



UCL POLICY BRIEFING – SEPTEMBER 2014

AUTHORS

Oliver Marsh

UCL Science and Technology Studies

oliver.marsh.13@ucl.ac.uk

Lajos Maurovich-Horvat

UCL Statistics

lajos.horvat.09@ucl.ac.uk

Dr Olivia Stevenson

Public Policy Impact Facilitator

Office of the UCL Vice Provost (Research)

o.stevenson@ucl.ac.uk

KEY FINDINGS

• **Building an educational community around Big Data in Europe should be a priority**, together with setting up the necessary collaboration and dissemination infrastructure to link learners, teachers, systems and leading user organisations.

• **Europe lags behind the US in developing Big Data strategies.**
A clear European education strategy that provides leadership on data quality, management, utilisation, privacy, and training is needed to increase European competitiveness in this area.

• **Big Data technologies have the potential to revolutionise learning, but it is essential that clear policy directives on maintaining privacy protections** are developed alongside technological innovation.

• Massive Open Online Courses (MOOCs) allow large quantities of information to be shared via the Web, but they are not a substitute for teaching.

Big Data and Education: What's the Big Idea?

Summary

This policy briefing discusses some of the key questions around Big Data and education, in particular:

- how Big Data can help to reform educational delivery and enhance learning;
- how academia can partner effectively with industry to exploit Big Data in order to improve learning; and
- how those involved in education can make the most of Big Data whilst being alert to its limitations.

Introduction

Education has always been high on the policy agenda and data has become a key asset for Europe's economies. Big Data (BD) however has been more of a priority in scientific, industrial and public sectors than the education sector. There are a number of pressing issues for education, including how to:

- increase educator effectiveness;
- harness insights from learning experiences;
- deliver education for all that is also tailored to individual learners needs;
- equip students with relevant skills for their future careers.

This policy briefing discusses the challenges of BD for education, learning and innovation in an emerging competitive data economy.

Key issues

What is Big Data?

A clear definition of BD is elusive. BD projects are an emerging field where innovative technology combined with a myriad of data techniques offers the potential to tackle a wide range of issues that appear when collecting and working with a large volume, variety and velocity of data. Whether it is geographical information, crime data, research data, transport data, energy consumption data, or health data, the potential to reuse data and extract value in real-time to gain insight into policy problems and business opportunities and challenges is leading industry to embrace data tools and further their development.

Common examples of the uses of BD can be seen in global corporate giants such as Amazon and Google who employ recommendation engines to match and suggest products, people and advertisements to users based on analysis of their user profile and behaviour data.

“Eventually data will surpass crude oil in importance.” (Conference speaker)

Where does Big Data come from?

BD often relates to unstructured and semi-structured data that appears in all kinds of places, from different sources and formats; for example web content, Twitter posts, Facebook photos and free form comments. Their outputs (website log files, social media sentiment analysis, video streams) can be analysed to provide a range of insights that previously were not possible. For example, decoding the human genome originally took 10 years to process; now it can be achieved in one week. Currently, a key concern is that **BD projects are highly dependant on skilled developers and computer programmers** and in order to avoid limited diffusion and **to democratise use, user-friendly developer tools are needed**. In equal measure **appropriate education and skills training from an early age** to encourage users to play with BD datasets in a fun and informative way **is required**.

Does Big Data have a role in higher education?

There are considerable opportunities for using BD in higher education. It is important that universities use BD to continue to deliver the very best learning environments for the good of society. There may also be possibilities around using BD to link research to education – both by making better use of latest research practices and outcomes to inform teaching and in enabling research activities to be undertaken as part of education.

CASE STUDY: COLLABORATIVE ONLINE SOCIAL MEDIA OBSERVATORY (COSMOS).

An Economic and Social Research Council (ESRC) and Joint Information Systems Committee (JISC) investment bought together social, computer, political, health and mathematical scientists to develop a 'social computational tool kit' that captures user-generated freely available Big Data to answer big social questions. A further educational outcome is the development of social science research training in new methodological tools for Big Social Data.

What are the possibilities and the limits of Big Data in improving education?

One of the most pressing educational concerns is increasing potential of technologies to make learning available online to students globally.

The internet, cloud computing, live stream, and other comparable technological developments bring new forms for delivering and increasing access to learning. In the process the widespread ability to collect, integrate and analyse BD data from these activities is generating important opportunities for improving education, but it also poses new challenges.

“Big ideas don't have to solve education's problems. It is about objectives, using data and data mining to connect content and objectives.” (Conference speaker)

A key part of the higher education agenda is to harness cross-disciplinary intelligence to improve curricula, content and delivery, enhance learners' experiences and create an environment that equips them with the skills and knowledge necessary to meet the changes and challenges posed by the BD revolution. In such complex systems as education it is hard for humans to see patterns, but BD projects have the capabilities to integrate and link traditional and new sources of data. Such capabilities can **generate deeper insights into students' learning and improve classroom teaching** as a clear picture of their capabilities and needs are developed earlier.

How can Big Data help to reform educational delivery and enhance learning?

With innovations in technology and increasing student mobility, educational delivery is changing constantly. Rising in popularity are Massive Open Online Courses (MOOCs). Currently 70 US institutions offer MOOC courses on one of the big US platforms, Coursera and with more than 470,000 student registrations for UK institutions MOOCs there is the ability for academia to gather insights from data in the process of learning itself.

Online teaching can use BD to reform educational delivery and enhance learning in numerous ways, for example to: adapt and improve delivery through personalising learners' experience; create communities of practice; and standardise the presentation of knowledge. In some instances, advances have led to lower dropout rates and greater flexibility in the system, saving money in lost tuition.

However, while MOOCs offer large amounts of data one of the key issues is **robust anonymisation** of individual students, which presents technical difficulties. **Data ownership** can also make publishing results problematic. Also, the **creative thinking** that is fostered in higher education generates complex, and therefore harder to track, information. Investment is needed to ensure the development of new pedagogical approaches to take advantage of the scale and possibilities of BD without losing sight of the complexity involved in tracking creative thinking. In order for Europe to enhance its competitiveness in BD, there is a need to consider the ethical and moral dimensions, especially how to develop robust data anonymisation tools.

Where can academia partner with industry to develop, mine and analyse legal and meaningful data sources in the hope of delivering radically improved learning?

While industry in the main has been driving the BD revolution it is recognised that partnerships between industry and academia hold potential. One of the most pressing issues for R&D is around the storage of BD, the curation of scientific information, and the production, disclosure and consumption of research information. Collaborative discussions on the challenges and opportunities for BD will help to define mutually beneficial agendas and generate transferable skills.

BACKGROUND

The 'Big Data and Education' conference, held at UCL on 13th May 2014, was organised by the UCL Big Data Institute in partnership with Reed Elsevier and run under the umbrella TECY (Technology, Education, Culture and Youth), a series of policy events organised by the European Commission and DIGITALEUROPE, which aims to elucidate the key questions around BD. The conference explored current considerations for the possibilities and limits of BD to help reform educational delivery and enhance learning.

Entitled 'Big Data and Education: What's the Big Idea?', it comprised presentations from speakers across business, education and academia. The presentations looked at institutional, industry and student perspectives to consider the importance of BD as a research, collaboration and communication tool, as well as where academia can partner up with industry to develop, mine and analyse meaningful data sources to deliver improved learning experiences. It also considered broader issues concerning Europe's position, specifically how to stay at the forefront of this worldwide revolution. This policy briefing is based on the conference and supplementary research.

Partnership between academia and industry should not be restricted to short-term gains. A key commitment to ensuring the appropriate delivery of training for students should be around **learning to ask "big questions"** – and being prepared to deal with answers derived from using analytic techniques, which also reflects wider stakeholder interests.

"Big Data is not the answer, and it is not the question. It is an opportunity."
(Conference speaker)

Problems and risks

One of the most important problems is the challenge posed to privacy through the ability to collect, integrate and analyse personal student data. There are also difficulties in obtaining informed consent from users when there are hundreds of data-collection sources, including many machine-generated sources of data. Computational techniques exist that can help to protect individuals from harmful uses of data, yet this is not well understood by the public and fears persist around loss of privacy. **In order to mitigate the risks greater efforts are required** to:

- raise public awareness around BD and its potential for education;
- reduce public fears by developing the science base and providing user generators with the tools and skills to enable them to be in control of their data;
- develop privacy laws that respond to the needs of society with inbuilt long term planning;
- ensure accountability at all stages of the BD revolution.

FURTHER RESOURCES

Find out more about UCL Public Policy and Big Data on our website: www.ucl.ac.uk/public-policy/events/big-data-and-education and watch some films from the event: www.ucl.ac.uk/public-policy/events/big-data-and-education/big-data-and-education-videos

The unique challenges of Big Data

BD transforms services and organisations in rapid time, which makes working out policy implications extremely complex. Some of the unique challenges are:

- **Data integration is crucial** to managing BD, but **governance and data quality** need to be key areas of focus for BD projects.
- BD is highly dependant on skilled developers and computer programmers at present. In order to avoid limited diffusion, **user-friendly developer tools** are needed to grow adoption beyond the few.
- BD involves learning **complex technology**, which restricts the rate at which BD resources are available for education users.
- The focus is on **near real-time data**. Legacy or historical data is at present outside of current data capture; IBM estimates that 90% of all data has been created in the last 2 years.
- **Poor data quality** can significantly impact the effectiveness of BD projects. As analysis on BD grows, so too will the need for validation, standardisation, enrichment and resolution of data.
- There is a **lack of a common language across platforms** that should be addressed to ensure clear data governance and strong communications.

Conclusion: a focus for Europe

Addressing the issues of Big Data policy responses for education, teaching, learning and innovation is a **political issue as well as an educational and technical issue**. It will require action by political leaders in consultation with teachers, learners, institutes, organisations and communities to adequately support developments.

BD issues are not just matters for IT departments and computer scientists but are also a matter for **education policy**. This includes developing strategies that link research to education, developing meaningfully large data on real world problems for teaching tools, and encouraging user confidence and user democratisation.

The stance that Europe takes on developing a clear education strategy that provides leadership on data quality, management, utilisation, privacy and training needs further consideration. **European citizens need to learn to challenge privacy risks and ensure accountability in the system.**