Insight

Centre for Data Analytics

Towards a Magna Carta for Data

Insight Centre for Data Analytics Discussion Paper

Centre for
Data Analytics

Insight

Towards a Magna Carta for Data

Insight Centre for Data Analytics Discussion Paper

DCU NUI Galway OÉ Gaillimh UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh UCD DUBLIN SFI Science Foundation Ireland

Executive Summary

This discussion document has been created by Insight Centre for Data Analytics. With over 350 researchers, Insight is Science Foundation Ireland's largest ever research investment and one of the biggest data analytics institutes in Europe.

Big Data is a frontier with significant implications for citizens, governments and business. There is growing public unease about the pace of growth in Big Data and lack of transparency about its use. Data use by public and private entities raises important questions about ownership, privacy, individual rights and societal progress.

So far discussion about the protection of individuals in the data age has focused on privacy. Legislation and political discourse has largely been limited to this one aspect of the ethical issues that arise in a Big Data age.

Big Data poses significant ethical challenges. The field is constantly changing. Data ethics theory spans disciplines that have traditionally had little or nothing to do with one another. Collaboration across areas such as data science, ethics, law, philosophy is not easy and expertise that spans these disciplines is rare. There is a prevailing attitude that ethics should be dealt with as issues arise – after the technology has been developed. Practitioners are increasingly aware this model is backwards.

We cannot allow technology to outstrip thinking on data ethics. When it comes to data, we need to define the rights of all stakeholders before we put the technology into use. We cannot afford to wait until the technology is in place. The field of data ethics is more developed in some spheres than others. Insight wants to progress the discussion in the data science field and facilitate the cross-disciplinary discussion that needs to happen.

Legislation alone is unlikely to be nimble or all-encompassing enough to protect individuals and their data. We need to define the rights and wrongs of data collection and use. Insight believes that a bill of rights; a set of principles adopted and adhered to by nations and industry; could be the way forward. It would be a living, evolving document, but what exactly it would entail is yet to be decided.

Insight believes that one positive step is to bring people from different disciplines together. There are pockets of expertise in different fields and our first step is to listen and learn.

We can balance the needs and rights of individuals in the Big Data age, but collaboration and cooperation will be key.

A Magna Carta for Data would go some way to providing assurance to citizens that they can share their data without making themselves vulnerable to abuse.

The era of Big Data is upon us. We must develop a bill of rights for data use. We need a Magna Carta for Data.

Contact: Barry O'Sullivan – barry.osullivan@insight-centre.org

For more information: https://www.insight-centre.org/magna-carta-for-data

Towards a Magna Carta for Data: a bill of rights for Europe's citizen

Big Data promises significant societal benefits; a better life for citizens, more and cheaper food production, better transport, education, health and better resource utilisation. But citizens and businesses have concerns. Data companies are collecting data that directly influences them, their families and their livelihoods. These concerns are not only about privacy, but go much deeper. Corporate owners of data can develop lock-in mechanisms to control entire sections of society. This impacts not only those who find themselves to be locked in, but also the wider society.

Big Data holds huge potential for society. If citizens begin to mistrust those to whom they entrust their information, they will soon decide against sharing their data. Fear would lead users to reject the benefits of data analytics.

Tim Berners-Lee, the creator of the world wide web, has suggested the development of a Magna Carta, a bill of rights for European citizens, for the web and data on the Web. Insight Ireland is suggesting that we broaden the Magna Carta for Data beyond the web, as there are many forms of data that are not web accessible but have societal and personal relevance.

Insight has already proposed the concept of a Magna Carta for Data. It has been well-received at a European level. Policy makers, while still focused on privacy and data protection issues rather than the general concept of data ethics, are nonetheless aware of the shortcomings of legislation and the need for something more. What is still unclear is the form such a document would take, and indeed what would be in it.

As data researchers, we are conscious of the fundamental tension that exists between data, its uses and its users. Citizens have legitimate concerns that data collected by third parties, relating to their actions, beliefs, preferences or intentions, may be used without their (informed) consent. They fear that its misuse may have a detrimental impact on their interests, intentionally or not, because third parties can use their data as well.

Researchers have reason to be concerned as well. In January 2012, the European Commission proposed a new framework that aims to address current and future developments in personal data protection. A notable aspect of this framework is the case of personal data that is processed, analysed and combined with the data of others (and with data from other sources) for health and wellness benefits.

The complications that arise are a growing source of unease for researchers.

In 2013 Science Europe, a non-profit organisation representing more than 50 major research funding and research-performing organisations throughout Europe, published an opinion paper: *"The Benefits of Personal Data Processing for Medical Sciences in the Context of Protection of Patient Privacy and Safety."*[1] This paper warned of the devastating implications of amendments, if passed, to the European Commission's proposal for a General Data Protection Regulation. The Commission's proposals contain a number of provisions and exemptions crucial to facilitating vital medical and health research within a framework of protection of individual rights to privacy.

In June 2015, Insight presented at a Day of Action for Data in Health and Science in Brussels. The aim was to meet with MEPs to emphasise the importance of making an exception of health research when introducing data protection legislation. The idea of making such an exception is one that has few opponents but we must be diligent in ensuring that the issue is not overlooked.

While the legislators wrestle with issues of how to develop a framework for the use of Big Data in personal applications, technology and society continue to develop and basic ethical questions have still to be tackled, especially in the data space.

We are data scientists. We have attempted to examine the ethical issues that we believe are pertinent and relevant when it comes to data research. However, we need to consult with people from different disciplines and learn about the issues that are arising in relation to data and ethics.

We want to provide a platform for cross-disciplinary discourse. We want to consult and see if we can devise the framework for a Magna Carta for Data, but to begin with, we must listen, learn and collaborate. This is an exciting project, and one that Insight Ireland is committed to. We look forward to progressing the project in Ireland.

---

[1] http://www.scienceeurope.org/uploads/PublicDocumentsAndSpeeches/ScienceEuropeMedicalPaper.pdf

Questions for the data-driven society

Technology has moved at such a pace that questions are now arising that would not have occurred to us a decade ago. The very notion of ownership is in question.

Insight envisions the Magna Carta for Data as something that will protect the rights of the individual, while enabling research and use of data for the good of societies and economies.

Is such an ambition even realisable? Is our vision of the Magna Carta the right one? Is the idea of a Magna Carta for Data realistic or even necessary? All of these questions are open for discussion.

Following internal discussion, Insight's members have come up with a number of questions that need to be addressed in the context of a discourse on data ethics. The following list is a start, a platform for further analysis. We intend to build on these questions by drawing from a wide pool of stakeholders.

1) Why a Magna Carta for Data?

Individual should be entitled to certain guarantees when it comes to their data, but the internet age has made an individual's rights regarding data difficult to legislate for. The legal protections currently in place are accessible only to those with adequate resources to negotiate them.

One possible way of protecting such rights is a Magna Carta for Data, akin to the Universal Declaration of Human Rights, that could be adopted by nations and industry. It would lay out a set of principles to provide assurances to individuals that, when they click the Terms and Conditions button, they are signing up to fair and agreed set of principles that observes their rights as well as those of the user.

Data itself needs a Magna Carta because data has value. It is a major currency. We cannot limit a Magna Carta for Data to data on the internet. People's rights regarding their data should be universal no matter how the data is collected or stored.

A Magna Carta should be a living and evolving document. Technology develops so rapidly, a Magna Carta for Data should be capable of adapting if necessary.

It should acknowledge the potential value of data and the benefits for society in the use of Big Data. Whether and how and if we balance the value of data versus the rights of the individual is something that needs careful consideration. Data is a valuable currency in research, industry and society. It potentially holds the key to better, safer, more efficient societal systems and much more besides. However a balance is needed. The temptation is ever-present for industry and others to exploit the data of individuals for profit and yet, without the trust and cooperation of the originators of that data, any sustainable system of research would be impossible.

If we don't acknowledge and examine the different interests involved, the rights of the individual will be swept away in this data gold rush. We believe that we must ask the difficult questions. And we believe firmly that the individual, not the value of their data, must be at the centre of any discourse about a Magna Carta for Data.

2) What does it mean to "own" certain data?

Ownership is often seen in the context of a set of entitlements, e.g. to use, store, publicise, sell, transport, or destroy something. Which entitlements are involved in ownership of data? How can these entitlements be justified? Who should own data? What does co-ownership of data mean? Are current interpretations and arrangements of ownership adequate to deal with Big Data? Does the traditional definition of ownership hinder innovation and research? Would the consent of the individual to whom the data belongs be required in every single situation? Can consent for one piece of research be implied in follow up research for example? What are the costs and what are the benefits of protecting ownership rights over research priorities? The economic model of Internet giants is predicated on accessing vast amounts of data in return for services. What are the rights and wrongs of this? How about informed consent? Can consent really be informed when it is attached to an exhaustive list of terms and conditions?

3) What can we learn from Big Data? What doesn't it tell us?

There is also the problem of people who are not captured by the Big Data net. Big Data is often spoken of as holding the key to society's problems, but those who speak like that forget that data footprints are not equal. Older people, disadvantaged people, people on the margins of society are likely to be underrepresented in any data gathering exercise. In relying on personal data to solve societal problems, we may only be capturing the issues of a certain section of society.

The limitations of Big Data need to be examined. What does it tell us? What doesn't it tell us?

4) How do we demand that contracts between individuals and powerful Big Data companies or governments are fair? What constitutes fairness?

Can we assume that the individual, who may not be very well informed about Big Data or privacy, is autonomous when signing the contract? (Particularly when all that is required for agreement is the click of a mouse.) This may be the most useful aspect to a Magna Carta for Data. If we can assure a level of fairness in the principles, and a company has signed up to that set of principles, the individual could tick the terms and conditions with a level of confidence because the company adheres to the Magna Carta for Data.

5) Are there technological solutions that could answer some of the questions around Big Data and the individual?

Is there something that we as data scientists could contribute to? What are the questions that we should be seeking to answer?

6) Where does responsibility for the security of data lie?

Does it lie with individuals or with companies or governments? This will have implications for cost.

7) The overwhelming focus on security and privacy is distracting from the wider issues around data ethics. What are the other things that we need to consider in a data driven society?

Data ethics is an enormous field, spanning many disciplines and areas of expertise. We need to engage in the issues, in the philosophy of big data, if we are to truly understand what's involved. What rights are people entitled to when it comes to their data? This is a huge question, but one that must be answered if our realisation of the possibilities of Big Data is to be sustainable.

Feeling safe in a data-driven society

Big Data has an image problem. It's a fact that today only 12% of European web users feel that they are entirely secure while making online transactions.

In order to ensure the sustainable and reliable functioning of our data-driven society, it is essential to build trust. Article 16 of the Treaty on the Functioning of the EU stipulates that *"everyone has the right to the protection of personal data concerning them*", however the complexity of the issue when it comes to private data and its protection needs a more elaborate policy and legal framework than we currently have. Moreover, citizens require practical tools and services that will provide them with secure access based on trust and control of their own personal data.

The analysis of personal data brings with it not only risks but also many opportunities in various fields from health to urban efficiency. Big Data analysis based on the distributions of DNA markers can better identify and predict disease patterns, for example. Third party companies including Microsoft (HealthVault), Google (Google Fit), Apple (Health Kit) and Samsung (SAMI) are able to compose a holistic overview of a person's state of health and wellness by using data analytics to process personal data uploaded by the individual.

There is huge potential to use data to improve the lives of people in cities. Take the solutions for Smart Cities, for example, involving citizen sensing and collecting data from smart grids, leap cards and more. Research such as this aims to improve the quality of life in cities as well as the use of resources. It can lead to improved safety and transportation, with a clear economic impact on the city ecosystem.

But without trust and buy-in from citizens, we will never fully realise the possibilities of Big Data. Citizens who are not confident in those who wish to use their data will withdraw it. In many ways Big Data research and use is a partnership, but it is one that must have the generator of the data at its heart.

The Strategic Research and Innovation Agenda

The Strategic Research and Innovation Agenda (SRIA) has done some work in defining the overall goals, main technical and non- technical priorities and a research and innovation roadmap for the European contractual Public Private Partnership (cPPP) on Big Data Value. It is a useful resource as we move towards a Magna Carta for Data.

The SRIA has several core activities in data privacy:

- In the technical priority area "Privacy and Anonymisation Mechanisms" the SRI suggests the development of a complete technical data protection mechanism.

- In the non-technical priority area "Social perceptions and societal implications" the SRIA aims to address privacy-by-design principles and create a common understanding amongst the technical community, as well as to identify key privacy concerns and develop answers based on new solutions.

- In the non-technical priority area "Policy, Regulation and Standardisation" the SRIA suggests the setting up of dedicated projects to address the circumstance of new data, with the intention, among others, of establishing an inventory of roadblocks inhibiting a flourishing data driven economy, and to make and collect observations about the discovery of new legal and regulatory challenges along with the implementation of state-of-the-art technology and the introduction of new technology.

The Magna Carta for Data will complement the SRIA. The substantial risks to the economic well-being of Europe from rapidly emerging data oligopolies and monopolies in industry need to be tackled and mitigated.

The Big Data Analytics Lifecycle: Capture – Process - Use

No discussion of data ethics is complete without an examination of the processes involved in data analytics. Each step throws up ethical challenges of its own.

Capture. To date we have only scratched the surface of what might be possible to "sense" in the future. The current trend in wearable technology opens up the possibility of a much more comprehensive approach to personal sensing in the future, for example, beyond tracking coarse-grained activities (such as walking, running, sleeping) to capturing real-time data about intimate physiological signals.

Such capabilities will be possible because of new approaches to biological sensing from the material science that is enabling new kinds of always-on LOAC-based wearable sensors and driving the capture of real-time data from our blood, sweat and tears to a new generation of ingestible sensors that are capable of recording the details of our a variety of biological processes.

As an example, personal data alone is set to explode over the coming decade as more people explore the benefits of personal sensing and the health and lifestyle benefits that it promises. Capturing this data is just be beginning of the Big Data analytics lifecycle. These new sources of data will be described, stored, and formatted in a way that facilitates understanding and usage/prediction. Technologies like Semantic web and linked data technologies create a real-time web of personal data - a world of structured personal data - that can be understood, searched, and otherwise processed.

Process. Unlocking the value of the data is about leveraging the second-order effects of this data, understanding the patterns hidden within, not just at an individual level, but also aggregated across groups, communities and populations of people, to exploit known-patterns and regularities within aggregated data streams. A good example is the use of mobile phone location information to infer traffic congestion in social mapping applications.

But equally it is the third-order effects of data process where a new opportunity exists: the recognition and understanding of previously undetected patterns within the data, from simple periodicities to more complex patterns. It is this vision that is, in part, driving Google X's new Baseline Study to identify a genetic baseline for a healthy human among aggregated genetic data.

Underlying all of this is an assumed willingness of individuals and communities to participate in a level of data exchange and sharing that will facilitate these types of applications. This cuts to the very core of the ethics and

regulation of Big Data analytics and will necessarily require advances in areas such as anonymisation, encryption, and security, leading to a "privacy-by-design" approach.

The engine that enables this includes recent advances in statistical analysis techniques and machine learning from Bayesian methods to Deep Learning. In particular it is necessary to understand the relationships that new machine learning techniques identify, and in doing so explaining the implications and impact of such patterns. Indeed the next major challenge for machine learning techniques is to deploy them as part of an integrated analytics platform, to exploit structured knowledge from the real-world in real-time. This will introduce a new level of automation during data capture thereby closing the loop of the data analytics pipeline.

Use. Deploying Big Data insights in the real world starts with getting the right information to the right user at the right time. But more that this it is about driving the type of behavioural change that can disrupt traditional industries, from health and nutrition, to transport and energy. Understanding the possible disruptions and preventing societal lock-in mechanisms is key to ensuring the benefits of Big Data.

Towards a Magna Carta for Data

A challenge in dealing with personal/private data in a Big Data analytics setting is assuring that policies, queries, access and so-on are governed in a transparent manner and that control lies in the hands of the data owner, subject to due regard for the public interest. While this is governed by policies, standards and systems that create the necessary infrastructure, there is also a requirement for a co-evolutionary relationship with regulations.

Another challenge is to create the infrastructure and standards of using and reusing of public information - Open Data. Thanks to international agreements and European regulations more public data is becoming available, but cross-country regulation of private data sharing (for example, for citizens moving across countries) is still a grey area.

Data (both public and private) is becoming easier to access and integrate. Data owners have control over a portion of their data corresponding to their personal data, and can grant access to other individuals or organisations to use data in pre-specified and well understood ways.

Our vision implies a highly distributed physical infrastructure as a direct reflection of the Magna Carta. This has significant implications for how systems and tools are built for enabling Big Data analytics.

Therefore, a variety of infrastructure elements is required for realising the benefits of data analytics while still being policy-compliant: some at the level of data management protocols, to algorithm for data processing and aggregation, to data exploitation and usage.

Insight is calling on all interested parties to engage with the Magna Carta Project and to make a meaningful contribution to the global conversation on data ethics. Ireland must take up its responsibility in this space. The majority of the world's tech sector giants have headquartered their European operations in Ireland and over the coming years Ireland is likely to be the site of many disputes around data use. We must engage with data ethics, Now is the time.