Data as a Resource

Monday 15th January 2018

The Parliamentary and Scientific Committee's meeting on 'Data as a Resource' couldn't have been more timely. The Data Protection Bill is passing through the House of Commons and the General Data Protection Regulation (GDPR) is on the horizon, so this was an excellent opportunity to explore some of the exciting opportunities – as well as challenges – concerning the collection and use of data.

The event began with a warm welcome from the Chairman, Stephen Metcalfe MP, who introduced the four speakers for the evening: Professor Mandy Chessell, Distinguished Engineer and Master Inventor at IBM, Dr Bissan Al-Lazikani, Head of Data Science at The Institute of Cancer Research, Dr Sian Thomas, Head of Information Management at the Food Standards Agency and Simon Burall, Senior Associate for Involve. The speakers each shared their views on the great opportunities and significant challenges that data presents, and what we need to do to best capitalise on this new and rapidly emerging field of research.

"Some say that data is the fuel for the fourth industrial revolution in the same way that oil, gas and coal fuelled previous ones, and they may well have a point"
Stephen Metcalfe MP

Professor Mandy Chessell explained that despite collecting data at an exponential rate, most companies are struggling to do anything with it. This is because data is often sent between systems without context, making it difficult to organise and structure. Context is provided by metadata, detailing how the data was collected, what assumptions have been made and the purpose of the data collection. While the current metadata standards are good, there are also thousands of them, each covering small aspects of the data. Professor Chessell is part of an international multi-company group working on the Apache Atlas open source project[1] that is looking to stitch all the metadata standards together and create an open source reference implementation that demonstrates how we bring the standards together. Professor Chessell is also building a consortium of subject-specific experts through the ODPi[2], attempting to build best practice around using data. The team is also looking to provide compliance tests to ensure technology that is processing data supports the metadata open standards correctly She encouraged everybody to help by explaining to others the importance of metadata and encouraging the use of the open standards created by Apache Atlas.

"Our guilty little secret is actually that most companies are seriously struggling to do anything with data" ***Professor Mandy Chessell, Distinguished Engineer and Master Inventor, IBM***

Dr Bissan Al-Lazikani gave a medical research perspective on data, highlighting how she uses Big Data in real-life cancer research. At The Institute of Cancer Research (ICR), scientists have created canSAR – the world largest cancer database, which contains billions of data points, including genomics, images, characteristics and 3D models of drug molecules, and clinical notes. Dr Al-Lazikani's team have used this data in various ways, including applying machine learning to identify 46 new 'druggable' gene targets, several of which are now being investigated in active drug discovery projects. Dr Al-Lazikani's team have also used data to predict responses to radiotherapy in the clinic. Applying big data approaches to patient data highlighted a number of features, such as genetic mutations or physiological characteristics, which predict a patient's chances of suffering side-effects. This information could be used in the future to personalise a patient's treatment, ensuring they get sufficient doses to treat their cancer while avoiding adverse effects. Dr Al-Lazikani called for three things to help us capitalise on use of data in the future:

1. To make sure that we more routinely collect and integrate key data such as genomics from the NHS.
2. To create intelligent, open contracts with patients about the use of their data for research – allowing us to safeguard their data while empowering them to contribute to research.
3. To train people with the necessary skills to analyse large datasets and contribute to data science projects.

"We need to have an intelligent, grown-up, open contract with patients around the use of data. We want to safeguard their data but also empower them to give it to research"
Dr Bissan Al-Lazikani, Head of Data Science, The Institute of Cancer Research, London

Dr Sian Thomas talked about the various ways the Food Standards Agency (FSA) uses data to protect public health. Data is viewed as a valuable asset at the FSA and as such it is collected, managed and used in a way that allows the public to trust it. Dr Thomas explained how, by labelling and making their data freely available, the FSA has unlocked new ways of using the data, both for protecting public health but also for purposes that had previously not been imagined. By looking at data on allergy incidences, they were able to create a system that better alerts people to make safe food choices in real time, keeping people safe and potentially saving lives. Dr Thomas also talked about the well-known food hygiene ratings system and how the scores are helping drive up food standards through competition. People outside of the FSA have also found novel uses for the food hygiene data, linking low hygiene scores and non-compliance with paying business rates, allowing a local authority in Northern Ireland to recoup £350,000 in otherwise lost business rates in a two-week pilot scheme alone. A key message of Ms Thomas's talk was that data can be used in a myriad of ways, which we often cannot predict. By making data open and accessible we unlock opportunities that can potentially benefit society.

"Data is not the new oil – it's much more renewable than that. We need to use it to get value from it, but we need make it relevant and accessible" ***Dr Sian Thomas, Head of Information Management, The Food Standards Agency***

Simon Burall spoke about how Involve, through the Sciencewise programme, both engages the public on matters of science and policy and gets the Government to listen to their views.

[1] https://cwiki.apache.org/confluence/display/ATLAS/Open+Metadata+and+Governance
[2] https://www.odpi.org/data-governance

By organising spaces with experts, policy makers and the public, Sciencewise works to uncover the trade-offs and tensions associated with society's willingness to accept innovation. Mr Burall argued that the public are more likely to accept technological change so long as it is distributed equally, suitably regulated and has a primary aim of benefiting the public. Mr Burall made the case for engaging the public in this way for issues surrounding the use of their data, but such engagement must be approached in the right way. Organisations need to understand the unique inputs that only the public can provide. Organisations also shouldn't approach the public assuming they lack the knowledge to accept new technologies, as the issues that dictate acceptance are often more political than scientific. Mr Burall also stressed the need for taking time for public engagement, allowing for six to 18 months for projects to get meaningful input from the public on issues surrounding science and technology.

"If you want to really know what the public thinks then take your time about it"
Simon Burall, Senior Associate, Involve

A lively question and answer session saw a range of issues covered ranging from the need for investment in skills, the risks of sharing data, the unintended consequences of making data freely available and the implications of Brexit.

Dr Al-Lazikani argued that starting to teach children as young as five years old about data is important to engaging the next generation with data science. The panel agreed that skills associated with data, such as mathematics and IT skills; need to be improved in everyone. Technology plays an increasingly important role in the day-to-day lives of people, and having these skills will be vital to creating a working relationship between humans and technology.

The panel discussed that there will always be occasional misuse of data and unintended consequences. However, Dr Thomas highlighted that the FSA had a largely positive experience of making their data freely available, with issues being corrected quickly. Professor Chessell warned of data potentially affecting biases in society, with something as simple as who selects and inputs the data potentially leading to different outcomes. This is something that needs to be considered to prevent human biases being translated into new technology.

A key aspect of the discussion was that regulation should focus on the use of data rather than on its collection and sharing. Professor Chessell highlighted that with the interconnectivity of metadata and the associated downstream implicit consent for data use, it is possible to trace how data is transferred and used. By overregulating data collection we risk losing the benefits that the data can bring. Dr Al-Lazikani explained that currently patients are consenting for their data to be used without realising that the use is limited further down the line by regulations. We need dynamic consent for data that allows data to be used for new uses, particularly as technology develops. By creating a system that safeguards against misuse both now and in the future, we have the potential to do great things for society.

Stephen Metcalfe MP then brought the event to a close by remarking that part of the answer to the questions over data lies in continuing to talk about them. Events like this are an important part of ensuring the Government has the right information about the positives and the negatives to make the right choices.

Dr Sam Dick, Science Information and Policy Officer
The Institute of Cancer Research, London