drug imports. Geospatial analysis is applied commercially via start-up ‘GeoSpock’ which is investigating delivery optimization through the scheduling of deliveries in order to reduce congestion and accidents.

Elisabete Silva described work carried out by LISA (Lab of Interdisciplinary Spatial Analysis) at Cambridge, which researches using spatial analysis in urban planning models. Dr Silva reflected on the development of large-scale urban planning models, most notably the limitation of traditional ‘rational’ models in favour of an understanding of complexity and adaptation in soft artificial intelligence urban planning models. Systems models, underpinned by complexity theory, are adaptive and can better understand so-called ‘wicked’ policy problems in urban planning. In terms of planning and computation, Elisabete presented arguments in favour of more adaptive policy making, where models are able to take account of past and future expectations rather than a single ‘snap-shot’. Research at the LISA lab uses spatial analysis in research using satellite data in China to track workers and firms as development occurs and this data is used to model the movements in order to understand which firms might be more sensitive to policy changes.

In Mike Bithell’s presentation, complex, process-based, global ‘earth system models’ are used to model complex policy problems such as climate change, urbanization, deforestation, disease, conflict, water, and energy use. Whilst constructing these models in social science has always been challenging, the opportunities from new sources of social data - such as personal data, population data, special data, placial data, financial, intellectual, political, psychological and behavioural – with which to populate such models, present considerable challenges. The linkages between these new data are currently largely unknown; data is fragmented; deciding what new data should be included; how to calibrate and validate models with new types of data such as psychological and behavioural data. Researchers in this field have to consider how complex the models should be when modelling social systems that are themselves composed of many interacting individuals; are heterogeneous; spatially distributed; and have complex dynamics, amongst other features. Researchers also face challenges in communicating the results to policy makers who look for reliable advice, when models are not well understood and incomplete.
David Stillwell’s research at the University’s Psychometric Centre uses Facebook data within a specially constructed app, ‘myPersonality’ to present psychometric tests online and invite Facebook users to participate in tests and share data to be used by academics. To date, over 6 million Facebook users have opted to share their data in this way. The tests use Facebook ‘likes’ (with an average of 225 likes on a Facebook page) to predict individual characteristics. Study conclusions have found behaviour-based personality can be a better predictor than language based (self-reported) questionnaires.

Joseph Chancellor reflected on the dichotomy between ‘data driven’ and ‘theory driven’ social science research, which has been associated with the development of big data. Theory-driven research involves small but structured data, includes only what the researcher wants to measure, is slow and methodical, has low predictive accuracy, narrow focus and is highly interpretable. In contrast, data-driven research, which is favoured by industry, involves unlimited, unstructured data, measures everything (except what you actually want), is fast, aims for the highest possible accuracy, has a broader scope but limited interpretability. In research which uses online behavior data in psychology research, for example in researching personality associated with narcissism, a hybrid approach is sought which combines theory-driven and data-driven methods by a process of sampling, which is then used in a model to make predictions for data gathering involving large data sets.

Rui Sun’s presentation focused on the question: does kindness pay for everyone? The research uses big data to test correlation between different theories, which address the issue of social gains from altruism. The research suggests that acting kindly promotes social capital gains, but only for the poor. In an early study, the research found that donating amongst Americans was positively related to having more Facebook friends, but only among people with low income. In a second study, in an online getting-to-know-you chat paradigm, people tended to enjoy interacting with partners they perceived as prosocial than non-prosocial, but only if they believed the partner to have low income. In a third study, the findings were expanded to the global level, demonstrating that having a higher percentage of the population donating is positively related to number of friendships on Facebook, but only for low GDP per capita nations.

Kai Ruggeri’s research group at the University’s Department of Engineering represent a growing research trend using social media for data collection to explore social science topics. Here social media data is used to inform global health, which is a policy area heavily affected by the blurring of national borders and rapid globalisation. Consequently, more studies involve diluted samples of participants around the world who have access to social media sites where surveys are promoted. To test what implications this may have on using social media to elicit data intended to inform global health policy, the group used a variety of sites to generate survey data. The focus of study was decision-making regarding medical travel, given that patients are increasingly crossing borders in order to receive care. The pilot involved data from over eight hundred participants in forty countries. The researchers found that the nuances in the data were relevant to policy makers, and raised broader questions in the use of data sets. For example, research found significantly uneven distribution of representation of nationalities that was not simply resolved using propensity weights, resulting in the minimising of highly relevant factors of minority groups and choices. So, small but important factors and outcomes could be missed due to large majority factors. This was particularly the case for medical choices between age groups, for individuals from regions highly affected by conflict, and also by under-represented countries in the data whose patterns did not fit those of highly represented countries, which also presented a proxy for national economic and health service standards.
Government and Census Data – Linkage, Search and Analysis

Andrew Means from the Impact Lab, a consultancy which conducts research projects in the social sector, took delegates through some of their projects which use data analytics. Means questioned why data analytics had largely been ignored in the social sector when it had so far transformed sectors such as: transport, entertainment, advertising, on-line shopping, and politics. Means proposed that a cultural gap existed between data scientists and practitioners in the social sector, where data is largely used as output, for reporting and evaluation purposes. Means described case studies that the Impact Lab is working on to explore the potential of data analytics being used as an input to work with policy makers. For example, using data analytics to identify housing stock at risk of lead poisoning in Chicago in place of testing individual school children, predicting potential school drop-out so that intervention can take place earlier in a child’s schooling and targeting ‘at-risk’ mothers so that services can then be offered to them.

Mihály Fazekas explored the challenges of working with administrative data sets within the Digiwhist research programme which is using government procurement data to research public corruption across European countries in order to assist public bodies in reducing the extent of corruption and its negative effects. The programme identifies missing data, such as the percentage of public procurement tenders that are advertised, in order to investigate corruption potential as well as ‘red flags’ in the data such as single tenders. Three specific examples using data to explore corruption were offered through an analysis of: lack of electoral competition in English local councils, mafia infiltration in Italy and the role of the European Court of Justice as an anticorruption actor. The research is being used to lobby governments to better manage administrative data and to build risk assessment software for public servants.

Miguel Morin presented on research using historical US census data to explore the question of whether developments in workplace technology will lead to the end of work. Current concern over technological developments replacing individuals in the work place are explored by examining the effects of electrification in the United States in the early twentieth century. Using US census data the research compares workers in electrifying counties to those in static counties. Some of the research challenges involve identifying individuals and tracking them over time. Studying changes in occupation reveals that people displaced by electrification were subsequently up-skilled and received increased earnings by the next ten-year census. The results are counter-intuitive to the predicted effect including the impact of the great depression. The results show technology benefits in the long-run although there are short-term costs of recession. Results indicate correlation and further research is looking to explore causal effects, for example the effect of geographical terrain, and the development of hydro-electric power.

Kakia Chatsiou and Aidan Condron presented from the ESRC -funded Big Data Network. In Phase 1 the Administrative Data Research Network (ADRN), an ESRC supported network offers procedures, process and training in the use of big data in research. In Phase 2, Business and Local Government, the work is sponsored by the Big Data Network Support (BDNS) team at the UK Data Archive. One of the main functions of the network is to enable researchers to gain access to administrative data for the purposes of carrying out social and economic research. On a case-by-case basis the network negotiates access for researchers, working with government departments, facilitating the linking of data sets so that anonymity of individuals is preserved, while specific information from individual data collections can be brought together for the purposes of research and analysis. In this way the network aims to benefit society by enabling research to be conducted on databases that are not generally designed for research and statistical use.

Building an open source data platform and browser tools that could enable researchers to securely link data sets were discussed as important procedures and tools for social scientists for the future. The ADRN and BDNS groups are keen to engage with researchers who might be able to contribute to these developments. Workshop participants were keen to find out how anonymity was managed, about linking data sets for research purposes, and practical issues involved in working with large data sets for social science research including preserving data sets for future use.
New Tools and Methods

Gabriel Recchia’s presentation took delegates back to the work of statistics quantifying the degree of association between two words in a text. According to Recchia, given the current excitement over complex and computationally intensive machine learning techniques such as deep learning and conditional random fields, it may seem unlikely that much useful information could be obtained simply computing normalized counts of the number of times that pairs of words co-occur (within a sentence or a lexical window of a particular size, for example). However, co-occurrence based methods have proven surprisingly useful in a wide variety of tasks in natural language processing. Furthermore, their simplicity and transparency affords them certain advantages in particular contexts. Recchia described studies that demonstrate the surprising success of co-occurrence-based methods in a variety of circumstances: predicting the degree to which two words are similar in meaning, predicting the excavation sites of archaeological artifacts, predicting grammatical classes of words in an unsupervised fashion, and others. These models have been applied in disaster relief and in epidemiological modelling.

Liang Wang’s presentation described developments in techniques for searching in high-dimensional data. The Internet is overloading its users with excessive information flows, so that effective content-based filtering becomes crucial in improving user experience and work efficiency. Latent semantic analysis has long been demonstrated as a promising information retrieval technique to search for relevant articles from large text corpora. Researchers have built Kvasir, a semantic recommendation system, on top of latent semantic analysis and other state-of-the-art technologies to seamlessly integrate an automated and proactive content provision service into web browsing. The research utilises the processing power of Apache Spark to scale up Kvasir into a practical Internet service. Researchers have identified both algorithmic and engineering solutions in Kvasir to tackle the scalability challenges by finding a good tradeoff between the accuracy and efficiency in searching within large and high dimensional datasets.

Accurate and timely collection of facts from a range of sources is crucial for supporting the work of experts in detecting and understanding highly complex diseases. In the final talk of the workshop Nigel Collier illustrated examples from the group at the Language Technology Laboratory who are working on Natural Language Processing (NLP) on large-scale Web and linked data. First, in the BioCaster project, working with global public health colleagues, high throughput text mining on multilingual news was employed to detect and map infectious disease outbreaks in near-real time on a global scale. Second, in the SIPHS project, researchers are beginning to explore techniques for encoding personal health reports from a variety of social media sources to support real-time knowledge discovery about infectious diseases and adverse drug reactions. In the LION project the group plan to explore how text mining could support literature-based discovery in cancer biology. The aim is to develop a tool, which cancer biologists can use to test and generate novel research hypotheses on the basis of knowledge already published in scientific literature.
Conclusions

The relevance and contribution of social science researchers to big data studies was emphasized; as one contributor put it, “social science data is big data” because social science already constructs complex models to try and map and understand the world and our place in it – how we work, how we move, how we interact as people.

Social science models are now global social models that can map traffic flows, land use, but there are many times of social science data – spatial, population, social, behavioural, financial, political – developing models that attempt to understand these data is important because many of the most significant policy challenges and developing models are global problems such as food security, conflict, disease, pandemics and climate change. Social science provides important context, and theoretical insight to explain and understand big data and new insight to enduring social problems will most likely come through collaboration between researchers linking innovative method and data gathering with established knowledge and understanding of the research problem.

To facilitate fruitful collaboration and make the most of the potential that big data offers for social science research and public policy, the workshop highlighted the following challenges for social science researchers:

- The challenge of big data causes us to reflect on what are social science research questions: the data do not ‘speak for themselves’.
- The importance of multi-method studies, an established tradition in social science – in order to ‘triangulate’ and verify findings.
- Make data as open as possible – this leads to improvements in the data; more studies carried out enables replication and verification of findings.
- Respond to data generating new research questions and models – be open to new learning from data.
- Social scientists and policy makers may not need to become computer scientists but do need to develop some computational skills in order to understand data and the decisions that are being taken with data.
List of Attendees

Anne Alexander | Centre for Arts, Social Sciences and Humanities (CRASSH)
Kathryn Berger | Veterinary Medicine
Mike Bithell | Department of Geography
Alan Blackwell | Computer Laboratory
Katherine Boyle | Department of Archaeology and Anthropology
Brendan Burchell | Department of Sociology
Sally Chan | Department of Psychology
Joseph Chancellor | Department of Psychology
Kevin Chang | Department of Engineering
Kakia Chatsiou | Administrative Data Research Network, University of Essex
Vasiliki Chatzi | Kings College London
Nigel Collier | Department of Theoretical and Applied Linguistics
Aidan Condon | University of Essex
Max Conway | Computer Laboratory
Graham Copekoga | Cambridge Public Policy Strategic Research Initiative
Elizabeth Cory-Pearce | The Tavistock Institute of Human Relations, St Edmunds College
Adam Coutts | Department of Politics and International Studies (POLIS)
Jon Crowcroft | Computer Laboratory
Simon Deakin | Faculty of Law and Cambridge Public Policy Strategic Research Initiative
Tanvi Desai | Administrative Data Research Network, University of Essex
Clare Dyer-Smith | Cambridge Big Data Strategic Research Initiative
Mihály Fazekas | Department of Sociology
Simon Frost | Veterinary Medicine
Lu Gao | MRC Biostatistics Unit
Satinder Gill | Computer Laboratory/Music Faculty
Charlotte Hamilton | Department of Land Economy
David Howarth | POLIS and Cambridge Public Policy Strategic Research Initiative
Pui Hung Hui | Department of Psychology
Maria Iacovou | Department of Sociology
Sonia Ilie | Faculty of Education
Nick Jones | Institute of Public Health
Hannah Jongsma | Behavioural and Clinical Neuroscience Unit
Sarah King-Hele | University of Manchester
Alex Kogan | Department of Psychology
Justin Koh | Department of Psychology
Maria Liakata | University of Warwick
Keren Limor-Waisberg | Cambridge Networks
Claudia Lopes | Centre for Governance and Human Rights
Wilson Lu | Land Economy
Below is a list of presentations from the Workshop, all are available to view online at:
www.bigdata.cam.ac.uk/events/events-archive/big-data-methods-for-social-science-and-policy

**People and place – Location and geospatial data**

Cecilia Mascolo | Computer Laboratory
*Mobile Sensing and Geo-Social data analysis for Social Science*

Steve Marsh | Computer Laboratory
*Real-world insights through geospatial analysis*

Elisabete Silva | Department of Land Economy
*Soft artificial intelligence, linking socio-economic and land spatial-led data analysis for urban planning*

Mike Bithell | Department of Geography
*Is social data big data? Challenges for global social models*

**Social media in social science and policy**

David Stillwell | Psychometrics Centre
*Predicting Psychology from Social Data*

Joseph Chancellor | Department of Psychology
*Combining data- and theory-driving insights using large, anonymous datasets of expressive online behavior*

Rui Sun | Department of Psychology
*Donations Predict Social Capital Gains for Low SES, But Not High SES Individuals and Countries*

Kai Ruggeri | Department of Engineering
*Using social media to inform global health policy: An example of major considerations regarding data for policy and policymakers*
Government and census data – linkage, search and analysis

Andrew Means | The Impact Lab
The Role of Prediction in the Targeting of Services

Mihály Fazekas | Department of Sociology
Exploring government administrative data to hold governments accountable in the big data era

Miguel Morin | Faculty of Economics
Adapting to Workplace Technological Change over the Long Run: Evidence from US Longitudinal Data

Tanvi Desai and Aidan Condron | ADRN
Facilitating access to Administrative and big data in the UK: the Administrative Data Research Network and the Big Data Network

New tools and methods

Gabriel Recchia | CRASSH
The Unreasonable Effectiveness of Co-occurrence Based Models

Liang Wang | Computer Laboratory
Needle in a Haystack: Understanding the Tradeoff between Accuracy and Efficiency in Searching of High Dimensional Big Data

Nigel Collier | Department of Theoretical and Applied Linguistics
Natural Language Processing for Digital disease detection in a fast-moving world
University of Cambridge Strategic Research Initiatives and Networks build on areas of existing research strength by bringing together a critical mass of expertise from across the University to address large-scale multi-disciplinary research challenges.

The Cambridge Public Policy Strategic Research Initiative (SRI) aims to support public policy research across Cambridge University, working with colleagues in science, social science, the arts and humanities, to apply new thinking to public policy problems and promote research and analysis into the public policy process.

The Cambridge Big Data Strategic Research Initiative brings together researchers from across the University to address challenges presented by our access to unprecedented volumes of data. Our research spans all six Schools of the University, from the underlying fundamentals in mathematics and computer science, to applications ranging from astronomy and bioinformatics, to medicine, social science and the humanities. In parallel, our research addresses important issues around law, ethics and economics, in order to apply big data to solve challenging problems for society.

For more information, find us at:
www.publicpolicy.cam.ac.uk  www.bigdata.cam.ac.uk
@CamPubPolSRI  @CamBigData

Contact the Public Policy SRI Coordinator:
Dr Charlotte Sausman
Research Coordinator, Cambridge Public Policy Strategic Research Initiative
Rm 219, Department of Politics and International Studies (POLIS),
Alison Richard Building, 7 West Road, Cambridge, CB3 9DT

tel: +44 (0)1223 334572
e-mail: cs738@cam.ac.uk