2012 SOGETI EXECUTIVE SUMMIT

RECORDED FUTURE

In October 2012, an international group of CIOs, senior executives and world renowned speakers gathered at one of the premier hotels in London to discuss Big Data in all its facets. This summary will give you an overview of what was discussed and describe some of the insights that were shared amongst the attendees. The slides used in the presentations can be found at www.sogeti.com/summit2012

→ *Reports, blog posts and continuing discussions can also be found at*

blog.vint.sogeti.com

Contents

Introduction - Michiel Boreel
Sogeti CTO

Michiel Boreel, host of the day, opened the event by asking the big question: "If you could know it all, what would you do?" During the two days of the Big Data event, the program explored what it will really mean when your insight into what is going to happen in the future will be as clear as your insight into what has just happened in the past.

Presently, social, mobile, cloud and now advanced analytics have come together and are starting to have major impact on society. Here it's not just about technological possibility but just as much the social desirability and economic feasibility of the new capabilities.

The topic of Big Data is interesting as it has an inherent tension: it promises great new insights, but as a consumer the hairs in your neck may start to stand up: 'do I really want big companies to know this much about me?' Are we getting to a stage where companies know more about you than you know yourself? On the other hand, companies already have great challenges in 'little data', to get their internal data in order using traditional Business Intelligence. Will we be able to overcome these challenges in time for Big Data?

Setting the Stage - Menno van Doorn
Director VINT Labs

Some people wake up every morning with a monitoring device on their head. The so called Quantified-selfers live and breathe data. They even call themselves 'data-sexuals'. They measure their life, their behavior and every little statistic that they can … everything they can measure and record in their life is recorded for later analyses. Are we now all ready to become data-sexuals?

2

What is big data? What is different? Why is it important?

In the old fable of the king who promises to double a grain of rice for every square on a chess board, the number of grains of rice quickly becomes too much to grasp. In the fable, the large amount of rice became the problem.

That is also how Big Data was initially perceived: big as part of the problem. But now, can we make the turn so that Big actually becomes part of the value, part of the solution? And while the word 'Big' in Big Data draws the attention to the size, it is of course about more than just volume. It's just as much about variety and velocity: tapping into more sources and doing so close to real-time.

Today, the information available is like the exhaust-gas of human activity; analogous to what comes out of the car exhaust: data that is simply the waste, a byproduct of our lives, leaving the digital footprints of who we are. And since Big Data is still evolving, expanding, there are really no *best practices* yet, at best there are evolving 'next practices'.

The case of supermarket chain Target became famous because it perfectly illustrates both the opportunity and the challenges around Big Data. Target set out to predict important life events, to come up with more suitable offers and interactions. As part of the campaign, it started to send targeted advertising to newly pregnant women. One of these was a young girl who lived with her parents, which caused her father to notice these pregnancy targeted ads. Himself not aware of the pending family expansion, he called Target to complain, only to later find out that Target in effect was 'on target' with their educated guess: the girl was in fact pregnant. The case led to great discussions in the media about the power of data and the privacy issues that immediately surround it.
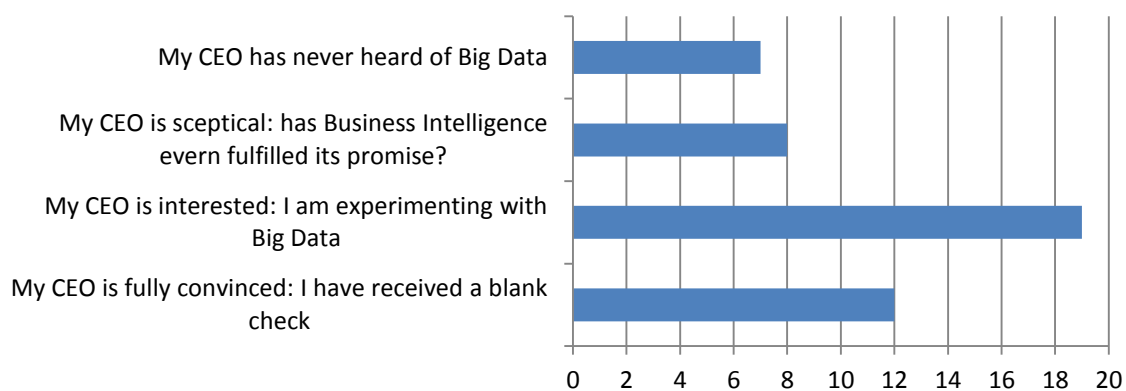
Other cases include PrefPol, which tries to predict crimes before they have happened and Schufa, a credit score company which announced to start analyzing Facebook and Twitter data to

create more reliable credit scores of people. Ilsa Aigner, a German minister, quickly jumped into the discussion and was quoted saying 'Schufa cannot become the Big Brother of the business world'. It illustrates how the forces of insight versus privacy are shaping the field.

What is your CEOs mindset regarding Big Data?
Attendee Poll

During the event, several activities gave insight into the experience of the attendees. They will be presented in these text boxes throughout the document.
In this case, the outcome of a poll on the question above was:

Response	Value
My CEO has never heard of Big Data	7
My CEO is sceptical: has Business Intelligence evern fulfilled its promise?	8
My CEO is interested: I am experimenting with Big Data	19
My CEO is fully convinced: I have received a blank check	12

(Horizontal axis: 0 2 4 6 8 10 12 14 16 18 20)

The Dawn of the Big Data Age – Carlota Perez
Professor of Technology and Socio-Economic Development

Carlota Perez provided the big picture, the context that surrounds Big Data. As an economist, she couldn't ignore the big theme in the economy today: the global crisis. She related the global crisis to the ICT revolution that has taken place in the past decades. And there are clear similarities between this revolution and previous industrial revolutions relating to steam, automobiles, steel etc.

Every technical revolution matures from first 'just' providing new cheap inputs into enabling new products which ultimately lead to new processes. These are all inter-related and based on a new infrastructure network which provides the medium, which transports and connects products, energy, information and people. In that last phase, they truly transform the economy as a whole, creating new markets, new

organizational principles and practices. This, in turn, creates a radical change in business opportunities and direction of innovation.

To illustrate, some of these paradigm shifts that the IT industry has enabled since the seventies:

Before	Now
Mass production	Flexible production
Closed pyramid	Open networks
Stable routines	Continuous improvement
Human resources	Human capital
Suppliers and clients	Value network partners
Fixed plans	Flexible strategies
Three tier markets	Hyper segmented markets
Inter-nationalization (relatively closed countries interacting globally)	Globalization (where business is globalized and can chooses where to do what. Freedom for business to move around)
Neglect of the environment	Environment as a challenge
Information as a burden	Information as an asset

A bit of an optimist

Ms. Perez is undoubtedly an optimist, seeing the opportunities that should come from the revolution. When the bubbles and speculations of the early phases make place for the production capital of the deployment period, she predicts a golden age will unfold.

This also has an effect on ICT. During the 'installation period', ICT was in constant supply push. Now, ICT must transform and learn how to respond to a true demand pull. The paradigm of ICT itself and its potential are now understood by all. Demand, both in volume and in nature, becomes the driving force. Consequently, the role of CIO then changes. He initially focused on being the topmost intermediary to 'modernization' – aimed at productivity, agility and flexibility for renewed competitiveness. Now he has to become

the expert strategic intermediary. Using this modern company to innovate for strategic expansion, finding new markets and better positioning for growth in a globalized world. The CIO is truly a strategy role.

As a conclusion, Ms. Perez introduced what she sees as the three main forces for innovation the coming decades:

1. The deepening and widening of the ICT revolution itself.
2. Spread of the uses of ICT across all other industries and activities (probably 'green').
3. The gestation of the next revolution, for example nano-technology or biotechnology.

It's in this context that Big Data will play its major role.

The Recorded Future - Christopher Ahlberg
CEO Recorded Future

A scientist, starting out from data visualization and human-computer interaction, entrepreneur Christopher Ahlberg founded the company called *Recorded Future*. The company provides forecasting and analysis tools to help analysts visualize temporal data and predict future events. The platform the company developed scans sources on the Internet, and extracts, measures and visualizes the information to show networks and patterns in the past, present, and future. Early in its development, the company caught the eye of Google and the CIA (in a good way), who both invested in the company.

Mr. Ahlberg positioned his challenge by pointing to the many revolutionary developments across the globe: revolts which lead to dictators falling, ultimately changing the order of society and business reality. But could we have predicted them better?

The Big Data revolution is along the axes of growing from internal to external data, from past to future. Technology on databases and data storage has advanced, making ever larger data stores possible. We've moved from structured to unstructured data and from corporate data to Internet/Web, as that is were more and more relevant data can be found.

Finally, the time is right to do better analyses: we have the processing capacity to do fast math, we have large and fast storage and there are there is web-harvesting and advances in linguistics. Combined, they provide the toolbox to do powerful analyses of practically any dataset.

If we could better predict protests in the world, it would surely be interesting for politicians. It's now possible, picking up hints from Twitter or Facebook and mapping them to dates and locations. But what else can we already predict? Perhaps we can find out when competitors will open new offices, when senior managers are expected to step down, when companies are planning to launch new products, what types of patents are they developing etc.

Some examples of what applications of these insights could be:

Monitoring unrest. FoxConn, the major Chinese supplier of Apple, has been dealing with unrest in their factories. Tracking publicly available data you can gain insight into what is really going on. If you aggregate what happens at FoxConn, it seems as if they are moving the majority of their production facility to the coast, while most of the protests are happening at their inland facilities. Are these facts related? The information was picked up from Chinese media, which in itself may be hard to get insight into, so finding these patterns is very novel. Another example is an application for monitoring what is happening surrounding US embassies and their people, especially in the aftermath of recent attacks on embassies. Or, a different type of 'unrest', you can think of Medical companies that have to deal with the discussions around science, medicine, side effects, ethics etc.

Analyzing cyber threats. Plotting patterns and trying to predict what may happen to the company in a couple of weeks, to prepare the right counter measures. It would have been helpful for the big American banks which have been under sustained attack in the past weeks.

Competitive intelligence. For example: how to pick a moment to launch your product so that it best stands out between the marketing actions of competitors? This is very relevant information to for example tablet producers or car companies: you don't want to launch your new flagship product a day before or after someone else, as it would minimize your effect. Another way of gaining competitive advantage could be by getting insight into 'who is suing whom' or 'who is working with whom', to get an idea about new alliances or mergers ahead of their public announcement.

Lead generation. Basically any detail about changes in a company can be a lead. Employment changes for CEOs, market plans, consumer activism etc.

Trading on patterns. Or more in general: gaining a market or trading advantage by analyzing complex events. For example by looking at trading by people who have insight into the inner workings of an organization may give significant returns compared to only deciding on official published press releases, even when such trades are well within the legal boundaries. Normally these would be too complex for a human to oversee, but with the use of technology, the pattern can be found and acted upon.

Summarizing in one sentence what the goal is of all these variations, it's this: What we're aiming for is to shift from surprise to foresight.

→ *An interview by businessinsider.com with the speaker can be found at:*
http://articles.businessinsider.com/2011-03-07/tech/29999286_1_intelligence-agencies-business-insider-spotfire

Is Big Data a fundamental change or is it all hype?
Table discussion

Attendees were challenged to come up with three arguments in favor of either viewpoint. Some quotes that came out of this discussion:

Big Data is just a hype, because…

"It's more about small data, with focus on the human"

"Getting access to the data is more important than big data itself"

"It may be too hard to take cultural differences into account"

"You will not find anything that is truly disruptive through Big Data, as these disruptions don't happen enough"

"Integrity issues around private, personal data will stand in the way of Big Data's success"

Big Data will turn out to be a fundamental change because…

"You can't change the market. The market will drive the data to you."

"A combination of data will definitely change your business. I can help you more if I know more about you."

"It's human nature to share"

"Big Data allows faster interaction, much more detailed"

"It will enhance interaction between provider and consumer, making it beneficial to both"

"We are driven by the customer needs, the need to build trust and lower risk. Information is at the core of everything we do"

Digital Narcissism - Andrew Keen
Author, a.k.a. 'the antichrist of Silicon Valley'

Andrew Keen is known for his view that the current Internet culture and the social media trends may be bad for culture in general. Keen is especially concerned about the way that the current Internet culture undermines the authority of so called 'learned experts' and the work of professionals. He talks about these ideas extensively in his book 'Digital Vertigo'.

According to Keen, today's online social revolution is dividing, diminishing and disorienting us. Like in the Hitchcock movie Vertigo, we are tempted to fall in love with an illusion. In the social media and Big Data space we're falling in love with something too good to be true. A slew of books describe the positive sides of social networks, sharing and connecting, but Keen is not convinced.

Many people, including other speakers today, see the network as very central to the world. Like in the movie about Facebook where the movie-Mark-Zuckerberg is saying 'we are going to live on the Internet'. The network is positioned as something positive, as something that will solve some of the eternal human problems. It will 'eliminate loneliness'. That alone is an area where plenty startups are popping up in large numbers. They create social reading, social television, social everything. For free, even.

We're voluntary playing a part in the 'Truman Show'. We are giving up privacy and becoming more transparent, we are merging our public and private lives. Yet mystery and secrecy are essential parts of who we are. These dimensions of being human could be lost if we end up only living public lives.

Visibility is a trap, and the hypervisibility of the day is a hypertrap. We are led to believe that we are interesting, while for many it merely reveals triviality. The Facebook model is very clear: we are giving our data away for free and our data is essential for their business. When the quote 'Data is the new Oil' holds any truth, the people who own the data are the ones becoming rich, they are the oil-barons of the time.

As a counterpoint, Susan Cain has written an interesting book called "Quiet" about how introverts drive innovation. The premise of sharing and collaboration is that it increases innovation. Yet there are many examples of soloists achieving great breakthroughs. Steve Wozniak, who has practically invented the personal computer, is a great example of someone who did not like to share but did great inventions. How do people like him fit in a social-network world?
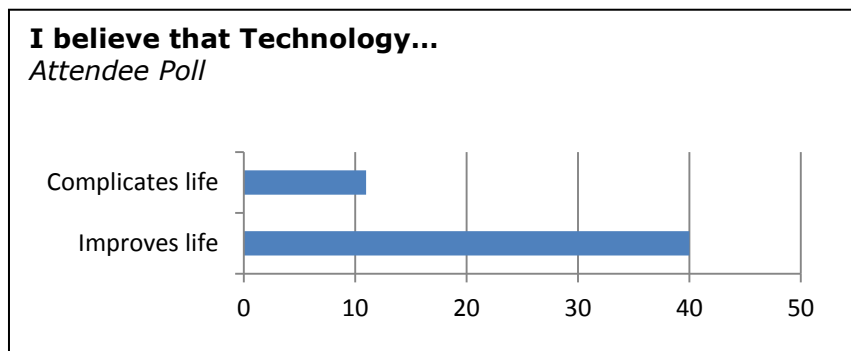
Keen is very outspoken in how to become more innovative: "Discourage people from sharing, chatting. Prevent group-think. Let them think for themselves, take away their devices, put them in dark rooms!". And what about governments who spy on their citizens? We have to find a way to protect the individual from the spies of the government, the big companies, etc. There is a role for government to protect the individual, to give people back the right to be forgotten, as today we've built a monster that doesn't know how to forget.

But can we still change course, or are we fighting windmills? According to Keen, we have to keep trying. Technology doesn't create itself; there are people at the helm. We should stop repeating the 'privacy is dead' mantra, as it's only dead if we allow it to be. He's mildly optimistic about the chances that we will find our way out of the Truman Show and that in the end, not all privacy will be lost.

→ *A recent article by Keen on Facebook and privacy has additional thought provoking insights:* http://edition.cnn.com/2012/05/30/tech/keen-technology-Facebook-privacy/index.html

I believe that Technology...
Attendee Poll

Complicates life	11
Improves life	40

(x-axis: 0, 10, 20, 30, 40, 50)

Too Big to Know - David Weinberger
Harvard Berkman Center for Internet and Society

How does the computer help human understanding? There are examples in Biology where the exact workings of the cell, the biological system of signaling across cells, can only be understood with the help of a computer and computer visualizations. Also, there is a software tool on the market,

11

Eurequa, which you can feed with large volumes of data and the tool will derive equations that are valid for the dataset. What's interesting is that the formulas work, they describe the data well, but we don't necessarily understand them or how they were derived.

This inspired David Weinberger to pose the following hypothetical situation: Imagine a world where some system exists which taps a lot of data so that it can give you the answer to big societal questions. Now, if you were to ask that system a question such as 'how do we revive the global economy' and the system would answer you with the following instruction: 'set fire to all the oil wells', what would you do? If you do not have insight into the reasoning or assumptions that lie behind the answer, would you actually turn the answer into action? Weinberger states that unless you have true understanding, the answer will be un-actionable.

Weinberger's book 'Too big to know' deals with the topic of knowledge and especially networked knowledge. In his talk, he shares a few examples of networked knowledge and how our understanding of 'knowledge' itself may be shifting.

Traditionally, knowledge is rare, it's 'what we've agreed upon' and it's orderly. Knowledge has for years been to create tree structures, taxonomies of things, animals and plants. Also, true knowledge is where traditionally the discussion ends: once we reach something we 'know', we no longer question it, we don't discuss alternatives. Knowledge is where we have driven out all disagreement, disbelief and doubt so that there is a certain finality to it. From 'it's true because' to 'it's just true'. Incidentally, the characteristics that apply to knowledge also applied to the medium that we used to document and spread knowledge: books are equally rare, orderly and final.

But now enter networked knowledge. What does that look like?

1. **Networked science.** In older days, whatever news or information came to you, that was it. You were limited to the book or newspaper article that informed you of some fact. No answers to your specific questions, no discussion, no dialogue. Now, science is very much networked. For example when some scientists discovered that perhaps there were exceptions to the rule that nothing could travel faster than the speed of light, they choose to post it on arXiv.org, an open website instead of waiting for a traditional peer-reviewed publication. This way, they would stimulate discussion, attract counter-arguments and tap into a large

network of people who would discuss their discovery at many different levels. Anyone who learned of this news could completely satisfy their curiosity, as there were many ways to explore further.

So, if you really want innovation, don't take away people's devices and put them in a dark room in isolation (as suggested by Mr. Keen), but create a highly interactive but diverse environment. The value is created by the diversity. Then you can use flexible taxonomies or maintain multiple taxonomies side-by-side to make information accessible independent of your frame of reference. We can get past the differences in nomenclature and classification, and get to the discussion about 'what it is'.

Some lessons from networked science:

- o Peer review doesn't scale
- o Networks flood the ecosystem
- o Knowledge contains differences
- o We are learning how to maintain fruitful differences

2. **Networked Learning.** In the software world, an education system has emerged which is more effective and efficient than any system before. It is based on contributions from many people, iterations on content, reuse of assets and a certain humility and generosity: people are open to discussion, improvement and enhancements. In effect, knowledge in this ecosystem is always a work in progress. The somewhat unordered nature of the environment invites participation and makes sure answers are shared, or a Weinberger calls it: 'Messiness scales meaning'.

Then, Weinberger came to two conclusions:

One - Knowledge is evolving

Knowledge itself is changing, from curated to inclusive, from scarce to overwhelming, from orderly to messy, from discrete to linked and from settled to unsettled. Networked knowledge may or may not be more true about the world, but at least it's more true about human knowledge.

Two - What we have in common isn't the knowledge about which we agree, but a shared world about which we disagree. Weinberger doesn't think that we will reach a utopian world where everybody in the world agrees on the same facts and where conflicts are of the past. But could more data at least lead to less disagreement?

What should we do first?
Group Action Items after the first day

Based on the first speakers, the following actions were mentioned as 'to do' in the field of Big Data:

- Find out where the Big Bucks are. Define the benefits cases.
- Find the core uncertainties. Find which questions need to be answered.
- Build in the ability to forget. Expect regulation or clients demanding this.
- Create incentives for good quality of information.
- In the public sector: be aware of the risk that Big Data will become one of the tools for true digital warfare.
- Find the value in better informed decisions.
- Focus on saving time as a currency.
- Start with the values and the principles.
- Management of Big Data should leave some decisions to others.
- Make the customer central to your Big Data initiatives.

What are the opportunities and challenges for Big Data?
Table Discussion

The attendees discussed to gather arguments why Big Data is good for you, or why it could be bad for you. Some of the arguments:

Why is Big Data good for you?
- It affects me directly, it's about me.
- The knowledge we can get through Big Data is good, through sort and search we already have a wealth of information, imagine what else we can find.
- Think of data services for local government, improving traffic flows, improve the outcomes of local projects.
- Big Data is good for your operating margin: where there is mystery there is margin.
- It could improve medical outcomes.

Why is Big Data bad for you?
- Of course the privacy issues that surround Big Data, as currently a system doesn't 'forgive' or forget.
- Becoming more data reliant creates an opportunity for disinformation too.
- Every 'era' so far has ended or resulted in some form of violent conflict. Will that be the fate of big data too?
- It takes time and money, big data doesn't deliver ROI. We're simply drowning in data.
- Perhaps it's unnatural. Ignorance is bliss. Do you really want to know everything?
- Fatalistic attitude. If we can predict more, will that leave us all depressed? Thinking that perhaps we don't own our own destiny?
- If Big Data really becomes part of the business world, it may drive people to decisions that they HAVE to take in a certain way, because the answers to important questions are all provided by big data. What would be the effect?

Networked Data - Albert-László Barabási

Professor at Center for Complex Network Research

Professor Barabási is the author of the books *Burst* and *Linked*. It surprises him how people are now suddenly interested in Big Data, while he has been talking for at least a decade about large and linked systems, which is very close to Big Data. The essence is that we are dealing with networks and networked data. Facebook is perhaps the most popular example of a network, but any CIO knows that the networked nature of technology is everywhere. Similarly in biology or understanding life in general: it's all networked.

If you look at any networked system, a simple model may seem to suggest that it is pretty random. But start connecting and you'll see the pathways and the networks that it is part of. Then, doing more research, there turn out to be some simple rules or 'laws' that hold true for networked systems. The first law is perhaps most well known:

1. *Six degrees*: In any network, there will be a short connection between practically any two individuals. Short in comparison to number of participants, at least. This is a fundamental network law: via connections you get to others. (To some this is known as the 'Kevin Bacon'-theory, as through associating movies to one-another you can jump to Kevin Bacon within at most six steps from any movie)

2. *Scale free property*: Random networks are uniform. In contrast, if you look at a map of airline routes or the structure of the Internet, there are highly connected hubs and outlying smaller nodes. If a map were truly random, they would be much more uniform and each node would have roughly the same number of connections.

3. *Rich gets richer*: Networks expand by addition of new nodes. If nodes can express some kind of preference during this process, they will often choose to connect to other highly connected nodes. This actually helps to create nodes.

4. *Achilles heel*: If airport disruptions are truly random, shutting down random airports will not quickly collapse the entire network. In fact, if we take out nodes in a truly random fashion, up to 95% of the nodes could be down, and the network would still function. A much different effect would be if through targeted attacks, the biggest hubs/nodes would be shut down first. In that case, the network would collapse much quicker.

5. *Vulnerability via interconnectivity*: If you would kill one node, it will lead the load to another node, possibly causing a cascade effect. This has been very visible in some failures in electrical grids in the United States.

6. *The power of maps*: When investigating anything that is networked, mapping the network is an immensely valuable tool. For example when investigating a company which was looking for the root of a communication problem. Mapping showed that NOT the top management was a highly connected hub in communication but another 'regular' employee who happened to travel a lot between locations while looking at health and safety. This one person was much more important in spreading organizational news and information than any manager or corporate function!

In the second part of the talk, professor Barabási showed how things in networks are happening in real time, and how we can start to predict. An interesting animation showed a map of Paris, displaying movements of cell phone users, giving insight into heavily trafficked areas, movement patterns etc. This mobile location data is now available and it tells a lot about the user. Cell companies own this data, but are still very cautious with exploiting it.

Looking into this location data, and looking at the patterns over multiple years, one of the conclusions is pretty remarkable: people don't travel very much. We make many individual choices that may vary widely, but as a society we fall within a nice statistical

formula describing how much we travel. Most people are moving between a limited number of locations and only a few true frequent flyers visit new destinations frequently. Based on only three months of cell phone data, it's already possible to make predictions for most people that are correct 93% of the time. For hardly anyone it's so difficult to predict that we can 'only' be correct 80% of the time. So not only are we highly predictable, we're also equally predictable. An application of this kind of research is to predict how global contagious diseases would spread or when a pandemic could occur. This can be done with great accuracy.

It shows how we can take human behavior and apply scientific methods to find patterns that are then highly predictable. And what can we predict? Almost anything that defines who we are:
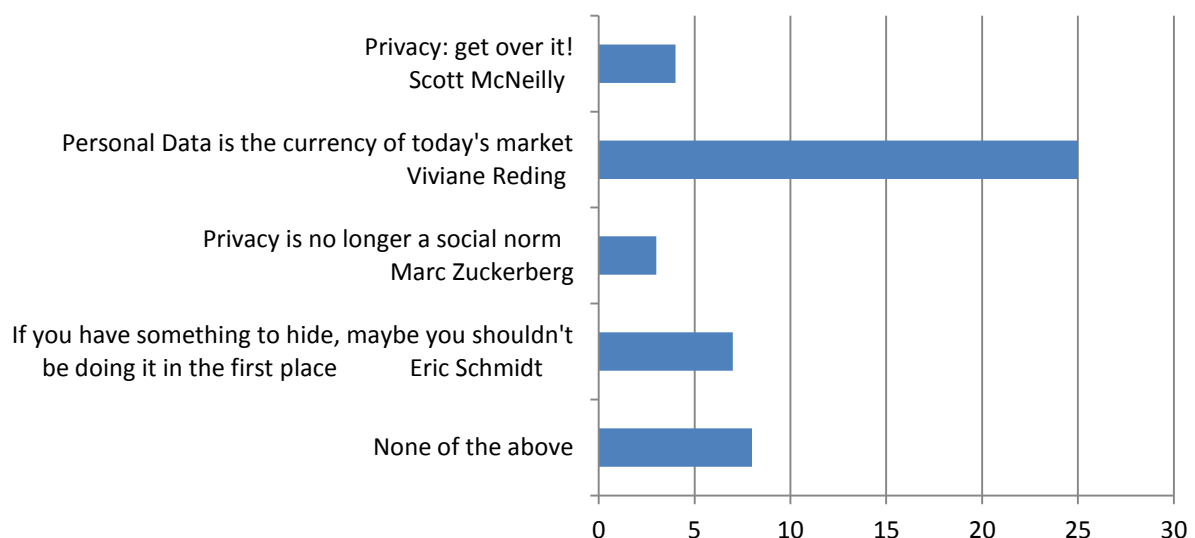
1. With whom – mostly with hubs, or others similar to us – Network science
2. Where – Mostly nearby, as we rarely go too far – Human Mobility
3. Why – Because we did it yesterday – predictability

→ More of Professor Barabási can be found at Barabasilab.com

With whom do you agree?
Attendee Poll

Option	Value
Privacy: get over it! Scott McNeilly	4
Personal Data is the currency of today's market Viviane Reding	25
Privacy is no longer a social norm Marc Zuckerberg	3
If you have something to hide, maybe you shouldn't be doing it in the first place Eric Schmidt	7
None of the above	8

The Intention Economy - Doc Searls
Fellow at the Center for Information Technology & Society at the University of California

Doc Searls has written 'The Intention Economy – when customers take charge,' which describes what could happen when customers truly take charge of their data. He opened his talk by putting the discussion about privacy and 'Big Brother' into perspective by saying "the important thing about a conspiracy theory is that it assumes competence…", followed by a well timed pause. Still, not all is well with how companies treat customers. Companies say they have the customer at heart, but it's often not true. IBM has a story it calls Chief Executive Customer, but it is mainly about connecting data points: the customer is not involved in it.

The Wall Street Journal has published a series of articles called 'what they know' and sites like Rapleaf.com say they can give you real time data on 80% of US emails. So what do they know? Doc Searls took a look to see if what they claim to know is actually correct. And it wasn't. Most of the data was not correct or incomplete. And you know this, because you can see what kinds of ads you are getting online. A lot of it is not working that well. And then there is the interesting distinction between consumers versus customers: the consumer may not actually be paying. Who is buying, who is being bought?

Part of the problem is visible in how the use of the web-browser has changed. It started out as a car: you owned it to navigate the Internet and the ecosystem adjusted around it. Nowadays the car has changed into more of a shopping cart with everyone dumping their cookies in it, following you around. But people are starting to fight back: the number one downloaded browser extension is AdBlock. 180 million downloads by people who no longer see any ads on the Internet.

The real answer would lie in a total reversal of the model: Vendor Relationship Management (VRM), where the customer is in the driver's seat. With VRM, you control your data and how it's used, not the corporations. You set your own policies, preferences and terms of service. For example, you set conditions such as 'don't track me outside your site of service', or 'give me my data in a usable form'. With VRM you can 'intentcast' to the whole market, have the open

market place come to us. If you're looking for a product or service, you simply announce the intent to buy something and vendors will compete for your sale. With VRM you can manage many relationships in your own way. It is how, in theory, free customers will prove more valuable than captured ones.

Does Big Data create competitive advantage?
Table discussion

What are examples of how Big Data can create competitive advantage in your industry? Some examples are:
- Microsoft: Getting better insight into partners, their activities and chances of success.
- Banking: Come up with better predictions, using external data to predict when there is a customer life event, where to place a new ATM or simply support marketers better.
- Telecommunication provider: addressing the issue of churn. Focusing on behavior of their top clients. A total of 800.000 Euros were saved per month by retaining important customers. They looked at the network instead of collection of individuals. (Already a case from 2004!)
- Healthcare: Get better health outcomes, find out how to select good trials, mine medical records to get better targeting for medicine and find out when medicine will and will not work, get trials done faster which equals a much shorter time to market.
- Postal services: Get better predictability to deliver packages, profiles of users receiving packages.
- Credit card company: very obvious, but important nonetheless, is to get better insight into the credit worthiness of people, their financial behavior.
- Aircraft industry: Keep production line machines running, predict when they will break down and predict material failure before it occurs.
- Media: Publish tomorrow's newspaper today

The Feedback Economy - Mark Plakias
VP Orange Institute - Silicon Valley

The context of Big Data is that it is part of the five big tech trends of the next three years: Mobile, Social, Cloud, Online Video and Big Data. We are now squarely in the middle of the post-PC era, where different rules apply. Hardware becomes irrelevant, as for example Amazon is offering Kindle devices at a low price saying "We want to make money when people USE our devices, not when they BUY the device."

IT itself seems to have disappeared, as it moved online, or at least changed beyond recognition. Earlier, IT was expensive, custom built and had a strong tie to physical hardware that we had to buy. Today, IT is cheap, has push-button availability and is made available as a service. For the ones providing the services,

in the post-IT era, it's no longer a mono-culture of IT but a patchwork of tools and technologies: python, scala, mysql, mongodb, jquery and many more. Basically 'whatever works', but not very often one of the big traditional platforms such as IBM, Microsoft or Oracle. Even though startups still like to target the enterprise space, as it remains an enormous opportunity: enterprise focused startups are much more likely to hit a billion dollar revenue compared to consumer oriented ones.

→ Read more on the post IT era at http://www.orange.com/en/press/OSV/postitera

When it comes to Big Data specifically, Michael Schrage, editor and research fellow of the MIT Technology Review and blogger for the Harvard Business Review, was quoted saying 'Executives need to understand that big data is about deciding what kinds of data should enhance or transform user experiences. Properly managed, IT becomes a new medium for shaping how people and their technologies interact'. And 'technology' by now is everywhere and many startups are looking to combine (big) data, technology and social into the next killer app.

Bitly, a link-shortening service that sits on a ton of Internet browsing data gained visibility when Hillary Mason wrote a blog post called 'Help I'm the first data scientist at my company' (→www.hilarymason.com/blog/help-im-the-first-data-scientist-at-my-company). Facebook is tapping into 2.7 billion 'likes' per day, Rypple (now Work.com and part of Salesforce.com) wants to 'socialize' employee reviews, Glympse is offering video sharing with short-time-to-live videos and Gymball wants to offer a framework for 'interest sensing' using for example geo-fencing. The innovation is even bleeding into physical products. There are self-learning and interactive home thermostats, interactive coffee tables and a Pandora musical chair. But what is the risk of all this? Is it all good fun, or could things go wrong?

Here Mr. Plakias introduced a four-quadrant model, with the different uses of data and potential risks or 'things to look out for' when pursuing data oriented innovations. In the model he combines 'where the data comes from' (1st party or 3rd party) and if the use is real-time/transient or 'persistent'. In the model you can find what a specific quadrant covers (for example traditional 'content' or 'behavior') and it shows the potential downside or risk ('creepiness' or 'unintentional exposure') when doing it wrong.
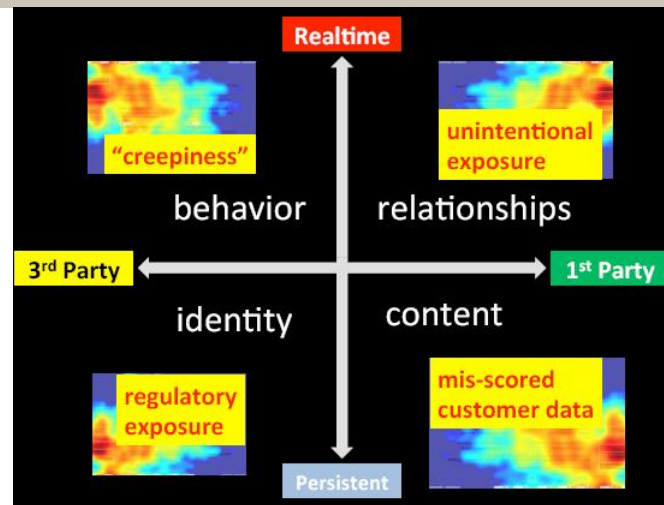
Fascinating Future - Jason Silva
Film maker and Futurist

Jason Silva brings energy to a room, as he dazzles the audience with a speech and 'visual espresso' videos which combine concepts, introduce far reaching hypotheses and ask thought provoking questions. He speaks of how exponential growth in technology is sling-shotting us into the future. Where technology will augment what it means to be human. Because that is the interesting question: what will it do to the human condition, as people are afraid of change.

Big Data is surely going to affect our business, it will improve the game. We never had so much data about everything we do. Data which could give us great new insights into who we are. We can create visual arts of the flight patterns of airplanes, which looks a lot like brain patterns, human migrations look like weather maps. We are free agents, but in a group you see the patterns.

And these are recurring patterns, patterns of patterns – the model of Internet is like the brain, technological systems looking like natural systems. Technology that serves as an expander of human consciousness. Big Data can serve as a global nervous system. These technologies are expanding exponentially, and 'exponential' is a concept that people don't intuitively grasp. For example, any smartphone today is cheaper, better and faster than the very best computer 40 years ago. Someone in Africa now has better access to information than the United States president had 25 years ago.

People worry about privacy, but there is nobody behind the curtain spying on the individual person. We are all part of a global tapestry. Perhaps Big Data doesn't belong in one company, but should be shared among companies. I love the concept of Radical Openness. The free exchange of information, where ideas can have 'sex'. Ideas are just as real as the neurons that create them. They leap, infect. They are the new replicators and they are undergoing evolutionary change. Here, Big Data doesn't just provide us

with a big map, it also gives us opportunities to tweak the map. Our mind exists in a feedback loop between mind, the outside world and our behavior. Now think of Big Data and the age of biology, of software that can write hardware. What could happen?

People have this fear that there are these big bad companies that will have all this data about us, but technology always has good and bad. And yet, the world is moving forward by all measures. I believe people will adapt, and we'll take the opportunity to know ourselves, to reconnect. In that respect I'm a rational optimist: I don't see the risk, I don't see the man behind the curtain who will use it for evil.

Big Data. New Physics - Jeff Jonas
Chief Scientist at IBM Entity Analytics

Jeff Jonas has a primary focus on making sense of data streams but with special attention towards privacy and civil liberties protections. His key position is that for us to deal with an ever increasing amount of data, the data must find the data and the relevance must find the user. This sounds somewhat Zen-like, but turns out to be fairly practical.

The fact that we have a problem is pretty obvious: the available observation space is exploding, thanks to many new data streams, but the sense-making algorithms are only evolving slowly. The effect is enterprise amnesia: we could know things, but we don't.

To look at the mechanics of Big Data, Mr. Jonas used a jigsaw puzzle metaphor. He even went so far as to really mix up some jigsaw puzzles to record how real people would deal with the complex task of sorting through many unconnected, duplicate, incomplete puzzle pieces. Anyone who's ever made a jigsaw puzzle knows how to do it: find similar

pieces, link the borders, sort by color etc. Sometimes new observations reverse earlier assertions and some observations produce novel discovery.

Many of the observations from making jigsaws apply directly to the world of Big Data: connecting more data together actually makes it easier to connect. By combining sources, we can build a more complete view of people and distinguish one from another. More data makes for better predictions and makes it easier to 'compute'. Mr. Jonas used the example of counting people: are you seeing one person with five accounts or are they five accounts of individual people? Or in case of illnesses: is it one case reported several times, or is a true epidemic starting? And while counting sounds like a trivial exercise, it lies at the basis of patterns and predictions: if you cannot count well, you cannot predict well.

When trying to count, more data sources help: from name, city and date of birth to type of car, IP address used etc. If we can map all data, we can try to find out if our multiple records actually link to one person. This is very relevant in many situations, not in the least to prevent fraud and crime. One dimension that is of ultimate importance, but often still overlooked: location. One thing (or person) cannot be in two physical locations at the same time. Also, two different things cannot occupy the same space at the same time. This means that 'when' and 'where' are the most important data!

Mr. Jonas calls it Space-Time-Travel and states that it is the ultimate biometric. Based on space and time, you can predict with 87% reliability where you will be next Thursday at 5:57 PM. Or most likely you can for example list off the names of the top 10 people you co-locate with in places that are not at home and not at work. The consequences will be enormous: as a huge opportunity, unraveling secrets, challenging all existing notions of privacy. And it's here now: the data is available.

But, to make it work, data has to connect with data. Whenever new information comes available, any connections should be made automatically and, secondly, relevance should

find the user: interesting observations should be noted and fed back to the decision maker, who then, through deep reflection can redirect and look for new patterns.
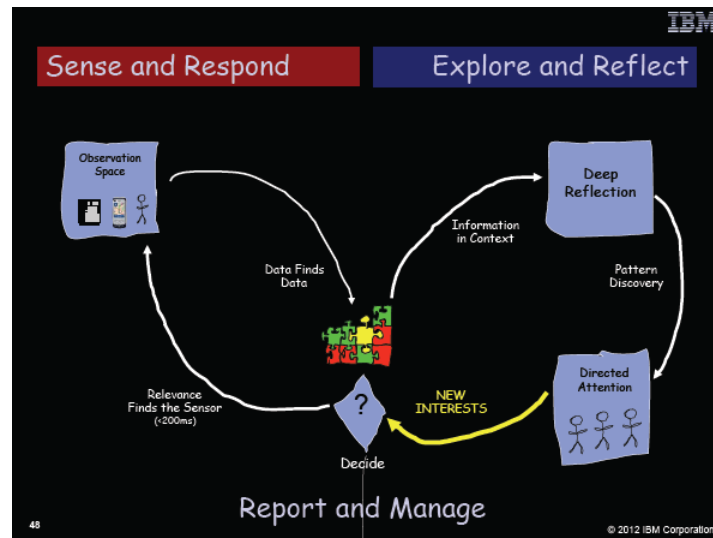
Finally, we have to get fantasy out of analytics:

- Often, the contemplated observation space is insufficient. If we don't tap enough data sources, some of the expected insights will never be realized.
- Often, it is necessary to re-focus an organizations collection interests. We may need to focus more broadly, in different areas or from different sources to complete our puzzle.
- Doing this within the policy and law while keeping adversaries in the dark can be tricky.

Conclusions - Pierre Hessler
Chairman Delegate at Capgemini Board

Pierre Hessler closed the event: "This has been an extremely interesting event, and it's been great to be together with all these people, speakers and audience. The program has brought us from science to art and all the way back to science again, intermixed with business. Very enriching!

I have a four point conclusion (which is one more than usual):

If anyone still doubts if this is real – that would be the wrong conclusion. Big Data has been exploding in our faces. There are interesting new aspects of big data such as the temporality or time-space-boxes. There is the notion of structuring unstructured data. The reality of big data is tremendous. We heard of ways on how to start, like the

example of how you can now know about when competitors are launching new products or opening new locations, which can be a small and pragmatic way to start.

When thinking big data, we should not forget small data, the more traditional BI. The way we treated small data is not very promising for how we will treat big data. We need to find a way to correlate small data and big data initiatives in most useful way. This may not be of great interest to big thinkers, but it is of great urgency for us and our companies.

Every big phenomenon triggers counter movements. There will be a big drama in the coming years. Large organizations will have the power to draw conclusions from big data, which some of us will not like. Some will start a countermovement. To simply wave away any worries by saying 'this will all be for good' is in fact quite funny, because in reality Big Data gives companies the ultimate weapon to make more money. There is no proof that such power has even been constrained by philanthropic reasons. Companies will use big data for themselves only. This suspicion is universal. So let's aim to not forget the value, rules, ethics and acceptable states. We should urge all of us to not throw that away, just because we see the possibilities of big data, because if we do the counter movement could be all the more powerful and could destroy the trust in the company. Hard to build, easy to destroy.

I believe we have talked about Big Data mainly as consumers, talking about phone data, consumer data. Yet what we shouldn't forget is that **one of the things we as companies can do is generate data for our own purposes**. E.g. a A consumer electronics company's toothbrush with sensors in it: they will sell millions of toothbrushes, which will be data generators. Not just on when they will be used, but on data on mouth in general: how you brush, how dentists see how you brush, how your children will brush. A company like that will be in a position to exploit big data with the communities they work with. It will change how that company works, and not just in the obvious ways but also by providing data on which parts of the toothbrush design are prone to failure, as direct input into the development of the company, their manufacturing, logistics, etc.

Let's not forget that Big Data is only one of the phenomena that urge us to create digital enterprises. The digital enterprise is not just exploiting the fruit of big data, its changing everything the company does, every process, and ultimately the very

position, the role the company has. And we, the CIOs have to have a fitting role, because there is no transformation without the CIO claiming a big part in it. Previously he was consulted in a supportive role. Business was looking at IT as a handicap to transformation. Now this has changed, as the CIO is getting into the driver's seat of transformation. The CIO is no longer the one merely answering questions about technology. No, he's asking questions on the use of technology. This makes the CIO an integral part of executive team. And as part of the executive team, the CIO will realize that the executive team is not interested in the *recorded* future; it's interested in SCRIPTING their future."

Appendix – Vendor Relationship Management Workshop
Can you use VRM to solve the privacy issues around Big Data?

After the main event, attendees were invited to take part in an optional workshop on the topic of Vendor Relationship Management, and how it could be a solution to the privacy issues that surround Big Data.

Vendor Relationship Management explained

Marcel van Galen, CEO of Qiy (pronounced as 'key'), introduced the concept of VRM and also showed how Qiy is already serving customers with an offering that is somewhat similar to VRM. Doc Searls, later in the workshop, added some of his thoughts and provided a bigger context: over time, we see silos crumble, a decline in hoarding of data and an increase in openness. It is in this context that VRM makes perfect sense and isn't some odd theoretical invention. VRM in concept is fairly straightforward even thought the technical implementation (and legal ramifications) may prove quite complex in some instances. The idea is that any individual is the owner of his or her own data, and can decide if and how to share data with others. The owner would know whenever someone had accessed or used the data. The owner could also change the data and give and

revoke access to the data and use the data to approach vendors. In the ideal situation, anyone willing to buy something (or subscribe to a service, go on vacation, rent a house, etc) would share the minimum amount of data necessary for vendors to come back with a proposal. In this scenario, the user would not even need to share name or address or any identifiable information to receive a quote. In a sense, it would be a reversal of the current model where vendors announce their wares and individuals have to shop around.

True VRM may not yet exist, but first glimpses are starting to be visible. Perhaps Facebook, Google+, LinkedIn or any other social service could become the central place to manage all your data and share it with others. It could become your personal data dashboard!

The Workshop

In the workshop, the discussion was centered on two questions: 'would VRM be attractive for your customers?' and 'are you willing to turn over data back to your customers?' During the discussion, a lot of remarks focused on the barriers and problems that might be related to VRM, even while acknowledging that the concept could be attractive to consumers if done right. Is this a sign of Cold feet? Fear of change? Or are there some kinks to be straightened out before VRM can really take off?

Below are some highlights of the insights that were shared during the discussion:

"What would happen if someone were to steal data related to my bank account in the case of VRM? Today, if a bank's site fails, it is clearly the fault of the bank. In the new situation, who would then be responsible or liable? Would ownership of data also **increase our personal risk** of something happening to our data?"

"There is little confusion about 'What' is VRM, but **the confusion starts when we talk about the 'how'**. How could it work? How would the technology work? How would companies play into this without losing something that is now very valuable to them?"

"Today, there are clear **regulations that would prevent** some of the data to be handed back to the customer. For example any bank has to legally keep your records for seven years"

"The data that our company would want to give back to the customer is not very interesting. The data that I don't want to give back IS interesting, but it's my competitive edge. I'm willing to share back the data that someone originally gave us, but **not the analyses and extrapolations** I did on top of the data: not all data about you is yours"

"How would you define personal data? **Is all data personal?**"

"Ultimately there could be two reasons for a company to move to VRM: Either it is demanded through **legislation** or it will be up to the market: if it could be a **competitive advantage** to give the consumer control, companies may choose to go there."

"Would it really change anything? With VRM it seems as if a new third party would maintain my data in my VRM data-store and as a consumer **I would now have to trust *them* not to abuse my information**. Or could it be implemented without one central data hub?"

"What about hacks? It feels a bit as if I'd be putting **all of my eggs into one basket**. Someone only has to hack into one place to have total access to all my life"

"It might work if there were such a thing as a **clean sheet** or fresh start, but as long as there are companies 'outside' the VRM, it wouldn't help. It would only work if every company is part of it, otherwise it would simply be one extra place to maintain my data"

"What would be the **economic model**, in the middle of a recession? Is it valuable enough for the customer to warrant the extra investments?"

What to do?

"**Keep an eye on what is happening**, as this may start to move quickly"

"In the next 12 months, **understand how the personal data store system could work**, and explore what would be needed to prepare the technology to be capable to give information back to the customers"

"An idea: do not start with your customers but **start with your employees**, as they have a pre-existing relationship built on trust"

"**Be pro-active**: as soon as you can give data back, give it back to your customers without being asked"