



RUSTAT CONFERENCES
JESUS COLLEGE, CAMBRIDGE

BIG DATA

Conference Report

Rustat Conference held on
Wednesday, 30 September 2015

Rustat Conference rapporteur: Eli Keren

www.Rustat.org
www.jesus.cam.ac.uk
info@Rustat.org

Rustat Conference on Big Data - Summary of Sessions and Discussions

30 September 2015, Jesus College, Cambridge

- Big Data is a relatively new player on the fields of both technology and law, and while there are a great many questions to be asked, not many can be definitively answered, partly because it's too soon and the space too young.
- Caution, uncertainty and legality were important features of the conversations. However, the closing discussions struck a more positive note with optimism for what Big Data is capable of delivering without necessarily compromising privacy:
 - Big Data has the potential to solve major problems in new ways e.g. spread of infectious diseases can be tracked from mobile phone data, as can the expansion of slums, allowing vital resources to be targeted where they are most needed. **Grace Cassy**
 - Young people have an entirely new outlook on privacy, actively wanting to be seen and for their data to be used to deliver targeted results. Big Data has great potential e.g. to revolutionise agritech, for home devices to be able to monitor their own usage stats and feedback to manufacturers. Smart Cities will apply Big Data to everyday issues e.g. light and traffic management providing huge cost savings. In areas where Big Data and IoT have the most potential, privacy could well be a side-issue. **Chris Doran**

Session 1: Application of Big Data in Research

- Governmental administrative data, specifically educational and graduate earnings data, can be used to identify skill gaps in the labour market. This identification can then be fed back into the educational system to plug the gap with the next wave of graduates. A public debate is required on the use of governmental administrative data. There is a need for a clear and publicly agreed framework for data use that eradicates the need for a project-by-project approach to data ethics. – **Anna Vignoles**
- Established businesses with big data needs and young start ups born digital can work together to the benefit of both, but only if old businesses learn to trust data, and new businesses manage to find the right business model to stay afloat. – **Mohamed Zaki**
- 'In public health data, people who actually operate in the health service won't do things with data that would be beneficial both to individuals and collectives, because of ethical concerns, while private agencies who have the same held data will exploit it to optimise insurance rates. We're at a very dangerous point.' – **Jon Crowcroft**

Session 2: Big Data and the Internet of Things (IoT)

Wearable tech is opening up exciting new applications to the world of professional sport and the healthcare industry by tracking the motion and performance of a body at all times. Complicated sensors that produce vast seas of data are not the key, but contextual awareness of a much smaller data pool can open up much larger possibilities.

The Internet of Things (IOT) similarly has applications to many other aspects of education and business. However, well-defined boundaries must be implemented to protect the individual from 'optimisation' that could leave those who struggle behind. Currently the democratic system of values may not be well applied to Big Data.

- 'It's not determined that a person will drop out [of university], it's probable, so if we act on that probability, we're actually denying agency.' – **Sharath Srinivasan**
- 'Some of these things are progressing out of pace, and there are some really fundamental questions we just don't have answers to.' – **Geoff McGrath**
- My complaint largely about the big data situation at the moment is that if you wanted to summarise it in general you'd say 'Big Data is the solution, now what was that problem again?.. and that's nuts... Somewhere in this mess, there is something about human values.' – **John Naughton**

Session 3: Can we share data and respect individual rights without disrupting new business models?

Big Data must not forget the human at the centre. 'Databoxes' (private devices that collect and manage a person's data, allowing them to see and understand better which items of personal data were being requested or sent where, and by whom) could give individuals a much greater scope of agency of the use and distribution of their personal data, if effectively implemented.

Personal data mined from social media has shown fantastic results in the fields of targeting and personalisation. However, it's encroaching on territory that users are uncomfortable with and more investigation needs to be done before boundaries can be definitively set.

- 'I'm not a cog in a machine to be optimised, I think of myself as a person with some autonomy and agency in the system, and yet a lot of the uses of data in these kinds of context seem to be thinking very much about 'How can I be made more efficient?', rather than 'How can I be enabled to improve myself?'. – **Richard Mortier**
- 'What makes people comfortable or uncomfortable with the way their data is used? ... It's not 'What is the data?', it's 'How is the data being used, by whom and for what?' – **David Stillwell**

Session 4: Big Data and Data Protection

There exist poorly-understood 'levels of consent' within society, by which some data can only be used with explicitly given informed consent, and others can be taken by authorised bodies because society has deemed them worthy of it on a greater level, superseding the individual. The law, as it pertains to Big Data, is murky and in need of educated modernisation.

- 'I wouldn't want a system where the only people who *couldn't* read my email were GCHQ. It seems a bit perverse if Chinese agencies or any other country could, but not GCHQ because they stick to the rules.' – **Julian Huppert**

Closing Comments:

- 'There are some very interesting, exciting technologies out there that do offer a potentially more effective way of addressing some of these very big challenges.' – **Grace Cassy**
- 'Most of the interesting applications of big data are not to do with people... they're what happens when we put data out in the field... In many of the areas where big data and IoT have the most potential, privacy could well be a side-issue.' – **Chris Doran**



RUSTAT CONFERENCES
JESUS COLLEGE, CAMBRIDGE

BIG DATA Conference Report

Table of Contents

Summary of Conference Sessions and Discussions	2
Conference Agenda	4
Jesus College Cambridge Rustat Conferences Background and Acknowledgements	5
Jesus College Cambridge Rustat Conferences Membership	6
Conference Report	
Session 1	7
Session 2	8
Session 3	9
Session 4	10
Speaker and chairs profiles	12
Participants list	16



BIG DATA

Conference Agenda - Wednesday, 30 September 2015

09.30-09.40 - Welcome and Introduction, Upper Hall

Professor Ian White, Master, Jesus College, Cambridge, Chair, Rustat Conferences

09.45-10.45 - Cambridge Big Data: the Application of Big Data in Research

- Large Scale Administrative Data for Social Policy Research.
- Data Driven Business Models: what business models that rely on data really look like; how they create value.
- What are the big questions Big Data can address next in social sciences?

*Chair: Professor Anna Vignoles Professor of Education, University of Cambridge, Fellow, Jesus College
Dr Mohamed Zaki Cambridge Service Alliance, Institute for Manufacturing, University of Cambridge*

10.45-11.10 - Break: Gallery, Upper Hall

11.10-12.10 - Big Data and the Internet of Things: KPMG and McLaren Applied Technologies. Stress, well-being and the use of data to improve individual and team performance

- Employees are leading increasingly busy and complicated lives with multiple competing priorities across both professional and personal spaces. This has led to an epidemic of stress which may be impeding well-being and impacting the performance of individuals and teams in the workplace.
- Can we use data to make interventions at both micro (employee) and macro (organisational) levels that would allow individuals and employers to manage well-being more effectively with the aim of improving individual and team performance?

*Chair: Professor Peter Williamson Cambridge Judge Business School, Fellow, Jesus College, Cambridge
David Fairs Partner, KPMG
Geoff McGrath Chief Innovation Officer, McLaren Applied Technologies MAT*

12.15-13.30 - Lunch: Master's Lodge - entrance via Prioress's Room, Cloister Court

13.30-14.30 - Can we share data and respect individual rights without disrupting new business models?

- Human Data Interaction HDI: placing the human at the centre of the data system.
- The myPersonality Project and benefits of sharing data. Making users comfortable with how their data is used without disrupting new business models.

Chair: Professor Jon Crowcroft Marconi Professor of Communications, Computer Lab, University of Cambridge

Dr Richard Mortier University Lecturer, Computer Lab, University of Cambridge

Dr David Stillwell Lecturer in Big Data Analytics & Quantitative Social Science, Cambridge Judge Business School; Deputy Director, Psychometrics Centre, University of Cambridge

14.30-14.50 - Break: Gallery, Upper Hall

14.50-15.45 - Big Data and Data Protection.

- Gathering and use of personal data - access, privacy and security. Legal, ethical, societal issues of consent and non-consent. The Investigatory Powers Bill.

Chair: Dr David Erdos University Lecturer in Law and the Open Society, Centre for Intellectual Property & Information Law, Faculty of Law, Fellow, Trinity Hall, University of Cambridge

Dr Julian Huppert Politician, University Lecturer, Fellow, Clare College, Cambridge

15.45-16.00 - Closing Discussion and Comments

Chair: Professor John Naughton Senior Research Fellow at CRASSH, Emeritus Professor of the Public Understanding of Technology at the Open University and Vice President of Wolfson College, Cambridge

Grace Cassy Partner and Co-Founder, Epsilon Advisory Partners

Dr Chris Doran Technology Entrepreneur, ARM; Director of Studies in Physics, Sidney Sussex College, Cambridge

Conference Close



Rustat Conferences Jesus College Cambridge

The Rustat Conferences are an initiative of Jesus College, Cambridge, and chaired by Professor Ian White FREng, Master of Jesus College. The Rustat Conferences provide an opportunity for decision-makers from the frontlines of politics, the civil service, business, the professions, the media, and education to exchange views on the vital issues of the day with leading academics.

Founded in 2009, Rustat Conferences have covered a variety of themes including: *The Economic Crisis; The Future of Democracy; Cyber Security; Manufacturing in the UK; The Geopolitics of Oil and Energy; Drugs Policy; Organisational Change in the Economic Crisis; Cyber Finance; The Understanding and Misunderstanding of Risk; Food Security; Transport and Energy*, and *Inequality*. In addition to acting as a forum for the exchange of views on a range of major concerns, the conferences provide outreach to a wider professional, academic, student and alumni audience through the publication of reports. The conferences are named after Tobias Rustat (d.1694), a benefactor of Jesus College and the University.

Acknowledgements

We thank the Rustat Conferences Foundation Members for their generous support and the following for their advice and assistance with the Rustat Conference on Big Data: Dr Clare Dyer-Smith of Cambridge Big Data, Professor Anna Vignoles, Dr Julian Huppert, Dr James Dodd, Professor Peter Williamson, Professor Bernard Silverman and KPMG and McLaren.

Cambridge Big Data

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. 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, research addresses important issues around law, ethics and economics, in order to apply Big Data to solve challenging problems for society. Cambridge Big Data supports collaboration and knowledge transfer in this growing field. www.bigdata.cam.ac.uk

Chatham House Rule and Rustat Conference Report

Please note the conference is conducted under the Chatham House Rule. The conference report will however reveal the identity and affiliation of speakers and discussants unless they request otherwise. The procedure is as follows: the Rustat Conference rapporteur will circulate a draft version of the report and anyone mentioned or quoted in it may remove the attribution. Once this procedure is complete the report is published via the Rustat Conferences website.

Twitter #BigData #Rustat @JesusCollegeCam





Rustat Conferences Foundation Members

The Rustat Conferences are supported through a mix of sponsorship and a membership scheme that was launched in 2013-14 - details of this can be found at www.Rustat.org. We are very grateful to the Rustat Conferences Foundation Members for their generous support:

Dr James Dodd - James's career has concentrated on the financing and management of companies in the fields of telecommunications and technology. He serves on a number of boards in these areas and is active in supporting a number of academic projects and charities.

Harvey Nash is an executive recruitment and outsourcing group. Listed on the London Stock Exchange, and with offices across the world, it helps organisations recruit, source and manage the highly skilled talent they need to succeed in an increasingly competitive and innovation driven world.

KPMG is a global network of professional firms providing Audit, Tax and Advisory services. It has more than 155,000 outstanding professionals working together to deliver value in 155 countries worldwide.

Laing O'Rourke is a privately owned, international engineering enterprise with world-class capabilities spanning the entire client value chain. It operates an integrated business model comprising the full range of engineering, construction and asset management services, delivering single-source solutions for some of the world's most prestigious public and private organisations.

McLaren Racing Ltd has a reputation for efficiency and professionalism. Working within a fast-paced environment and to the highest standards, its highly skilled workforce operates primarily in the areas of manufacturing, engineering and race team as well as logistics and support.

Mr Andreas Naumann is a senior executive in the financial industry. Outside the professional sphere, he is keenly interested in subjects like urbanisation, youth unemployment, education and foreign policy. He supports the Rustat Conferences as a private individual.

Sandaire Investment Office - SandAire and Lord North Street came together in April 2014 to combine their businesses, both of which specialise in looking after the investment assets of very wealthy families, charities and endowments.

Maria and David Willetts



RUSTAT CONFERENCES
JESUS COLLEGE, CAMBRIDGE

BIG DATA

Wednesday, 30 September 2015

Session 1: The Application of Big Data in Research

Professor Anna Vignoles began with the fundamental question upon which all of the day's conversations must be based, the question that we must ask and re-ask ourselves regularly: *What is big data?* She stated that big data can refer to 'a very wide range of data – some are data that we've always had but now have on a larger scale... and some are genuinely novel,' but adds that the phrase has become ubiquitous to the point that it is 'almost meaningless'. Instead of tackling the entire realm of big data and attempting to define what big data is, Professor Vignoles focused on the kinds of data that are relevant to her own research: governmental administrative data, a data set that only the brave would argue against being labelled 'big'.

Professor Vignoles's research is geared towards utilising governmental administrative data to tackle the growing problem of skill shortages in the workplace. When workers with certain skills are scarce, some businesses find themselves paying vast wage premiums in order to secure the few workers who have the skills they need, while small businesses may not be able to afford them. Professor Vignoles and her team have been trying to address this problem by focusing on educational data. For all individuals who have entered the educational system since 1993, the government holds a complete data set tracking their progress, including further and higher education. This data has already been put to use to identify issues in the school system, but it is now possible for this data to be used to determine what happens to young people once they reach the labour market. Skills that are in short supply can be identified from these data sets by comparing earnings between individuals with different qualifications and observing which areas of expertise generally gave rise to the higher salaries.

Professor Vignoles took an aside at this point to assert that 'education is, of course, not simply about producing skills for the labour market. Education is far more than that, it is the ultimate endeavour for humankind.' However, she stresses that the need to understand the educational system and the affect it has on young graduates is not in contradiction with that endeavour.

Previous research that did not rely on big data has already given some insights and led to policy change. Having observed that businesses were paying high premiums for a minority of graduates with mathematical skills, more mathematics were added to social science curricula in order to fill the hole in the market some years down the line. With the help of big data on the entire population, there is even greater scope for understanding skills needs. Professor Vignoles pointed out that there are several barriers to implementing such research, most notably insufficient cross-departmental data sharing required to allow these methods to reach full fruition. Professor Vignoles concluded her talk by calling for a public debate on the use of governmental administrative data. She argued that there is a need for a clear and publicly agreed framework for data use that eradicates the need for a project-by-project approach to data ethics.

Dr Mohammed Zaki spoke about the increasing role of big data in business, specifically drawing a distinction between young 'digital born' companies and older, more established companies that rely on human experience and intuition to a much greater extent. There are twin challenges facing these two very different kinds of businesses: the barrier to allowing big data to optimise existing business models and the emergence of a flurry of new business models built around the incorporation of big data.

Relating to small businesses, Dr Zaki has identified several distinct small business models that centre themselves around the use of big data: (1) Free data collector and aggregator: companies collect data from vast numbers of different, mostly free, sources then filter it, enrich it and supply it to customers in the format they want; (2) Analytics-as-a-service: these are companies providing analytics, usually on data provided by their customers; (3) Data-aggregation-as-a-service: these companies aggregate data from multiple internal sources for their customers, then present it back to them through a range of user-friendly, often highly visual interfaces; (4) Free data knowledge discovery: the model here is to take freely available data and analyse it; (5) Multi-source data mash-up and analysis: these companies aggregate data provided by their customers with other external, mostly free data sources, and perform analytics on this data to enrich or benchmark customer data.

The advantage of these businesses being born digital is that they can supply data services to larger, more established companies that are struggling to modernise their systems. However, not all companies settle on the appropriate business model nor find their niche in the market; 10% of the small businesses investigated had ceased to operate a year on from Dr Zaki's original explorations.

Dr Zaki's explained there has been a shift in focus from products towards solutions. Customers no longer just want a vehicle, they want an optimised service route and live fuel management updates. This leads on to the issue of service-appropriate pricing: a company may install sensors in a product that feed back to the company and notify the user before a problem develops, but such technology would require data management on a very large and expensive scale while the customer may expect aftercare on the product to be free. It becomes clear that products are much simpler to price than solutions and on-going services that make use of big data.

The roundtable discussion highlighted the issue of privacy protection, particularly with regards to Professor Vignoles' research into the use of governmental administrative data, but Professor Vignoles reminded the conference that the data she was dealing with differed greatly from other mined sources such as social media data due to the fact that the governmental data was anonymised in accordance with the law. **Jon Crowcroft** disputed this distinction, arguing that there is 'no difference between Department of Education and facebook', stating that their use of public data is tantamount to the same thing and adding that the two should be treated with the same caution. **David Erdos** rebutted that the issue should be the *use* of the data, arguing that it appeared that the research would not affect identifiable individuals but would contribute to the public good, and it was not the purpose of European Data Protection to obstruct that. He argued that 'if you say European Data Protection means full informational self-determination, then a huge amount of that kind of data use would be shut down.'

Session 2: Big Data and the Internet of Things

Geoff McGrath, CIO of McLaren Applied Technologies, addressed the Internet of Things (IoT) and its application to human physical development. 'The Internet of Things is where the cyber world and the physical world converge,' he began, going on to explain the development of products that have the potential to change greatly the way in which we interact with technology. Embedded and adaptive intelligence can lead to products that sense their environment or status and adapt their behaviour or output accordingly, such as mountain bikes that adapt their suspension to the terrain.

This talk focused on the application of the IoT towards human performance. By using sensors to track aspects of human motion, McLaren developed a scheme by which they could use 'analytics of the body's condition to drive strategy and decision making' in the world of professional sports, wherein a relatively small physical injury could lead to staggering financial loss. By applying contextual awareness to their measurements, McLaren found that they could achieve maximum insight from a minimal amount of data. Rather than using a great array of sensors, they use a single cheap sensor for each human asset, coupled with deep analytics fuelled by machine learning techniques, to recognise the context of the measurements they're

making. For the first time, we are now able to identify exactly when to dial back the training to prevent injury before the early warning signs begin to develop.

Dr McGrath concluded by drawing our attention to the benefits of this technology outside of its birthplace in the world of professional sport. In patients recovering from disability, for example, sensors could check their movement for warning signs every time they moved rather than relying on a session with a doctor once in three months. He added, however, that the concepts offer a technological barrier: since the information generated cannot be aggregated but must remain personalised, the data generation has the potential to be staggeringly vast.

David Fairs expanded on this theme, arguing the virtues of applying the IoT to people in the form of similar wearable tech for more personal means. KPMG have applied big data techniques to identify students at university who are likely to drop out. Usually, these students cannot be identified until their situation has deteriorated too far, costing them a degree and the university a significant sum of money, but big data could enable these students to be identified well in advance of their situation deteriorating, allowing for intervention and extra support where needed. He spoke also of applications to recruitment: while it can currently take up to a full year to fully understand whether or not a new recruit suits their job, big data analysis and wearable tech could cut this time down considerably. The applications to business are clear and wide-ranging if the technology lives up to its hype.

These two talks led to a debate on the nature of ownership of personal information sourced from wearable tech, and whether or not employees should be under obligation to share this information with their employers. Boundaries proved to be a recurring theme. **Sharath Srinivasan** argued that it is one thing for elite sportsmen to be given bonuses for engaging in activity in their private lives that would benefit their careers, but beyond such cases, and where power inequalities can be far greater, boundaries must be placed on who has the right to decide the 'optimum' life. Similarly, with regards to the case of reducing university drop-out rates, **David Fairs** added that this technology could be used to identify drop-outs before they even join university; however, that could create socio-economic exclusions deemed unethical. Again, boundaries on the use of data gathered from wearable tech must be placed.

John Naughton observed that one of the implications of the 'datafication' of everything is the gradual intrusion of technocratic efficiency into areas of human life where it is inappropriate. To illustrate this he made an interesting point about the prison system: a system of 'innocent until proven guilty' is inefficient, costly, time-consuming and wasteful from a resource point of view; 'it would be much simpler to have people locked up based on a decision made by two senior policemen, but we don't do that because we have a criminal justice system which has embedded in it a set of values which we regard as being essential to a democratic system... and the big data movement ignores that fact.'

Session 3: Can We Share Data and Respect Individual Rights Without Disrupting New Business Models?

Richard Mortier's talk addressed his models of Human-Data Interaction (HDI). He explained that humans are being viewed as cogs in machines rather than as individuals with agency: 'A lot of the uses of data in these contexts seem to be thinking very much about 'how can I be made more efficient' rather than 'how can I be enabled to do things myself'.' This has resulted in a dangerous cycle wherein personal data is analysed, the analysis gives rise to certain inferences which then cause actions to be taken, generating more data in turn. This data is then fed back into the analysis without taking into account the fact that this new data may contain some flaw absent in the original human-produced data due to, for example, some flawed inference. This is just one of the issues that permeates data systems that lack adequate human input.

Dr Mortier explained that there are three functions he believes that current systems lack: *Legibility*, the ability of the individual to understand what data is being collected and what it's being used for; *Agency*, the ability to exert some control over the data created or facilitated; and *Negotiability*, the individual lacks the

ability to negotiate terms and conditions, offered only a binary 'accept' or 'decline', or to change their mind once they've accepted. Should they have legibility, they may have some agency, and armed with basic agency, they can try to achieve negotiability.

To achieve this higher level of HDI, Dr Mortier has been designing something he calls a 'databox', which would bring personal data within the agency of individuals who are not experts in the field and do not necessarily understand the complexities that the collection of personal data entails. Databoxes are private devices that collect and manage a person's data, allowing them to see and understand better which items of personal data were being requested or sent where, and by whom. The question then becomes: if databoxes were to become commonplace, how would this affect business practice?

Richard Mortier's talk prompted an open discussion about the nature of data ownership. Does owning a databox make you master of the data within? If the databox uses a cloud service, how is ownership shared between the cloud provider and the user, and if it rests heavily with the user, how much responsibility does the cloud provider have to keep that data safe? Richard Mortier assured the room that the databox was much closer to a research project than a reality at this stage, and that those issues would be addressed once the feasibility of databox implementation had been fully explored.

David Stillwell opened a discussion concerning who should have a say over what data is collected and how data is used. He used the example of a facebook application designed back in 2007 that collected a user's data with their consent through the medium of a personality quiz, resulting in a large data set. This information, he realised, could be used to predict someone's personality from their facebook 'likes' and, with enough material, he found that their personality could be estimated and matched to a refined degree. The information was put to use in the form of targeted advertising campaigns: people deemed from their likes to be extroverts were shown different adverts for the same product than those deemed to be introverts, resulting in an increase in sales.

This concept opened the door to a whole new method of tailored selling. Using the housing market as an example, Dr Stillwell spoke on the possibility of matching the personality of a seller to that of a potential buyer, explaining that the seller was likely to have decorated and customised the house according to their personality. This then extended to banks using social media data to calculate the level of risk in giving a potential customer a loan. This prompts a new question: at what point do people become uncomfortable with their data being used in this way? While 84% of surveyed users were comfortable with their data being used to optimise healthcare, this dropped to 66% for personalised adverts, 47% for car insurance premiums and just 34% of surveyed individuals were comfortable with their data being used to calculate the probability of a loan default. Interestingly, 27% of people said they would be happy to pay \$3 per month to prevent facebook from using their data for financial gain – the same \$3, the audience was told, that facebook makes on average per user per month from their data.

In the roundtable discussion that followed **Irene Ng** argued that distributing facebook revenues according to the number of users overprivileges user data and undermines the technology and capital that was employed to scale and provide facebook's services, while **Niraj Saraf** argued the fundamental feasibility of the idea, doubting that there would be any real safeguard against data use even after users had paid to revoke it.

Session 4: Legal Aspects of Big Data

Julian Huppert's talk introduced the murky area of the necessity of consent with respect to the accessing of personal data. He talked about the contrast between spaces in which we consent as individuals to share our data with another party, including spaces such as social media, and spaces in which society consents for us, for example our health records being shared without our consent when halting spreadable diseases, or our data being used in police investigations. In these latter examples, society deems the end goal of this effective violation of our privacy justification enough to circumvent the need for individual informed

consent. There also exists a third space in which our data is accessed with neither individual nor societal consent – cyber crime and cyber espionage.

Dr Huppert questioned how much control over their digital footprint it's feasible for an individual to retain, touching on the 'right to be forgotten' and how such rights are becoming increasingly difficult to secure. There is also a moral question attached to data retention: on one hand, it would be excellent if courts were more transparent, but on the other, a 60-year-old man should not necessarily be haunted by a conviction obtained at the age of 16, and so a certain trade-off is essential.

Dr Huppert talked about the much-held worry that some large corporations may act in a disfavorable or 'evil' way with our data. However he added that dire consequences occur even in the absence of malicious intent. An example of this is the recent damaging leak of files from Ashley Madison, in which data was leaked, even from individuals who had paid to have their data deleted. Another case mentioned was a federal department in the USA that lost a hoard of employee data, including fingerprints. 'The problem with losing fingerprint data is that while it's simple to change your password or issue new [social security] numbers, it's quite hard to change your fingers.'

Dr Huppert gave a case study on the subject of societal consent. Espionage has moved on from a small data issue to a big one. While previously, letters had to be opened physically, or a conversation could only be overheard by someone present in the room, the internet age has ushered in a set of new possibilities, and associated challenges, to the field of spying. At the moment, text messaging and phone call metadata is held for a minimum of 12 months. With the right clearance, the government can build a very clear and complete picture of where a citizen has been for the past year. It is very clear that this data could be useful, but Dr Huppert asks *how* useful, and what are the downsides? The Communications Data Bill, which Dr Huppert fought against in the UK parliament, was designed to allow the government to track online datalogs in a similar manner to its collection of cellular metadata. Dr Huppert argued that the problem is not that we lack sufficient data to, for example, combat terrorism, but rather that we are unable to utilise the large volume of data that we *do* have in an appropriate way. He used the Woolwich murder as an example: the data necessary to predict and potentially prevent the atrocity was available, and in fact 'the report found that intrusive surveillance of one of the two individuals was supposed to have been started on the morning of the killing, but that they were so busy processing all the other data, they didn't get around to filling the paperwork in on time. The problem was not that we needed more data to throw at this, it was that we needed to have the resources to look at the data that was already there.'

Finally, Dr Huppert highlighted the need to modernise data legislature and also addressed the challenge of data encryption – there exists no system that is very secure against criminals and cyber espionage, but that is open to 'good people'. Dealing with cyber crime is an increasingly difficult and complex task that has no easy cure in the foreseeable future.

The discussion then attempted to tease out some specifics: what constitutes communications data and what constitutes content? If you walk down a street with your phone, you create tracking data that you own, but if you walk into a shop and buy something, you and that shop have created the purchase data together, so who then owns it? Why do we privilege some kinds of consent over others? These questions were largely interesting for the asking, not for the answering since, as with so many aspects of such a young and amorphous subject area, no well-defined and universally-accepted answers exist.



Speaker and Chair Profiles

Professor Ian White FEng, Master, Jesus College, Cambridge, Chair, Rustat Conferences

Professor White is Van Eck Professor of Engineering, and the Master of Jesus College, Cambridge. He is also Head of Photonics Research in the Electrical Division in Engineering, and Chair of the Rustat Conferences. Ian White began his time at Cambridge by being awarded the BA in 1980, and the PhD in 1984. After being appointed a Research Fellow and Assistant Lecturer at Cambridge he then moved to the University of Bath to become Professor of Physics in 1990. In 1996, Professor White moved to the University of Bristol, becoming Head of the Department of Electrical and Electronic Engineering in 1998. He returned to Cambridge in October 2001 as Van Eck Professor of Engineering. Professor White's current research interests are in the area of high speed communication systems, optical datacommunications, laser diodes for communications and engineering applications and RF over fibre systems.

Grace Cassy

Grace Cassy is the Co-Founder of Epsilon Advisory Partners. Epsilon works with leading technology companies and investors, helping them navigate the intersection of government, enterprise, technology and society. Grace also co-founded CyLon, Europe's first cyber-security accelerator programme. She has been a TEDx speaker on balancing privacy concerns with the potential social impact of Big Data. Prior to this, she spent 10 years in the UK Diplomatic Service, working on Asia, Europe, the Americas and global security policy. Between 2004-06, she was a member of Prime Minister Tony Blair's foreign policy team in 10 Downing Street specialising in Asia and national security. She is a graduate of Cambridge University.

Professor Jon Crowcroft FRS

Jon Crowcroft is the Marconi Professor of Networked Systems in the Computer Laboratory, of the University of Cambridge. Prior to that he was professor of networked systems at UCL in the Computer Science Department. He has supervised over 45 PhD students and over 150 Masters students. He is a Fellow of the ACM, a Fellow of the British Computer Society and a Fellow of the IEE and a Fellow of the Royal Academy of Engineering, a Fellow of the IEEE, and in 2013 was elected a Fellow of the Royal Society. He was a member of the IAB 96-02, and went to the first 50 IETF meetings; was general chair for the ACM SIGCOMM 95-99; is recipient of Sigcomm Award in 2009.

He is the Principle Investigator in the Computer Lab for the EU Social Networks project, the EPSRC funded Horizon Digital Economy project, hubbed at Nottingham, the EPSRC funded project on federated sensor nets project FRESNEL, in collaboration with Oxford; and a new 5-year project towards a Carbon Neutral Internet with Leeds. Jon's research interests include Communications, Multimedia and Social Systems, especially Internet related.

Dr Chris Doran

Chris Doran is the Technology Entrepreneur for ARM Research. In this role he is responsible for evaluating new business and technological opportunities for ARM and forming strategic partnerships with universities and other organisations.

Prior to his current role Chris was the Founder and Director of Geomerics, a company specialising in graphics software for the games industry. Geomerics technology has been used in many iconic titles, including the recent releases in the Need for Speed, Dragon Age and Battlefield series. In 2013 Geomerics was acquired by ARM, the company responsible for designing the chips that go into most mobile devices. Chris is

also a Director of Studies in Physics for Sidney Sussex College, Cambridge, and prior to founding Geomerics Chris held a number of research posts, including an EPSRC Advanced Fellowship and an RSE Enterprise Fellowship. He is the author of many research papers and a book on Geometric Algebra.

Dr David Erdos

David Erdos is University Lecturer in Law and the Open Society in the Faculty of Law and a Fellow in Law at Trinity Hall, University of Cambridge. Before joining Cambridge in October 2013, David spent six years as a research fellow at the Centre for Socio-Legal Studies, Faculty of Law and Balliol College, University of Oxford. David's current research explores the nature of Data Protection especially as it intersects with the right to privacy, freedom of expression, freedom of information and freedom of research. In addition, David continues to have a research interest in bill of rights and related constitutional developments, especially in the UK and other 'Westminster' democracies.

David's Data Protection and the Open Society project has developed arguments about the nature, substance and operation of the law by drawing on rigorous comparative empirical analysis using both quantitative and qualitative methods. This analysis, which draws on his background in both law and political science, has demonstrated that in terms of the application of Data Protection law to journalism, literature and the arts, large differences continue to be apparent between European Union countries. David's consolidated research focuses on the origins and impacts of bills of rights especially in the UK and other Westminster-styled democracies (Australia, Canada, New Zealand). This work resulted in a publication of a single-authored OUP monograph, *Delegating Rights Protection*, in 2010. David's academic work has received funding from a range of sources including the British Academy, Council of Europe, Economic and Social Research Council, European Union and Leverhulme Trust.

David Fairs

David is Partner, People Powered Performance, at KPMG in the UK. David leads the People Business which provides clients in the UK and globally with workforce and HR solutions that address their key people and talent issues to drive business performance. A respected commentator on UK pensions, David was previously a partner in KPMG's Pensions team. In that role he has advised on pension strategy and design for public and private sector clients. Passionate about his subject, David is a member of numerous industry bodies, including a working group set up by the government to restore confidence in workplace pensions. He is currently Chairman of the Association of Consulting Actuaries and a member of the Pension Regulators Advisory Panel. David has lead a number of large benefit change programmes including the largest pension communication project carried out in the UK for Lloyds Banking Group which included the development of bespoke modelling tools, delivery of approximately 400 seminars and webinars to 12,000 staff and delivery of over 12,000 one to one meetings. David has delivered similar large-scale communication projects to GSK and IBM.

Dr Julian Huppert

Dr Julian Huppert is the former Liberal Democrat Member of Parliament for Cambridge (2010-2015), having succeeded David Howarth, who stood down after one term at the 2010 election. Unlike most MPs he has a science background, having been an active researcher at the Cavendish Laboratory studying biophysics and bioinformatics of nucleic acids. His political interests include science policy, innovation, human rights, technology policy, immigration, education, transport and climate change. As an MP he chaired the Humanist Group, an all-party parliamentary group, and was vice-chair for Local Government and for Digital Economy; he also sat on the Home Affairs Select Committee.

Julian gained his PhD in Biological Chemistry at Trinity College, Cambridge in 2005. He was elected a Junior Research Fellow of Trinity in 2004, became a fellow of Clare College in 2009, and a lecturer in the Department of Physics in 2012. During his eight years serving as a County Councillor for Cambridge's East Chesterton area, Julian helped open Brown's Field Community Centre, chaired the Cambridge Traffic Management Committee and served on the Regional Assembly. He has also set up a very small biotech company, winning a DTI SMART award for innovation.

Dr Geoff McGrath

Geoff is Chief Innovation Officer and Vice President, McLaren Applied Technologies. He has over twenty years of international leadership and management experience in technological innovation from conceptual engineering design and development through to concept sales to create viable business lines for blue chip and start-up companies alike. Geoff has a proven track record of leading strategy development and execution to deliver rapid and sustained growth in technology, design and engineering businesses. A mechanical engineer by training his career spanned most branches of engineering before changing focus to join pioneers in the mobile internet world for telecoms, media and entertainment. Now in his fifth year with McLaren Applied Technologies, Geoff seeks to capitalise on the convergence of data management, analytics and simulation to deliver high performance design of products and processes in fields as diverse as health, energy, transport and consumer brands. McLaren is renowned for its racing heritage and, more recently for its super-car business. The next big thing at McLaren is McLaren Applied Technologies.

Dr Richard Mortier

Richard Mortier is a member of faculty in the Systems Research Group at the Cambridge University Computer Lab. Past work includes Internet routing, distributed system performance analysis, network management, aesthetic designable machine-readable codes, and home networking. He works in the intersection of systems and networking with human-computer interaction, and is currently focused on how to build user-centric systems infrastructure that enables people to better support themselves in a ubiquitous computing world through Human-Data Interaction.

Professor John Naughton

John is Senior Research Fellow, CRASSH; Emeritus Professor of the Public Understanding of Technology, Open University; Director, Wolfson College, Press Fellowship Programme; Columnist, the Observer; and Adjunct Professor, University College, Cork. John was elected a Fellow of Wolfson College in 1992 and is now an Emeritus Fellow and Vice-President. By background a systems engineer with a strong interest in the social impacts of networking technology, he has written a weekly column for the *Observer* since 1987. He has written extensively on technology and its role in society, is the author of a well-known history of the Internet – *A Brief History of the Future* (Phoenix, 2000) – and is currently working on changes in our information ecosystem brought about by technological change. His latest book - *From Gutenberg to Zuckerberg: what you really need to know about the Internet* - is published by Quercus Books.

He was the Academic Advisor to the Arcadia Project at Cambridge University Library, which ran from 2008-2012 and investigated the role of the academic library in a digital age.

He is currently a Senior Research Fellow in the Centre for Research in the Arts, Social Sciences and Humanities (CRASSH) where (with Professor Richard Evans and Professor David Runciman) he is a Principal Investigator on the Leverhulme-funded research project on "Conspiracy and Democracy".

Dr David Stillwell

David is University Lecturer in Big Data Analytics & Quantitative Social Science at Cambridge Judge Business School, University of Cambridge. He obtained his BSc in Psychology at Nottingham in 2007 before progressing to an MSc in Research Methods in 2008. He obtained his PhD in Decision Making from the School of Psychology at the University of Nottingham in December 2012.

In June 2007, between obtaining his first degree and embarking on his Masters, David began a personal side project designing applications for social networks. The result was the myPersonality Facebook application that allows users to take real psychometric tests and receive feedback on their results. Today myPersonality has collected data from more than six million people and the resulting database has become a priceless academic resource used by more than 100 researchers all over the world.

Since these early successes, the influence of David's work has expanded considerably, so that today its impact ranges from targeted online advertising, the real-time analysis of online digital footprints for online personalization, and behavioural prediction. David's research predicting personality and other psychologi-

cal traits from online behaviour has led to articles in *The Economist*, the *Wall Street Journal*, and an op-ed in *The Financial Times*, and has led to segments on CNN, BBC Click, BBC World Business Report, Bloomberg, and the Colbert Report.

Professor Anna Vignoles

Professor Anna Vignoles is Professor of Education (1938) at the University of Cambridge. She has published widely on widening participation in higher education, social mobility, the impact of school resources on pupil achievement and on the socio-economic gap in pupil achievement. Her research interests include issues pertaining to equity in education, school choice, school efficiency and finance and the economic value of schooling. Anna is a Research Fellow at the Institute for Fiscal Studies and a Visiting Professor at the Institute of Education. Anna has advised numerous government departments, including the Department for Education, the Department of Business, Innovation and Skills and HM Treasury. She provided advice to the Browne Review of Higher Education Funding, the House of Commons Education and Skills Committee investigation of higher education funding, the House of Lords Economic Affairs Select Committee, as part of their inquiry into education and training opportunities for young people, and Lord Leitch's Review of Skills. Anna is also the economist member of the NHS Pay Review Body.

Professor Peter Williamson

Peter is Honorary Professor of International Management Fellow Commoner and Director of Studies in Management at Jesus College. His research interests include Globalisation and its implications for corporate strategy; strategies for success in China and the internationalisation of Chinese companies and their global impact, merger and acquisition strategy and post-merger integration; business ecosystems and the management of networks of strategic alliances; strategies for a carbon-constrained world.

Professor Williamson has wide experience in research, consulting and executive education as well as serving as a non-executive director and chairman of both publically listed companies and new ventures in industries as diverse as whisky, textiles, hedge fund management and Chinese software. He has lived and worked in Australia, France, Hong Kong, Singapore, Switzerland and the United States, before returning to Britain to join Cambridge Judge Business School where he is Academic Director of the Advanced Leadership Programme and Jesus College, where he is Director of Studies in Management. He currently serves as a non-executive director of the global renewable energy firm Green Gas International B.V. and non-executive chairman of the technology-enabled training company Imparta Ltd.

Professor Williamson was formerly with the Boston Consulting Group in London, and Merrill Lynch. He serves on the editorial boards of *European Management Journal*, *Business Strategy Review*, and *Academy of Management Learning and Education*. Peter has acted as consultant on business strategy, mergers and acquisitions, and international expansion to numerous companies, governments and international organisations throughout the Asia-Pacific region as well as in Europe and North America. He has experience in China since 1983, assisting numerous multinationals and joint ventures, and more recently, Chinese companies venturing abroad.

Dr Mohamed Zaki

Mohamed's key research interests lie in the field of Big Data analytics implications for service science, business model innovation and customer experience. His research uses an interdisciplinary approach of Big Data technology to address a range of real organizations problems to make better business decisions within complex service network. He is currently leading the Big Data and analytics for service research at the Cambridge Service Alliance, University of Cambridge. Mohamed is working in many research projects in collaboration with business industries such as BAE systems, Caterpillar, IBM, Pearson and Zoetis. He is also a co-investigator in EPSRC research networks such as Consumer Goods, Big Data, and Redistributed Manufacturing (RECODE) and EPSRC-Digital Economy (NEMODE). He has experience of both academic and industrial research and he is a contributor to the international research community.



Big Data - Participants List

Wednesday 30 September 2015

Professor	Jean	Bacon	<i>Computer Laboratory, and Fellow, Jesus College</i>	<i>University of Cambridge</i>
Professor	Michael	Barrett	<i>Professor of Information Systems and Innovation Studies Cambridge Judge Business School</i>	<i>University of Cambridge</i>
	Jeff	Bauer	<i>National Intelligence University</i>	<i>USEUCOM JAC</i>
Professor	Peter	de Bolla	<i>Professor of Cultural History & Aesthetics, Director of Concept Lab, CRASSH</i>	<i>University of Cambridge</i>
	Christopher	Bradley	<i>Information Strategist</i>	<i>Data Management Advisors</i>
	Grace	Cassy	<i>Co-Founder and Partner</i>	<i>Epsilon Advisory Partners</i>
	Nick	Chism	<i>Foundation Member, Rustat Conferences; Global Chair, Government & Infrastructure</i>	<i>KPMG</i>
	Sinead	Connolly	<i>Privacy Policy</i>	<i>Facebook</i>
	Jonathan	Cornwell	<i>Director, Media Symposia; and Rustat Conferences</i>	<i>Jesus College, Cambridge</i>
	John	Cornwell	<i>Director, Science & Human Dimension Programme; Rustat Conferences</i>	<i>Jesus College, Cambridge</i>
Professor	Jon	Crowcroft	<i>Marconi Professor of Communication Systems, Computer Laboratory</i>	<i>University of Cambridge</i>
	George	de Courcy-Wheeler	<i>Foundation Member, Rustat Conferences, Deputy Chief Investment Officer</i>	<i>Sandaire Investment Office</i>
Dr	James	Dodd	<i>Financier and Technologist</i>	<i>Foundation Member, Rustat Conferences, Jesus College, Cambridge</i>
Dr	Chris	Doran	<i>Technology Entrepreneur, ARM; Director of Studies, Physics</i>	<i>Sidney Sussex College, Cambridge; ARM</i>
Dr	Lydia N.	Drumright	<i>University Lecturer in Clinical Informatics NIHR Career Development Fellow</i>	<i>University of Cambridge</i>
Dr	Clare	Dyer-Smith	<i>Coordinator, Cambridge Big Data</i>	<i>University of Cambridge</i>
Professor	Bryan	Edwards	<i>Defence, Security & Resilience Theme Leader</i>	<i>STFC - Science & Technology Facilities Council</i>
Dr	David	Erdos	<i>University Lecturer in Law and the Open Society, Centre for Intellectual Property & Information Law, Faculty of Law</i>	<i>Fellow, Trinity Hall, University of Cambridge</i>
	Charlie	Ewen	<i>Technology Director</i>	<i>Met Office</i>
Dr	David	Excell	<i>Co-Founder and CTO</i>	<i>Featurespace</i>
	David	Fairs	<i>Partner, People Powered Performance</i>	<i>KPMG</i>
	Mike	Ferguson	<i>Managing Director</i>	<i>Intelligent Business Strategies Limited</i>
Dr	Hamed	Haddadi	<i>Lecturer in Digital Media</i>	<i>Queen Mary, University of London</i>
	Nick	Halstead	<i>CEO and Founder</i>	<i>Datasift</i>
	Helen	Harris	<i>Communications Manager</i>	<i>Jesus College, Cambridge</i>
	Will	Heaven	<i>Technology Features Editor</i>	<i>New Scientist</i>
Dr	Rex	Hughes	<i>Visiting Professor, Canada Centre for Global Security Studies</i>	<i>Monk School, University of Toronto</i>
Dr	Julian	Huppert	<i>Politician; University Lecturer, Fellow, Clare College</i>	<i>University of Cambridge</i>
Dr	Rahul	Jaitly	<i>Technologist; former Chief Architect / CIO / CTO</i>	<i>Financial Times, Universal Music, Kuoni</i>
	Eli	Keren	<i>Rapporteur</i>	<i>Rustat Conferences, Jesus College, Cambridge</i>
	Manu	Kumar	<i>Head of Data Science</i>	<i>Vodafone</i>
	Martin	Lynch	<i>Head of Business Intelligence</i>	<i>AXA</i>
Professor	Robert	Mair	<i>Sir Kirby Laing Professor of Civil Engineering, former Master, Jesus College</i>	<i>University of Cambridge</i>
	Damien	Margetson	<i>Partner, Data Insight Services</i>	<i>KPMG</i>
Professor	Derek	McAuley	<i>Director of Horizon, School of Computer Science</i>	<i>University of Nottingham</i>
Dr	Geoff	McGrath	<i>Chief Innovation Officer</i>	<i>McLaren Applied Technologies</i>

Dr	Richard	Mortier	<i>University Lecturer, Computer Laboratory</i>	<i>University of Cambridge</i>
Professor	John	Naughton	<i>Senior Research Fellow at CRASSH, Emeritus Professor of the Public Understanding of Technology, Open University;</i>	<i>Vice President, Wolfson College, University of Cambridge</i>
Professor	Irene	Ng	<i>Director, International Institute of Product and Service Innovation, Professor of Marketing and Services Systems</i>	<i>WMG, University of Warwick</i>
	Eleanor	Potter	<i>Digital Economy Unit DCMS</i>	<i>Department for Culture, Media and Sport</i>
	Niraj	Saraf	<i>Lead Technologist, Urban Living</i>	<i>Innovate UK - Technology Strategy Board</i>
Dr	Carola-Bibiane	Schönlieb	<i>Lecturer in Applied and Computational Analysis, Department of Applied Mathematics and Theoretical Physics</i>	<i>Fellow, Jesus College, University of Cambridge</i>
	Chris	Sexton	<i>Foundation Member, Rustat Conferences Global Head of Business Systems</i>	<i>Laing O'Rourke</i>
	Jeff	Skopek	<i>Faculty of Law</i>	<i>University of Cambridge</i>
Dr	Sharath	Srinivasan	<i>Director, Centre of Governance and Human Rights</i>	<i>University of Cambridge</i>
Dr	David	Stillwell	<i>Lecturer in Big Data Analytics, and Deputy Director, The Psychometrics Centre</i>	<i>Cambridge Judge Business School University of Cambridge</i>
	Peter	Straw	<i>Customer Insight Manager</i>	<i>British Gas</i>
Professor	Anna	Vignoles	<i>Professor of Education; Theme Champion, Making Big Data Work, Cambridge Big Data; Fellow, Jesus College</i>	<i>University of Cambridge</i>
	Adam	Wethered	<i>Rustat Conferences Foundation Member; Senior Adviser</i>	<i>Sandaire Investment Office</i>
	Georgina	White	<i>Head of BGRS Insight</i>	<i>British Gas</i>
Professor	Ian	White	<i>Van Eck Professor of Engineering, Chair, Rustat Conferences</i>	<i>Master, Jesus College, Cambridge</i>
Professor	Peter	Williamson	<i>Rustat Conferences Advisory Board; Fellow, Jesus College; Professor of International Management</i>	<i>Cambridge Judge Business School University of Cambridge</i>
	Simon	Wright	<i>Strategic Governance Manager, Group Data Protection & Privacy</i>	<i>Sky</i>
Dr	Mohamed	Zaki	<i>Cambridge Service Alliance, Institute of Manufacturing</i>	<i>University of Cambridge</i>