Data Science: Exploring the Mathematical Foundations

November 2014

Knowledge Transfer Network

Smith institute for industrial mathematics and system engineering

Contents

1. Executive Summary

This report provides a summary of the opportunities, challenges and potential future strategies for mathematics and statistics within data science, and the added value that the mathematical sciences can bring to industry. It is based on the findings from a joint workshop on the application of the mathematical sciences to the underpinning foundations of data science, held by the Knowledge Transfer Network and the Smith Institute. The workshop was attended by a mix of academia and industry, giving a variety of perspectives from across a range of sectors, all with one thing in common – they want to get more out of data. Domains represented at the workshop included digital forensics, social media, environment, engineering, technology, security, defence and aerospace.

There has been an upsurge of commercial and academic interest in data science. It has become a crucial tool to handle, manipulate and analyse data on which many important decisions are based. Across all sectors of industry and academia, it is recognised that adding new streams of data or finding patterns in existing data can add value to business. With the ever increasing amount of data available, the role of data science becomes ever more important and the mathematical sciences are a key element for its success.

The table below summaries the opportunities, challenges and strategic priorities, including suggested future activities and noted key principles, for mathematics and statistics to add value to business within the area of data science:

Opportunities	**Challenges**	**Suggested Activities**	**Key Principles**
extracting maximum value from data	analysis gives actionable and compatible options	KTN Ltd and InnovateUK to help coordinate leadership	emphasise co-development of solutions between data users, analysts and providers
providing proof of concept to demonstrate value	balancing security, privacy and accessibility of data	improve access to high quality data	develop mechanisms that kick-start opportunities and new ideas
formalising, quantifying and refining hypotheses	ensuring analysis and visualisation are dynamic	incentivise collaboration between industry and academia	encourage commitment of industrialists to exploit opportunities and impact
capitalising on the advances made in computation	access to high quality data and expertise	encourage open research (results with data provided)	promote the sharing of experiences, good practice and data across domains
creating a data science community	finding both data science and sector specific expertise		prevent intellectual property or legal requirements from inhibiting progress

Each point listed in the above table is discussed in more detail within the report.

2. Introduction

The Knowledge Transfer Network (KTN) and the Smith Institute (SI) held a joint workshop on the application of the mathematical sciences to the underpinning foundations of data science on 23rd July 2014. The workshop was attended by delegates who were specially chosen for their wide-ranging and relevant expertise as well as their appreciation of the mathematical sciences and data science.

This report captures views and opinions that the delegates expressed during the workshop to determine the major mathematical challenges that academia and industry face within data science practice in the UK. Strategic activities and recommendations are given to overcome the challenges in data science and thus increase added value to industry.

3. Background

Industry is continually facing the challenges of dealing with large, complex and sometimes fast-moving data sets that are difficult to process and learn from. There is demand from UK business to know more about different types of data analysis, the methods involved and techniques applied. The demand is driven from the added value that using data more intelligently might bring. However, it is all too easy to misunderstand the data and its structure or to apply inappropriate analytical techniques which consequently draw flawed conclusions. The evolving field of data science is therefore gaining prominence.

Data science is the field of study which is concerned with the collection, preparation, analysis, visualisation and management of data. It is an area which builds upon and incorporates the expertise of many different disciplines to successfully extract meaning from data.

Whilst the computing infrastructure to store and handle data is a necessity, mathematical sciences are fundamental in underpinning the ideas, concepts and techniques required to analyse data and ultimately extract the useful insights that allow industry to make decisions that create growth.

The current position of the UK with respect to data is fast-moving. There are more data about the way we live than ever before and the volumes of data are continuing to increase. There is an increase in the deployment of sensors to track information across a vast range of areas from transport to banking, retailing to farming, science to energy. Social media alone is providing remarkable insights into customer behaviour and emotional preferences. Given the current situation, the mathematical science community have a great opportunity to support UK business.

4. Opportunities and Motivations

The following benefits can be achieved through appropriate analysis of data: new scientific discoveries, market-changing products, increased transparency, improved decision making and enhanced services. Couple significant volumes of data with the benefits of good data analysis and industry, academia and government have the opportunity to enhance the competitiveness of the UK and grow the economy. The UK is particularly well placed to do this because it has a very strong mathematics and statistics community.

The opportunities for the mathematical sciences to add value to business are:

- applying appropriate mathematical and statistical techniques to extract maximum value from data;
- providing proof of concept for new technical areas to demonstrate methods' value to industry;
- formalising and quantifying hypotheses about how the data arose and validate, compare and refine those hypotheses;
- capitalising on the advances made in computation which allow greater flexibility and choice of mathematical and statistical methods;
- creating a data science community, rich in mathematical and statistical knowledge, that can boost the economic growth of UK companies.

The above opportunities as well as examples of areas that were identified during the workshop are described in more detail below.

4.1. Mathematical and Statistical Techniques

Amongst the workshop delegates there was a feeling that many industries are suffering a data deluge. The data collected are often complex and the information to be extracted multifaceted; simple counting methods are insufficient to assemble the required information or produce only simple answers. There is a shift towards the use of more advanced and adaptable methods that have the ability to process structured or unstructured data, of changing sizes and dimensions, that is discrete or continuous, and with varying levels of complexity.

There is a wide variety of mathematical and statistical techniques and complex analytics that can extract value from complicated, multifaceted data. Throughout the workshop there was much emphasis on the successful use of probabilistic methods and the need to disseminate them more widely whilst encouraging understanding by a wider community. Other techniques are listed below with references for further reading gathered at the end of the report:

Probability Theory [1]
Hidden Markov Models [2]
Evolving and Multiplex Networks [3,4]
Deep Learning [5,6,7]
Bayesian Analysis [8]
Classification and Clustering [9,10]
Data mining [11]
Graphical Models [12]
Topological Data Analysis [13]
Tropical Geometry [14,15]
Dynamical systems [16]
Machine learning [17]
Sparse Tensor Methods [18]
Stochastic optimisation tools [19]
Large-scale Linear Algebra [20]

New areas of study – Topological Data Analysis (TDA) example

Many mathematical and statistical techniques are emerging which on paper provide numerous benefits to data science. As these methods are, however, new to the data science scene there is a limited number of applications and case studies of their use.

TDA for example is a new area of study aimed at having applications in areas such as data mining. TDA represents data using topological networks and uses data sampled from an idealised space or shape to infer information about it [13]. It essentially allows algorithms to analyse sets of data to reveal the inherent patterns within rather than showing correlations between preselected variables.

This method could be valuable to industry because it enables the user to discover insight into the data without having to ask the correct question of the data in advance. By working with industrial partners to have access to real data sets, case studies can be published to demonstrate value and promote new techniques to industry.

Not only can these methods be used to extract value from data but they can be used to judge the performance of competing data analytics tools, in terms of computational efficiency, accuracy, stability and scalability to larger problems.

4.2. Big Data

Big data is generally understood to refer to techniques developed to analyse data sets which are either too big, too complex or too lacking in structure to be analysed using standard approaches. A common misconception around big data is the expectation that acquiring powerful computer infrastructure will immediately provide a business advantage. Instead information technology, computer science and mathematical science must go hand in hand. Infrastructure is necessary, but achieving value from big data also requires more sophisticated data analysis methods.

New approaches to analysing data have to be found and, where appropriate, existing methods have to be scaled. This is where the mathematical sciences can make a considerable contribution: building on the foundations of current statistical methods and identifying new techniques to augment or replace old ones that are less appropriate, making the analytics efficient, and most importantly making sure the correct inferences are drawn from the data available.

For big data analytics to make the greatest impact the mathematical sciences need to apply data analytics to areas where the UK has competitive advantage. If there is a real need and a real problem, then this is where true value can be gained.

Big Data Opportunities

With the right analytical and strategic approach, big data has the potential to be of great social benefit by unleashing the power to tailor services, thus enhancing user experience and consumer satisfaction. The combined strength of data can lead to enriched information, enabling the extraction of ground truths and the mapping out of a ‘complete picture’. For example, tracing a drug from its creation to the end of trials and beyond allows for more advanced mathematical modelling of disease patterns and progression, which in turn will provide a better understanding of individual variability, improved risk and impact assessment, and enhanced decision support.

Further examples of big data opportunities for the mathematical sciences lie hand in hand with the fact we are living more and more in a service-oriented era. Smart cities and intelligent systems, such as buses tweeting service times and delays, or home heating systems texting temperature alerts for instance, are already important application areas. The design and development of mathematical algorithms which provide the intelligence for these systems is a key component.

Customer facing industries are data rich. Retail industries with their loyalty card schemes, utility companies with the implementation of smart meters to monitor homes, and phone companies with the rise of smartphones with GPS and internet connectivity are just three examples out of many which have growing markets on a global scale and an abundance of data.

4.3. Reducing and Understanding Uncertainty

Uncertainty exists in many aspects of data science and there is a need to better understand and reduce this uncertainty. Mathematics and statistics can formalise and quantify hypotheses about how the data arose and validate, compare and refine those hypotheses. In this way, we can get an understanding of the underlying “laws of motion” and also make predictions and explore what-if scenarios.

Perhaps more widespread is the opportunity to introduce probabilistic thinking, such that implicit uncertainty can be communicated, visualised and interpreted more accurately. Even with high volumes of data and the best analytical techniques, uncertainty cannot be eliminated. It is therefore important that we can measure the amount and understand the types of uncertainty that surround us in everyday life.

4.4. Bridging the Gap between Industry and Academia

The workshop delegates represented a mix of academia and industry and through discussions it was apparent that there is a need to bring the two closer together. There is enthusiasm to establish stronger relationships and create a data science community which covers research and commercial domains spanning many sectors.

As an example, customer-facing industries have potential for huge growth. With a natural shift occurring to become more customer-focused, mathematical sciences can further pave the way to help understand what customers want, and how to analyse large quantities of data to extract value and gain business and customer advantages. There is therefore the opportunity for industry to use the knowledge and expertise of academics; in turn, academics would gain access to real world examples where new areas of research can be explored and tested.

One challenge to bridging this gap is that incentives in the two communities are often different. Companies for example need to protect their IP in order to gain a competitive edge. Sharing ideas and forming collaborations with the community can therefore be difficult. However, provided these issues are recognised and discussed early when forming new collaborations, they can be managed.

4.5. Cross-Sector Knowledge Transfer

There were 34 delegates present at the data science workshop, with a variety of perspectives from across a range of sectors, all with one thing in common – they want to get more out of data, and they know that the mathematical sciences have the power to help. A community is forming. Domains represented at the workshop included digital forensics, social media, environment, engineering, technology, security, defence and aerospace.

There is an opportunity to take methods already in practice for one sector and test their suitability for application in another. Sharing ideas across the domains, from cars to health to transport for instance, introduces potential for increased growth, as opposed to just incrementally changing current practice.

5. Challenges

There are many challenges and issues facing the data science community. This section highlights the present challenges which caused concern to the delegates of the workshop and areas that could be improved. Priority challenges, issues and areas to be addressed are:

- ensuring data is analysed such that the output gives actionable, priority options which can form the basis for robust decisions;
- maintaining security and privacy of data whilst not discouraging collaborations and open sharing of ideas and methods;
- making big data analysis compatible with computational science to leverage advances in fast computation, data storage and distribution of data;
- ensuring data analysis and visualisation are dynamic, to encompass the continual stream of data and the real-time inference that comes with it;

- creating a data science community which has access to high quality data, case studies and expertise from various sectors across industry and academia;
- finding and hiring staff with the relevant expertise in data analysis as well as sector specific knowledge.

These challenges are described in more detail below.

5.1. Big Data

A prominent challenge facing big data is the need for more real-time streaming methods. Many data-driven sectors such as customer services have evolved to be 24/7 and therefore show a significant shift towards requiring real-time data analysis. Big data methodologies are required which can be tailored to take advantage of timestamps and the data stream. Research has been undertaken into streaming methods but these ideas have not yet gone beyond academic circles. Part of the challenge here is to communicate these mathematical advances to the wider community.

Another challenge facing the mathematical sciences within big data is the need to keep in line with evolving technological advances. New hardware and software platforms will continue to develop to deal with the evolving data landscape and the growth in data volumes. Bespoke data analysis tools written for a specific hardware configuration may not be compatible with future advances and will therefore become redundant if this is not accounted for.

Machine Learning

Machine learning encompasses a valuable set of techniques and methods that can be used to evaluate data. Its methods focus on the ability to learn without being explicitly programmed and can therefore evolve and change when exposed to new data.

Machine learning is used in many applications to process large amounts of data. It is therefore popular in the data science field and its range of applications continues to grow. For example, machine learning has had much success for predictive text messaging, Facebook photo tagging and Google filtering.

Caution should however be applied when relying on machine learning methods to inform significant decisions which can have high impact (positive or negative) on people’s lives, for example the decision of whether to give someone a loan or whether a patient qualifies for a heart by-pass operation.

5.2. Communication

It was evident from the workshop that communication between industry and the mathematics community can be improved. Firstly, there is an issue of problem identification, formulation and dissemination. There is a need to understand what it is that organisations want out of their data and how this can be described to the relevant skills base in order to analyse the data most effectively. Similarly, there is a need to elicit the benefits extracted from the data and help industry interpret the mathematical results.

Visualisation is a powerful tool because it can give a clear idea of the message within the data. It can assist the interaction between industry and the mathematical sciences, aid the explanation of the outputs from data analysis, and inform the decision making process. There is, however, a challenge in understanding the most appropriate visualisation methods and graphical displays to use. Most importantly, the visualisation needs to be dynamic such that the most recent insights of the data are presented.

During the workshop, it was articulated that in some cases industrial members will not be able to understand the mathematics behind a solution, and that they would rely on a bespoke "black box" solution to be produced by a mathematical scientist. Academics can only produce these solutions if they know the types of problems that industry face. Often this may only come from looking at their data. Communication through data is therefore important but can cause data security issues as discussed in section 5.5 below.

It was also recognised that data is often horizontally distributed in vertically structured companies. This shows the importance of effective communication within companies and across differing internal groups, not just between industry and academia. Data scientists with management consultant skills and the understanding to overcome communication issues are thus invaluable.

5.3. Validation and Verification

The area of data science is constantly evolving and new methods or techniques are frequently required to adapt to the changing conditions, it therefore becomes harder to validate and verify the methods and results. Mathematicians and statisticians are expected to create sound, defendable, and auditable conclusions from the data. This is difficult if there are no case studies, or examples to compare ideas, or processes in place to judge the outcomes. It is important to build a knowledge base which has case studies and examples, and provides testing information and evidence to justify results.

In addition, there are many mathematical methods which have not been widely applied in the area of data science but could have the potential to bring considerable advantage. These include:

- Neural computing
- Topological data analysis
- Algebraic statistics
- Tropical geometry
- Pattern theory

5.4. Skills Shortage

The demand for data scientists has risen dramatically in recent years and many companies are finding it hard to hire people with the relevant skills. This is partly due to the creation of new roles and skill sets in industry that did not exist before.

It is uncommon to find experienced individuals who have strong sector specific skills who can also apply cutting-edge mathematical methods of data analysis. There is great value in attracting mathematicians into industrial areas to gain sector-specific knowledge or introduce sector-specific expertise to data science techniques.

The industrialists at the workshop recognised the importance of the mathematical sciences when drawing information from data, and highlighted their frustration of not having (or knowing how to have) access to mathematicians and statisticians. The discussion uncovered that the implementation of tools was not a problem within industry; what they need is access to people who can provide insight. It also highlighted internal challenges within organisations, such as the need to break down silos between different teams.

For mathematicians and statisticians, there are challenges of how to take the mathematical sciences to industry and what are the best mechanisms for commercialisation of their knowledge.

There appears to be great demand in industry for the provision of advice on what the appropriate mathematical and statistical techniques are, and for mathematicians to engage with the domain experts to solve problems together. New collaborations can avoid non-experts taking on specialist work, and instead create and support an environment where each expert contributes from their own specialism.

5.5. Data Security and Anonymity

Collecting data for analysis carries with it the risk that the data might get into the 'wrong hands'. Often the data which contains the most valuable insights is the most sensitive. The importance of privacy and security is therefore a concern. This is not an aspect of data analysis that gets much attention, but its importance is growing and needs to be addressed.

For sectors where the nature of work is classified or commercially confidential, collaboration becomes harder. Operating in this manner can be a challenge because it reduces the opportunities for open discussions and limits the available channels for help.

Furthermore, when fusing data within or between different sources, anonymity can, perhaps inadvertently, be compromised. The stitching together of datasets by matching common features can, accidentally or not, unveil sensitive information which in turn creates a security risk. This risk needs to be managed to ensure the protection of individual and organisational rights.

Bringing together datasets also raises the question of ownership. Where do the boundaries lie when merging data?

For open data, it is important to consider the freedom of use of that data along with the unintended consequence of misuse. Legal requirements, such as those enforced by the Data Protection Act, may not be scalable to current or future data use scenarios. The challenge of law enforcement, who and how to monitor and maintain data usage and what a security process might look like, should be considered.

6. Strategic Priorities

This section identifies activities that can bring added value to industry through the use of mathematics and statistics within data science. The following are the priority areas that are recommended:

- KTN and Innovate UK to help coordinate leadership for this area;
- improve access to high quality data;
- incentivise collaborations between industry and academia;
- encourage open research where research results are published alongside the data on which they are based.

These are described in more detail below.

6.1. Data Science Community

It is recommended that a network of people is established who have a range of skills and expertise, covering multiple industrial sectors and across disciplines within academic institutes. The delegates at the workshop make a very strong and varied foundation for this network. At the base of this network could sit the KTN. It would be important to define who takes clear ownership of this area within the KTN and to define and understand their role. The network can provide assistance in connecting the required skills to the relevant problem and maintain links between academia and industry. Over time, this network can grow and become a clear and obvious place to seek help and discuss data science issues.

Links can be formed between research councils and centres of excellence to build upon the network and further improve the variety of skills, knowledge and expertise available. The network can offer relevant training, seek out the right skills, identify techniques and influence the direction for mathematics and statistics in data science.

In parallel, it is recommended that a central repository for various data sets and case studies is created. These datasets should include benchmarking and examples of good practice. Giving access to data, examples and case studies will improve our understanding, support new research, and create a better insight into how mathematics can bring value. Relevant open source repositories already exist and should be used as a source of complementary information (whilst being careful of differing terminologies between communities).

6.2. Conferences and Workshops

Conferences, workshops and meetings provide further networking opportunities, encourage learning across different sectors and disseminate updates on applications of methods. These events should be organised on a regular basis and advertised widely, attracting participants from multiple disciplines and a variety of industrial sectors to inspire cross fertilisation of ideas. Individual events would benefit from focusing on specific application areas or specific problem categories. Opportunities for different communities to interact in an informal way can also be of benefit.

6.3. Education

Demand for data science skills is increasing. To cover the skills shortage, it is recommended that we build capacity through education and encourage the relevant skills to be taught at University level. This requires participation and commitment from academics to influence the courses that are set across a variety of disciplines.

In the shorter term, reference material can be collated to teach and inform the community when dealing with data. A catalogue of data types could be created which note the different mathematical and statistical techniques that are relevant. This would be generated by mathematical scientists but in a way that can act as a guide and reference for industry. Because of the extensive nature of data science, it would be hard to cover all types of data and associated techniques but at least a catalogue would be able to provide industry with a breadth of knowledge that would not normally be accessible, with the opportunity to contact the relevant mathematical scientists who can help.

6.4. Encouraging Mathematical Activities

Open source competitions and challenges are recommended. Competitions that pose a data science problem and require mathematical techniques to solve it (similar to the Kaggle competitions for machine learning) can attract interest to the area, introduce new ideas and form relationships between relevant sectors. Moreover different approaches can be directly compared.

Another option is to host a Study Group, where industrial problems are presented and a variety of students and academics come together to discuss how they can be solved. Study Groups have been very successful not only at helping to solve industrial problems with mathematics but also with bringing groups together and bridging the gap between mathematics and industry.

Mathematical sciences should take advantage of current opportunities. For example there is an opportunity to engage with Europe through Horizon 2020 and there will be opportunities with the planned Alan Turing Institute to expand, research, educate and transfer knowledge in data science.

7. Recommendations

Section 6 recommends activities that can be jointly undertaken to increase added value for industry. It is important when undertaking these activities that consideration is given to the nature of engagement. Successful activities will share and manage expectations of timescales, intellectual property, the academics' roles, industrial responsiveness to the problems and solutions, how the data is to be distributed and stored, and how future engagements might develop.

With this in mind, we recommend that any initiatives undertaken to address the data science challenge should be directed by the following key features:

- emphasise co-development of solutions between data users, data analysts and data providers;
- develop mechanisms that identify and kick-start early stage opportunities, encouraging new ideas to be tested quickly;
- encourage commitment of industrialists to exploit these opportunities through further development, and document the impact;
- share experiences across domains, and promote good practice in data cleanliness and data sharing;
- operate in a way that does not get tied up in intellectual property or legal requirements before any progress can be made.

If these principles are followed, the mathematical sciences will be in a much stronger position to exploit the opportunities within data science and bring added value to industry.

8. Acknowledgements

This report was written by the Smith Institute, with many thanks to the delegates (Appendix B) for their attendance, input and discussions at the "The Application of the Mathematical Sciences to the Underpinning Foundations of Data Science" Workshop on the 23rd July 2014.

9. References

[1] Durrett, R. (2010) Probability: Theory and Examples. Fourth Edition. Cambridge University Press

[2] Rabiner, L.R. (1989) A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE, 77, 2, 257-286

[3] Grindrod, P., Parsons, M.C., Higham, D.J., Estrada, E. (2011) Communicability across evolving networks. Physical Review E, 83

[4] Gomez-Gardenes, J., Reinares, I., Arenas, A. & Floria, L.M. (2012) Evolution of Cooperation in Multiplex Networks. Nature Scientific Reports, 2, 620

[5] Arel, I., Rose, D.C. & Karnowski, T.P. (2010) Deep Machine Learning – A new Frontier in Artificial Intelligence Research. IEEE Computational Intelligence Magazine

[6] Ghosh-Dastidar, S. & Adeli, H. (2009) Spiking neural networks. International Journal of Neural Systems, 19, 04, 295

[7] Hinton, G.E. (2009) Deep belief networks. Scholarpedia. 4,5,5947

[8] Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J.B., Neyer, F.J. & Aken, M. (2013) A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research. Child Development, 1-19

[9] Michie, D., Spiegelhalter, D.J. & Taylor, C.C. (1994) Machine Learning, Neural and Statistical Classification. Cambridge

[10] Finley, T. & Joachims, T. (2005) Supervised Clustering with Support Vector Machines. Proceedings of the 22nd International Conference on Machine Learning, Germany.

[11] Hand, D., Mannila, H. & Smyth, P. (2001) Principles of Data Mining. Massachusetts Institute of Technology

[12] Koller, D. & Friedman, N. (2009) Probabilistic Graphical Models: Principles and Techniques. MIT Press

[13] Lum, P.Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M. Carlsson, J. & Carlsson, G. (2013) Extracting insights from the shape of complex data using typology. Nature Scientific Reports, 3, 1236

[14] Mikhalkin, G. (2006) Tropical geometry and its applications. Proceedings of the International Congress of Mathematicians. http://arxiv.org/pdf/math/0601041v2.pdf [21 November 2014]

[15] Maclagan, D. & Sturmfels, B. (2013) Introduction to Tropical Geometry. Citeseer. http://homepages.warwick.ac.uk/staff/D.Maclagan/papers/TropicalBook23.8.13.pdf [21 November 2014]

[16] Thelen, E. & Smith, L.B. (2005) Dynamic Systems Theories. http://www.iub.edu/~cogdev/labwork/handbook.pdf [21 November 2014]

[17] Smola, A. & Vishwanathan, S.V.N (2008) Introduction to Machine Learning. Cambridge University Press

[18] Schatz, M., Low, T., Van de Geijn, R. & Kolda, T. (2014) Exploiting symmetry in tensors for high performance. SIAM Journal of Scientific Computing

[19] Byrd, R. H., Hansen, S.L., Nocedal, J. & Singer, Y. (2014) A Stochastic Quasi-Newton Method for Large-Scale Optimization, arXiv. http://arxiv.org/pdf/1401.7020v1.pdf [21 November 2014]

[20] Gleich, D. F., (2014) PageRank beyond the web, Preprint on arXiv. http://arxiv.org/abs/1407.5107 [21 November 2014]

Appendix A - Workshop Agenda

10:00	10:30	**Registration**	KTN
10:30	10:45	Welcome	KTN, SI
		Session 1: Setting the Scene	
10:45	11:15	*Big Data in ICT*	Ian Osborne, KTN
11:15	11:45	*Data Analytics at GCHQ*	Bill Oxbury, GCHQ
11:45	12:15	*Customer Facing Industries*	Peter Grindrod, Oxford
12:15	12:45	*Social Media Data*	Peter Laflin, Bloom Agency
12:45	13:45	**Lunch and Networking**	
13:45	14:45	**Session 2: Challenges and Perspectives**	Delegates
14:45	15:15	**Coffee and Network**	
15:15	16:15	**Session 3: Solutions and Strategies**	Delegates
16:15	16:30	**Conclusions and Final Points**	KTN
16:30		**Coffee and Close**	

Appendix B – Delegate List

First Name	Surname	Affiliation
Ali	Anjomshoaa	Forensic Science SIG
Andrew	Barnes	Home Office
Christoph	Best	Google
Philip	Bond	Council for Science and Technology
Matt	Butchers	The Knowledge Transfer Network
David	Calder	The Knowledge Transfer Network
Felicity	Carlyle	The Knowledge Transfer Network
Simon	Cooper	Welsh Government
Mike	Dewar	NAG
Mark	Doutre	Lockheed Martin
Robert	Downes	CASA – University College London
Torran	Elson	Smith Institute
Hannah	Fry	University College London
Rob	Gill	Deutsche Bank
Peter	Grindrod	University of Oxford
David	Hand	Imperial College London
Des	Higham	University of Strathclyde
Jonathan	Hogg	Science & Technology Facilities Council
Dan	Jabry	CrowdEmotion
Zoe	Kelson	Smith Institute
Peter	Laflin	Bloom Agency
Michelle	Ledbetter	Smith Institute
Tom	McCutcheon	MOD/Dstl
Jonathan	Mitchener	Technology Strategy Board
Peter	Murray	The Knowledge Transfer Network
Ian	Osbourne	The Knowledge Transfer Network
Bill	Oxbury	GCHQ
Tim	Palmer	University of Oxford
Mark	Pitman	Medical Research Council
Sanjiv	Sharma	Airbus
Andrea	Sharpe	Natural Environment Research Council
Patrick	Wolfe	University College London
Jeremy	Yates	University College London