Big Data Case from Almere DataCapital

Version	Who	What
0.1	Frank	First draft
0.2	Oscar Wijsman	Text changes
0.3	Mark Filius	Added highlighted accent to the case. Added Oil and Gas case/attachment as an example
0.4	Oscar Wijsman	Case reshuffled, added some info
0.5	Frank	Inserted logo's and header

What is Almere DataCapital

Big data has enormous potential but harnessing that potential is still in its infancy. It needs a new area of expertise that requires new skills (knowledge) and applications (tools). The city of Almere has started to unlock the potential through the ambitious programme Almere DataCapital. The term DataCapital refers to a concentration of companies, services, knowledge and facilities all dedicated to the capture, storage, search, sharing, analysis and visualisation of big data.

The programme Almere DataCapital brings supply and demand together in a safe and efficient manner. The programme itself has no products or services; it brings small and large players together to enable the development of the big data services that the market demands, which is not limited to cheap, efficient and durable storage and fast data processing. The power of the approach lies in the ability to really explore big data and in doing so provide added value. Almere is building an eco-system of companies, education and research facilities that will provide knowledge and services about and for big data. The programme expects to generate new economic activity and with it, new jobs. Almere is on its way to becoming an (inter)national big data hub.

A major part of Almere DataCapital is to set up services for big data using the as-a-service model (XaaS). A consortium of companies has been put together to enable the service platforms and develop and provide added value with innovative services. The first platform to be developed is for the health sector.

What is Big Data and how does this fit into this project

'Big data' is the term given to data collections that are so big or complex that they are no longer manageable using the usual tools such as conventional databases. The amount of data grows too quickly, keeps changing and is by nature very diffuse and unstructured. To be able to manage such data collections, more and different knowledge will be needed in the future. Knowledge about standards, filters, meta-data, techniques for storing, finding, analysing and securing data, and sector-specific editing of data.

Increasingly, an extreme amount of computing power will be needed, such as that provided by a supercomputer. Furthermore, the massive storage combined with the computing of the large data amounts will consume huge amounts of energy, unless organised more efficiently so that there is a return on the energy costs. Big data is now

everywhere and is causing a real landslide, especially in sectors with a lot of video or images, measurements, mobile devices or with a high degree of information exchange.

Already some sectors have had to learn to deal with huge amounts of data so as not to drown in it. But it is not about drowning One of the major data producers is the medical sector, with a kaleidoscope of imaging from X-rays, MRI and other scanners, genetics, pathology, sensors and research data. The medical sector also has a problem. Often all imaging is stored in a vendor specific manner so that it can't be shared between disciplines and can't be used for research or educational purposes. In all cases, within medical and other sectors, big data means very large amounts of data from very different origins. To not be able to access or process the data results in a loss of knowledge and unused innovative potential.

As you have noticed, we have written Big Data with a capital B and D. This is because Big Data has fast become a concept over the past years that embodies a different way of thinking about data. It has evolved seriously since the amounts of data created and stored have risen exponentially the past few years. There has always been data, but what makes Big Data Big with a capital B?

- *"Every day, we create 2.5 quintillion bytes of data — so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. This data is big data. Big data spans three dimensions: Volume, Velocity and Variety.",* read about these three dimensions here:
- http://www-01.ibm.com/software/data/bigdata/

A 10 minute explanatory video about Big Data can be found here:
http://www.youtube.com/watch?v=eEpxN0htRKI&feature=relmfu

You can look at Big Data from different angles. The positive ones are things like more insight in a lot of sensor data, e.g. regarding child healthcare like the 'IBM Data Baby'. Or the analysis of raw data from breast cancer research. It can be used to detect fraud or warn before something goes wrong. It can be used to optimise processes.
Less clear are the 'benefits' for society like user profiling on websites, personalised advertising and filter bubbles that create an impression of the outside world catered especially to one single user profile. But what is the reality? What is beneficial to society, and what is Orwell's 1984 (http://en.wikipedia.org/wiki/Nineteen_Eighty-Four)? Is Big Brother watching Big Data?

The case

Big Data isn't just about big numbers but can be all kind of data. That makes it often difficult to comprehend. One way of understanding Big Data is to visualise it. But what is the best way to make people understand what's hidden in the data streams and files and how this evolves in time?

Imagine you live in 1991. The first internet browser was just released, the concept of the WWW had just been invented (before that "hyperlinks" were non-existent, strangely..., you could not "click" anything on the internet). Try to imagine what the impact the

Internet had on e.g. the media industry, journalism and the way we deal with collecting information and sharing it.

This is the situation that you are now in regarding Big Data. After the Internet and Mobile Devices it is often called the 'third big wave' in information technology. Every ICT company that deals with data has been thinking about big data for the past year(s), but no-one knows what it will bring us in the next 20 years except that it will (again) change our lives.

So, *what does Big Data mean for your profession*? What does it mean to you? Do you have any idea how much data is surrounding you and at what pace we are drowning in it? Where does it come from? What can a Media & ICT professional expect in this era of Big Data? What opportunities do you meet that challenges the traditional view on your profession? What knowledge and capabilities should you have as a Media & ICT professional to cope with it and profit from Big Data? Question, questions...

Just as important as knowing *what skills* are required to cope with big data, is knowing *how* you as a professional react to managing the problems. How do you go about learning to grasp the situation, understanding the problems, getting to know what has to be done so that you define the skills needed. *Please record the process you and your co-students go through during the MFW!*

Which areas of society and business are impacted by Big Data? How to explain the possibilities of Big Data to business developers? What tools can you imagine an analyst or technician or even an end-user will use to explain his findings? What skills would a Media & ICT professional need to be equipped with to be able to fully profit from what Big Data can bring? More questions....

To give a real life example, please see the attached presentation which describes the big data problems as seen by major oil companies. Why an oil company? It is not the obvious big data case like user profiling but still a straight forward example that e.g. shows that we are now living in a sensor world. The presentation gives you an insight into the type of challenges faced by these companies. Although it's a large presentation, the sheet numbers 2, 4, 8, 10 and 12 will give you sufficient insight.

It is up to you to make your own big data case during the upcoming days. Try to find out what you can do with big data, how to do it and how to present your findings to others.

Information sources

On the Internet information about Big Data is also big. In Google 'big data' has 88 million hits...

There are two documents that you must read

- http://www.almeredatacapital.nl/images/rapporten/ibm-understanding-big-data.pdf Read Part I chapter 1 and 2.
- http://www.almeredatacapital.nl/images/rapporten/big-data-the-next-frontier-for-innovation-competition-and%20productivity.pdf Read at least the executive summery and browse the document.

More information about Big Data can be found here:

- http://computerworld.nl/article/13654/13-misvattingen-over-big-data.html
- http://www.analytics.northwestern.edu/news/news-articles/Teradata-Assists-in-New-Analytics-Focus-at-Northwestern-University%20.html
- http://www.emc.com/about/news/press/2011/20111205-01.htm
- http://infocus.emc.com/william_schmarzo/big-data-mba-course-101a-%E2%80%93-unit-ii/
- http://www.quora.com/Career-Advice/How-do-I-become-a-data-scientist
- http://www.ini.cmu.edu/degrees/mits/curriculum.html
- http://bigdatasi.rice.edu/curriculum/