InImpact: The Journal of Innovation Impact : ISSN 2051-6002 : http://inimpact.innovationkt.org
Special Edition on Innovation through Knowledge Transfer : Vol. 5 No.1 : pp.146 - 156 : inkt13-022

From Linear to Crowd Innovation - Facilitating the process by compiling and processing hard and soft big data through social media

Jens Rønnow Lønholdt*, and Borivoje Boskovic**

**Corresponding Author, LYCEUM Innovation and Process Consultancy,*
lonholdt@lyceumconsult.dk
***M2-Fundraising*

Abstract. *This paper addresses the issue of Big Data in relation to knowledge transfer and the new concept of Crowd Innovation. The internet has made a huge amount of data available, which is exponential increasing. Different systems are already available for handling the colossal amount of data. However within the field of research driven innovation we do not yet see the full benefits of this. Consequently it is the aim of this paper to present a strategic perspective on the use of Big Data, when planning and implementing a research driven innovation process. The paper will discuss what is considered as respectively Small, Big, Hard and Soft Data. Furthermore, it will present cases in which the intelligent use of big data is put to beneficial use. The paper will strategically outline where and how the full innovation process from idea to market can benefit from an intelligent use of a combination of hard and soft big data. Finally, the paper will discuss and outline the perspective of how research-based innovation could be moving from push to pull and finally reach the concept of crowd-based innovation. Crowd-based innovation understood as the comprehensive and structured use of social media as important platforms for interaction, exchange of ideas and development of products. There are already cases exemplifying this new approach, which we could call Version 3.0 of research driven innovation.*

Key Words: Big Data, Hard and Soft Data, Innovation Processes, Demand-Driven Innovation, University Research Driven Innovation, New Innovation Concepts, Crowd Innovation

1. Background and Introduction

The internet and especially the ever increasing data harvesting facilities and systems have made a huge amount of data available. The amount is constantly increasing covering the public as well as the private sector. The health sector, traffic planning and marketing are all recent examples of areas where structured compilation and processing of big data have been used to benefit of a wide variety of users and customers. Still, a cursory screening of the internet shows that innovation processes are not widely involved in the concept of structuring big data. However, these processes could benefit from this with regard to hard and soft data and all stages of the innovation process; from idea and research to conceptual design, prototyping, product development and market maturing. This paper will discuss in overall strategic terms what is considered as respectively Small, Big, Hard and Soft Data. Furthermore, it will present cases in which the intelligent use of Hard Big Data and a combination of Hard and Soft Big Data are put to beneficial use.

Based on this, the paper will strategically outline where and how the full innovation process from idea to market can benefit from an intelligent use of a combination of hard and soft big data. Finally, the paper will outline the perspective of how research-based innovation could be moving from push to pull and finally reach the concept of crowd-based innovation. This will be based on personal experiences of the authors of this paper and a trend that has recently emerged on the internet.

Crowd-based innovation can be seen as using a Version 3.0 for innovation. It includes social media as important platforms for interaction, exchange of ideas and development of products. Knowledge, experience and expertise of numerous resource organisations and individuals are crowded in the Cloud for the benefit of a fast and efficient demand-driven innovation process.

2. What is Big Data and what is Small Data?

A variety of definitions of Big Data covering very diverse fields can be found on the internet including the following (see List of References for details): Douglas, Laney (2012); Search Cloud Computing; PC Magazine; Kate Kaye (2013); Srinidhi Melkote; IBM definition; TechAmerica (2012); White Paper: Integrate for Insight" Oracle; McKinsey Global Institute (May 2011).

The definitions naturally vary depending on the circumstances in which they are applied. However, a common feature runs through all these definitions: it is not so much a question about tetra bites, the capabilities of servers or the speed of the internet connections that characterises big data since all of these develop rapidly over time. What characterises Big Data is that they require new forms of compilation, processing and especially interpretation as they represent high volume, high velocity and high variety information assets. In this connection the veracity of the information is of crucial importance for decision makers. The veracity of data and information has always been a point in question. However, Big Data put a lot of focus on this as the number of sources and the amount of data makes it virtually impossible to apply traditional quality assurance systems as is done for Small Data. Consequently there is a need to develop veracity assuring mechanisms that focuses on Big Data in particular.

There are also a number of definitions of Small Data on the internet including the following (see List of References for details): Mark Whitehorn (2013); and Courtney Kettmann (Jan. 2013). Again the definitions vary but the common feature is that Small Data are very reliable data with a high veracity because the source is usually well known. The data can be checked and comprehensive quality assurance systems can be applied. Consequently, a way to deal with Big Data could be to segment them into a cloud of well-defined Small Data by means of computer software and algorithms.

3. What is Hard Data and what is Soft Data?

Hard Data is what is quantifiable and thus could be measured as one value. Hard data is data in the form of numbers or graphs. Soft Data is only qualifiable and encompass meanings, opinions, attitudes, anecdotes, stories, rumours, feelings, and is consequently very subjective and can change over time and from location to location. It can be measured and made into statistical information by different kinds of surveys, the result of which is very dependent on the formulation of the questions. Consequently, it is well known that the same survey can give completely different results depending on the survey company and especially the questions formulated. A good example is political surveys.

Today, the *Interaction and Individualism Society* has taken over evolved from the well know *Information Society in the beginning of this century*. This new society is widely facilitated by the social media and consequently soft data has become increasingly important, and hard data is losing ground. Mintzberg is a strong advocate of the use of soft data as he talks about the *soft underbelly of hard data*. The following is adapted from Mintzberg (1994:257-66):

- Hard information is often limited in scope, lacking richness and often failing to encompass important non-economic and non-quantitative factors. Much important information never becomes hard fact. The expressions on a customer's face, the mood in the factory, the tone of voice of a governmental official. Hence, while hard information may inform the intellect, it is largely soft information that builds wisdom.

- Much hard information is too aggregated. The problem is that a great deal of information is lost in such aggregating.

- Much hard information arrives too late. Information takes time to "harden" – time for trends and events and performance to appear as "facts" more time for these facts to be aggregated into reports.

- Finally, a surprising amount of hard information is unreliable. Something is always lost in the process of quantification – before those electrons are activated. Anyone who has ever produced a quantitative measure knows just how much distortion is possible, intentional as well as unintentional.

Of course soft information can be problematic too. It can be speculative, distorted, and so forth. But that only highlights the key point here: that all information must be scrutinised carefully. The danger with hard information is that once information becomes hard such as statistics, it tends to acquire "the authority and sanctity of Holy Writ" (Devons, 1950:155).

The following part is adapted from the TechAmerica report (2012), which include good examples of the beneficial use of hard data.

4. Examples where Processes Benefit from the Intelligent Use of Hard Big Data

4.1 National Archive and Records Administration (NARA)

The National Archive and Records Administration (NARA) have been charged with providing the Electronic Records Archive (ERA) an Online Public Access System for U.S. records and documentary heritage. As of January 2012, NARA is managing about 142 terabytes (TB) of information (124 TB of which is managed by ERA), representing over 7 billion objects, incorporating records from across the federal agency ecosystem, Congress and several presidential libraries, reaching back to the administration of George W. Bush. It sustains over 350 million annual online visits for information. These numbers are expected to dramatically increase as agencies are mandated to use NARA as of 2012.

In addition to ERA, NARA is currently struggling to digitize over 4 million cubic feet of traditional archival holdings, including about 400 million pages of classified information scheduled for declassification, pending review with the intelligence community. Of that backlog, 62% of the physical records stored run the risk of never being preserved.

The NARA challenge represents the very essence of Big Data: How does the agency digitize this colossal volume of unstructured data, provide straightforward and rapid access, and still effectively govern the data while managing access in both classified and declassified environments? NARA has adopted an approach that put it on the path to developing the Big Data capability required to address the challenge. This approach combines traditional data capture, digitizing, and storage capabilities with advanced Big Data capabilities for search, retrieval, and presentation, all while supporting strict security guidelines.

4.2 Vestas Wind Energy

Since 1979, this Danish company has been engaged in the development, manufacture, sale, and maintenance of wind power systems to generate electricity. Today, Vestas installs an average of one wind turbine every three hours, 24 hours a day, and the turbines generate more than 90 million megawatt-hours of energy per year. This is enough electricity to supply millions of households. Making wind a reliable source of energy depends greatly on the placement of the wind turbines.

For Vestas the process of establishing a location starts with its *Wind Library*, which combines data from global weather systems with data collected from existing turbines. Data is collected from 35,000 meteorological stations scattered around the world and from Vestas's turbines. The data provides a picture of the global scenario, which in turn leads to mesoscale models that are used to establish a huge Wind Library that can pinpoint the weather at a specific location at a specific time of day.

Vestas uses one of the largest supercomputers worldwide, along with a new Big Data modelling solution, to slice weeks from data processing times and support 10 times the amount of data for more accurate turbine placement decisions. This reduces information response time by 97%, lower the cost per kilowatt, reduce IT costs and decrease energy consumption by 40%.

4.3 NASA Johnson Space Centre

NASA Johnson Space Centre (JSC) manages one of the largest imagery archives in the world and has provided industry and the public with some of the most iconic and historic human spaceflight imagery for scientific discovery, education and entertainment. NASA's imagery collection of still photography and video spans more than half a century: from the early Gemini and Apollo missions to the Space Station. This imagery collection currently consists of over 4 million still images, 9.5 million feet of 16mm motion picture film, over 85,000 video tapes and files representing 81,616 hours of video in analog and digital formats. Eight buildings at JSC house these enormous collections and the imagery systems, that are growing exponentially, and its sheer volume of unstructured information is the essence of Big Data.

NASA has developed best practices through technologies and processes to: (i) Comply with NASA records retention schedules and archiving guidance; (ii) Migrate imagery to an appropriate storage medium and format destined for the National Archives; and (iii) Develop technology to digital store down-linked images and video directly to tape libraries.

In the following, examples show where the structured use of big data could be beneficial for processes and their end results based on a demand-driven and customer-oriented innovation processes.

5. The Potential Use of Big Data

5.1 Healthcare Quality and Efficiency including Early Detection

Politicians and public servants struggle constantly to improve quality and efficiency in the delivery of healthcare while reducing costs in these times of economic crises. But in a different light it can be seen as a significant opportunity to improve the lives of millions at a lower costs if the concept of Big Data is applied. The increased use of electronic health records (EHRs) coupled with new analytic tools presents an opportunity to mine information for the most effective outcomes across large populations. Using carefully de-identified information, researchers can look for statistically valid trends and provide assessments based upon true quality of care.

It may involve using sensors in hospitals or homes to provide continuous monitoring of key biochemical markers, performing real time analysis on the data as it streams from individual high-risk patients to an analysis system. The analysis system can alert specific individuals and their chosen health care provider if the analysis detects a health anomaly, requiring a visit to their provider or a “911” event about to happen.

5.2 Transportation

Big Data has the potential to transform transportation in many ways. Traffic jams are the nemesis of many drivers. They waste energy, contribute to global warming and cost individuals time and money. Distributed sensors on handheld devices, on vehicles and on roads can provide real-time traffic information that is analysed and shared. This information, coupled with more autonomous features in cars can allow drivers to operate more safely and with less disruption to traffic flow. This new type of traffic ecosystem coupled with intelligent cars has the potential to transform how we use our roadways.[1]

5.3 Education

Big Data can have a profound impact on education and the competitiveness of individual countries in the global economy. For example, through in-depth tracking and analysis of on-line student learning activities researchers can ascertain how students learn – down to the level of mouse clicks. This can help develop approaches to improve learning. This analysis can be done across thousands of

[1] http://www.forbes.com/sites/toddwoody/2012/09/19/automakers-on-the-road-to-self-driving-cars/

students rather than through small isolated studies.[2] Courses and teaching approaches, online and traditional, can be modified to reflect the information gleaned from the large scale analysis.

5.4 Fraud Detection – Healthcare Benefits Services

Big Data can transform improper payment detection and fundamentally change the risk and return perceptions of individuals that currently submit improper, erroneous or fraudulent claims. This challenge is an opportunity to explore the benefit of applying Big Data technologies and techniques, to perform unstructured data analytics on medical documents to improve efficiency in mitigating improper payments. The benefit is that the culture of submitting improper payments will be changed. Big Data tools, techniques and governance processes would increase the prevention and recover value by evaluating the entire data set and dramatically increasing the speed of identification and detection of compliance patterns.

5.5 Cyber Security

Government agencies face severe challenges associated with protecting themselves against cyber-attacks. They have to manage the exponential growth in network-produced data, database performance issues and the complexity in developing and applying analytics for fraud in relation to cyber data. Agencies continue to look at delivering innovative cyber analytics and data intensive computing solutions. However, cyber intelligence and other machine generated data are growing beyond the limits of traditional database and appliance vendors.

Agencies are looking to incorporate multiple streams of data to benefit both human and automated analysis. Cyber data such as host, network, and information from the World Wide Web, are being fused with human oriented information such as psychosocial, political, and cultural information to form a more complete understanding of our adversaries, motives, and social networks.

5.6 Fraud Detection – Tax Collection

By increasing the ability to quickly spot anomalies, government collection agencies can lower the tax gap – the difference between what taxpayers should pay and what they pay voluntarily – and profoundly change the culture of those that would consider attempting improper tax filings. Most agencies practice a *pay and chase* model, in which they accept returns and often pay out tax refunds, and only ex post facto review a sampling of returns in order to reveal unintentional or intentional underpayment. Big Data offers the ability to improve fraud detection and uncover non-compliance at the time tax returns are initially filed, reducing the issuance of questionable refunds.

5.7 Weather

The ability to better understand changes in our weather and climate can benefit millions of citizens and thousands of businesses including farmers, tourism, transportation and insurance companies. The Vestas case presented above is also an example. Weather and climate-related natural disasters result in tens of billions

[2] http://www.nytimes.com/2012/03/29/technology/new-us-research-will-aim-at-flood-of-digital-data.html

of dollars in losses every year and affect the lives of millions of citizens. Much progress has been made in understanding and predicting weather, but it's far from perfect. New sensors and analysis techniques hold the promise of developing better long term climate models and nearer term weather forecasts based on the concept of Big Data.

In the following is given examples of where the combined use of hard as well as soft big data can benefit processes and the end result.

6. Examples where Processes Benefit from the Intelligent Use of a Combination of Hard and Soft Big Data

6.1 Royal Institute of Technology of Sweden (KTH)

Researchers at KTH, Sweden's leading technical university, wanted to gather a wide array of real-time data that might affect traffic patterns in order to better manage road congestion. This real-time sensor data includes GPS from a large numbers of vehicles, radar sensors on high ways, congestion charging, weather and visibility etc. The challenge was how to collect, compile and process this wide variety of data at high velocity and assimilating it in real time for analysis.

Collected data is now flowing into a commercial off-the-shelf (COTS) Streams Analytics software. This is a unique software tool that analyses large volumes of streaming of real-time data, both structured and unstructured. The data is used to help intelligently identify current conditions and estimate how long it would take to travel from point to point in the city, offer advice on various travel alternatives, such as routes, and eventually help improve traffic in a metropolitan area.

The result has been a decrease in traffic congestion and accidents in the target cities. KTH is now looking to expand the capability to support routing of emergency services vehicles.

6.2 University of Ontario, Institute of Technology

The rapid advance of medical monitoring technology has done wonders to improve patient outcomes. Today, patients are connected to equipment that continuously monitors vital signs, such as blood pressure, heart rate and temperature. The equipment issues an alert when any vital sign goes out of the normal range, prompting hospital staff to take action immediately.

However, many life-threatening conditions do not reach critical level right away. Often, signs that something is wrong begin to appear long before the situation becomes serious, and even skilled and experienced nurses or physicians might not be able to spot and interpret these trends in time to avoid serious complications.

The University of Ontario's Institute of Technology partnered with researchers from a technology firm that was extending a new stream-computing platform to support healthcare analytics. The result was *Project Artemis* - a highly flexible platform that aims to help physicians make better and faster decisions regarding patient care for a wide range of conditions. The earliest iteration of the project is focused on early detection of nosocomial infection by watching for reduced heart rate variability along with other indications. Project Artemis is based on Streams Analytic

Software. An underlying relational database provides the data management required to support future retrospective analyses of the collected data.

7. Research Driven Innovation: From Push to Pull to Crowd Innovation

University knowledge was previously transferred to or exchanged with the surrounding society following the so-called push model. This meant that in relation to the need of the society, universities and research institutions by various means and measures *pushed* knowledge out (see Chart 1 below). As given in Lønholdt, J.R. et al (2012) it is now widely recognised that the most efficient form for knowledge transfer or exchange is the so-called *pull* based model. Here, the process is demand-driven determined by the need for knowledge articulated by the consumer. See Chart 1 below that illustrates (1) the more inefficient broad push model that is less targeted and based on undefined needs and (2) the targeted and narrow pull and demand-driven joint exchange model based on defined needs.

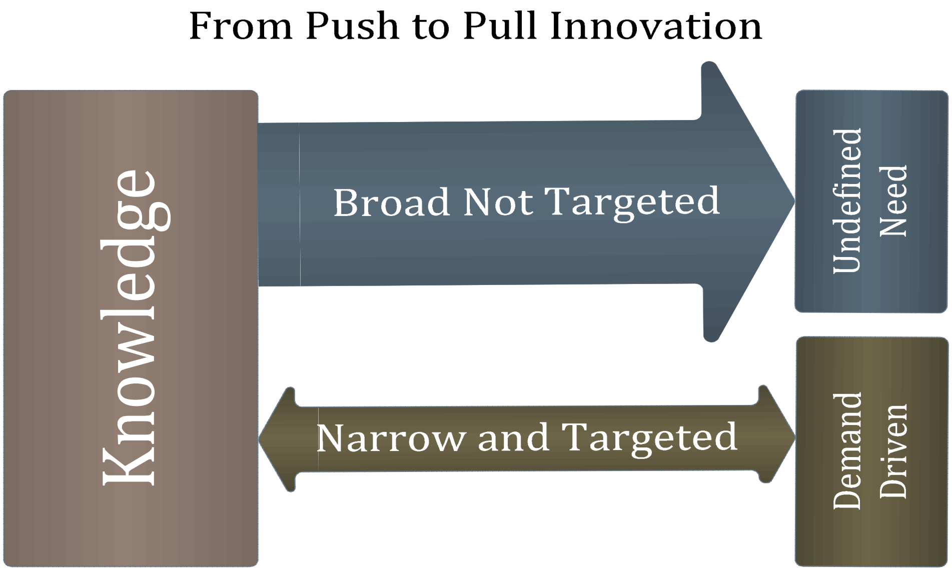

Chart 1: From Push to Pull Innovation

However, as society has developed from the so called *Information Society* into the *Interaction and Individualism Society* a new and even more efficient and especially fast knowledge transfer and exchange model has emerged. It is a combination of both technological *and* social development. The development of the capacity and numbers of soft- and hardware *and* the development of social media speed up the possibilities and ability for interaction. This is supported by the more individualistic culture in our society that has developed over recent years. The bottom line is that we have gone from the linear push and pull models to a Version 3.0 that is not linear. It could be called *Crowd Innovation* as it is non-linear and based on interaction in the *Cloud* through numerous social media and by using laptops, smart phones and tablets. And it includes a fertile combination of hard and soft big data.

As can be seen from Chart 2 overleaf, research based innovation version 3.0 is born in the Cloud where needs, knowledge, expertise and experience are crowded and clustered and out of this merger comes new ideas, new innovation, new products, new processes etc. The interaction and connection lines in this cloud/crowd universe are all kinds of professional and personal social media including and more to come: LinkedIn, Facebook, Twitter, Skype, Instagram, YouTube, and TED Presentations. This is slowly replacing formal brainstorming meetings, professional conferences, workshops and seminars, and not least it makes discussions about Intellectual Property Rights (IPR) obsolete in most cases.

It is worth to note that the ones involved in the interaction are a wild diversity of serious professionals and happy amateurs. They interact with the one goal of jointly drive an innovation process from idea to market. Including finding the funding of the processes, called Crowd Funding, where a number of platforms are available on both national and international levels raising funds from 1,000 to 5 million € based on a sort of gift concept for the ones contributing to the funding.

The corresponding author of this paper is in the middle of a fairly large innovation project based on the new concept of crowd innovation and the preliminary results are very promising. The findings so far show that it significantly speeds up the innovation process and widely broaden the resource base for the project. At present, funding is raised through venture capital but crowd funding is under consideration pending the outcome of the first phase of the project.

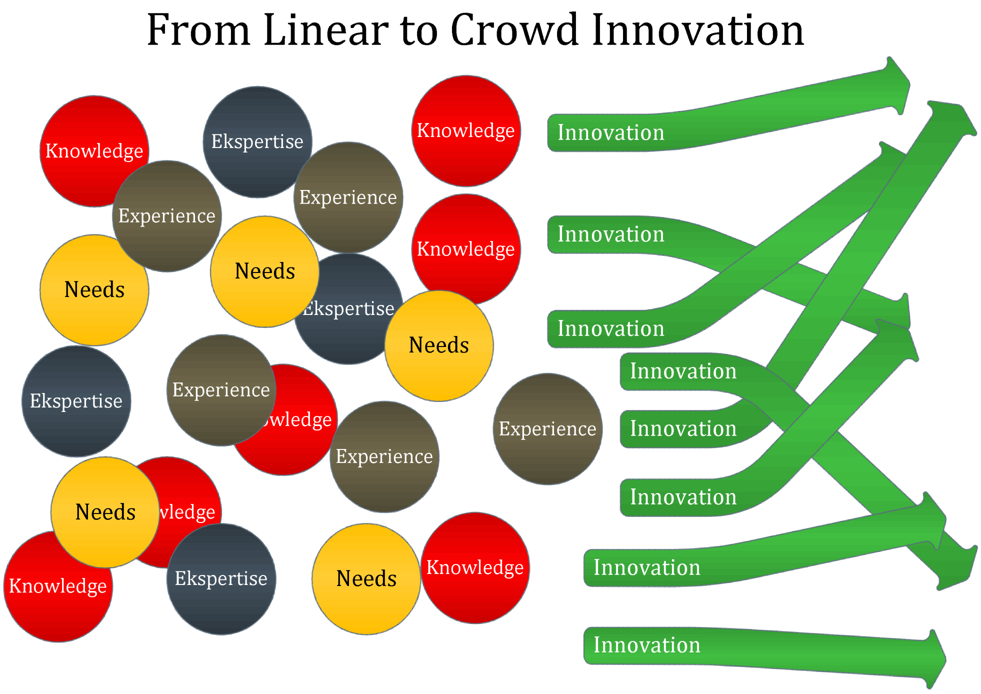

Chart 2: From Linear to Crowd Innovation

8. Discussion and Conclusion

This is not an academic paper on the subject as it is not based on research or case studies conducted by the authors, but has followed the basic principles of Crowd Innovation where knowledge, experience and expertise of resource persons has be harvested from the internet (see the List of References). However, it started as all good research starts, with a question and a certain curiosity related to one of the core expertise of the authors, namely university research-driven innovation. The simple question was: Why has the concept of Big Data and in particular the concept of the combination of Hard and Soft Big Data not been applied more widely for this type of innovation process? As shown above, it has been applied successfully within other areas.

Based on this, the intention of the authors was to give a strategic perspective on this question and hopefully identify where and how in the innovation process combined big data could be applied. In this connection our goal was to sketch a variety of *Google for Innovation* where all these big hard and soft data could be compiled and processed for the benefit of innovation processes. However, again as all good research, it took another turn. It became increasingly clear for the Authors that a new concept of innovation has developed during the last couple of years where social media are the engines and individualism the fuel. A process that is far from the well-known and comfortable linear process we are used to. Instead, it is dynamic, network-oriented, and total unpredictable.

We do not claim to be the first to discover or identify what we call *Crowd Innovation*. However to us, it was a kind of *Snow-White-Awakening*. Our screening on the internet did not reveal a more professional and comprehensive discussion of this. It is just happening. This Snow-White-Awakening was even stranger for us as we could see that without knowing we had actually applied the concept of Crowd Innovation to our most recent innovation projects, as well as the writing of this paper.

Consequently, what we found through the writing of this paper was that a new kind of innovation process has developed over the last couple of years and it is not based on academic research or professional review processes. Instead, it is a natural evolution process made possible by the internet and especially social media in perfect tune with our society of individualism. We are fully aware that a number of research studies have been done and are on-going with regard to important aspects of this. However, we have not found comprehensive work on the concept we have named *Crowd Innovation*. We are convinced, also based on our own recent experiences, that this is the future of innovation.

Therefore, we strongly recommend that this phenomena or concept be more professionally researched in the future and included as a key issue in coming INKT Conferences. In this connection it could be that we should reinvent the form and scope of coming activities and conferences with the KES-IKT family taking them into the Cloud and run them based on a Crowd Concept.

9. List of References

1. Lynch,C. “Big data: How do your data grow?” Nature, vol. 455, pp. 28–29, Sep. 2008.
2. “Oracle Big Data Appliance.” [Online]. Available: http://www.oracle.com/us/products/database/big-data-appliance/overview/index.html .
3. “White Paper: Integrate for Insight” Oracle. [Online]. Available: http://www.oracle.com/us/technologies/big-data/big-data-strategy-guide-1536569.pdf

4. "Big Data Analytics | Microsoft SQL Server." [Online]. Available: http://www.microsoft.com/sqlserver/en/us/solutions-technologies/business-intelligence/big-data-solution.aspx
5. "Big Data-Intelligence Begins with Intel." [Online]. Available: http://www.intel.com/content/www/us/en/big-data/big-data-analytics-turning-big-data-into-intelligence.html.
6. "Digging Into Data Challenge | National Endowment for the Humanities." [Online]. Available: http://www.neh.gov/grants/odh/digging-data-challenge.
7. Boyd, D. and Crawford, K. (2012) "CRITICAL QUESTIONS FOR BIG DATA," Information, Communication & Society, vol. 15, no. 5, pp. 662–679
8. Manovich, L. (2012) "Trending: The Promises and the Challenges of Big Social Data," in Debates in the Digital Humanities, M. K. Gold, Ed. Minneapolis, MN: U of Minnesota Press
9. Burrell, "The Ethnographer's Complete Guide to Big Data: Conclusions (part 3 of 3)," Ethnography Matters. [Online]. Available: http://ethnographymatters.net/2012/06/28/the-ethnographers-complete-guide-to-big-data-part-iii-conclusions
10. Boyd, D. and Crawford, K. (2011) "Six Provocations for Big Data," SSRN eLibrary
11. Anderson and Rainie, L. (2012) "The Future of Big Data," Pew Internet & American Life Project, Washington, D.C.
12. Cooley, T., Meizel, K., and, Syed, N. (2008) "Virtual Fieldwork: tree cases studies," in Shadows in the Field: New Perspectives for Fieldwork in Ethnomusicology, Oxford University Press, USA, pp. 90–107.
13. TechAmerica (2012) Demystifying Big Data: A Practical Guide to Transforming the Business of Government. TechAmerica Foundation: Federal Big Data Commission
14. Bollier, D. (2010). The Promise and Peril of Big Data. Aspen Institute, Communications and Society Program
15. McKinsey Global Institute (2011) Big data: The next frontier for innovation, competition, and productivity, by James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh
16. Lønholdt, J.R. et al (2012) Next – Practise in University Research Based Open Innovation – from push to pull: case studies from Denmark. Paper presented at the IKT12 Conference in Bournemouth, April 2012, Bournemouth Great Britain.
17. Whitehorn, M. (2013) Big Data versus small data: Unpicking the paradox
18. Kettmann, Courtney (Jan. 2013) Big v. Small Data… Why marketers today need to know the difference
19. *Douglas, Laney (2012). The Importance of 'Big Data': A Definition. Gartner*
20. *Search Cloud Computing http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data*
21. *PC Magazine http://www.pcmag.com/encyclopedia_term/0,2542,t=Big+Data&i=62849,00.asp*
22. *Kate Kaye (2013) What 'Big Data' Means for Marketers http://adage.com/article/dataworks/data-defined-big-data/239144/*
23. IBM definition, Bringing Big Data to the Enterprise http://www-01.ibm.com/software/data/bigdata/