



IMPROVING UPTAKE OF TEXT AND DATA MINING IN THE EU

Facts

Project No:	665940
Program:	H2020 CSA GARRI
Duration:	09/2015 - 08/2017

TDM Spotlight The International Perspective

Who?

Microsoft. Best known for the Windows operating system and Microsoft Office suite of productivity software products like Word, Excel, PowerPoint, etc. Less known are its many activities in the area of data analytics, artificial intelligence and TDM. Microsoft develops products and services that allow people to analyse text, sound, images and data in many different ways. Microsoft offers platforms (Microsoft Intelligent Cloud) that enable third party companies and organisations to innovate and take to market products which rely on Microsoft algorithms and technologies. For example, Microsoft Cognitive Services makes state-of-the-art face, speech and text recognition capabilities available for developers to build powerful AI applications and services. It is estimated that in various ways Microsoft is helping 41,500 European start-ups.

Microsoft does so much in the data analytics field! Is it possible to characterise what you do here in Europe?

There is so much in the field of data analytics going on. Microsoft works all over the world – selling a wide range of products and services, but also supporting and actively participating in the global tech innovation ecosystem. For example, in our Artificial Intelligence and Research organization

comprises more than 6000 people globally and some of these are based in Europe.

We employ 25,000 people in Europe, of which 3,000 are engineers and researchers. In addition to our sales operations, and the research we do here, we have two Microsoft Accelerators in Europe – London and Berlin – which support European start-ups and SMEs in many different ways. Europe has a really vibrant start-up community, and it is great to help and support home-grown talent here in Europe.

Tell us about Microsoft and why it is active in the space of data analytics / TDM.

With the explosion of data, the opportunities for everyone are huge – from launching new products based around data, improving your customers' experiences and discovering new medical treatments, to saving money by using the information assets you have intelligently. Organisations and businesses must effectively handle all aspects of data – managing diverse datasets, processing it, storing it, analysing etc. Microsoft Azure is the cloud computing platform that underpins many of our data services in this area.

As a global organisation, how do you compare Europe's Big Data environment with other parts of the globe?

There is great innovation going on here - from the start-ups we support, and the researchers we



employ here, to the numerous students, researchers and academics who we work with in the data analytics / Artificial Intelligence field.

However, for various legal reasons we find the environment more difficult and complicated here in Europe than, for example, in the US. This means that we have to be more cautious in Europe than we are elsewhere around the globe.

What are the legal issues you face here, which you don't in the US?

In the US, copyright law, combined with specific court rulings on data analytics, mean that people and organisations that mine and analyse third party materials, such as the open web, are on a solid legal footing. Similarly, the US protection for databases only acts to prevent wholesale copying of databases that qualify for copyright protection. This is important for data analytics and data mining, because it means that the facts and data which copyright has never protected and that are held "within" a work subject to copyright or database law remain clearly available for reuse in the US. The US courts refer to this type of use as "non-consumptive use."

Another example is that the US doesn't differentiate between commercial and non-commercial research in the way EU law does.

How might this affect Microsoft, or your customers and partners?

As I indicated previously, it means that we have to move slower in Europe than elsewhere. Also, unlike the US, the notion that facts copyright or database rights in the EU might taint the ability to analyse facts and data embedded within a work, means that the sources that can be used as part of Microsoft Azure will be less varied in the EU than in the US. This in turn will have an effect on European data analytics innovation markets.

Microsoft also supports others, through its accelerators and products and services like Azure where people innovate "on top" of the platforms we provide. The legal environment in Europe will have a ripple effect well beyond just what we do in this space. It influences not just large tech companies but start-ups and SMEs who have to navigate Europe's comparatively complex copyright / database and data protection laws.

For example, from a legal perspective it will be a lot easier for a US based start-up to create new and powerful insights via machine learning by combining open government data, with data sourced from the open web, and then combine it with their own proprietary data.

Tell us about working in partnerships.

We have many business partnerships, but we also work with lots of students, researchers and academics across the globe. Our innovation model is a hybrid one – we are part of an open research ecosystem.

We have the pleasure of welcoming 800 interns a year into our Microsoft research labs. We also have published more than 10,000 peer reviewed articles within academic communities and collaborate with other industry and academic research organisations. The way European copyright law divides research clearly into "commercial" and "non-commercial" categories really does not reflect how research and innovation occurs. The open research community that Microsoft participates in with academics means that research is fluid – there is a porous boundary between us and academia.

How do you find the pool of data scientists available to Microsoft in Europe?

The standard of people we come across is very high in Europe. The problem, like anywhere, is there aren't enough data scientists! To help with this, and the general up-skilling that is required in the workplace, we created Microsoft Virtual Academy. People can take free courses from beginner level and up in many different areas of technology. In the area of TDM / data analytics we offer courses on advanced analytics, Azure, Big Data, business intelligence etc. Why don't you enrol?

Based on interviews with:

Jule Sigall, Assistant General Counsel of IP Policy & Strategy

Kenji Takeda, Director of the Microsoft Azure for Research program / Visiting Fellow at the Alan Turing Institute

TDM BARRIERS

-  Legal issues such as copyright and database laws
-  Lack of data scientists

Discover more
VISIT OUR COLLECTION

