EXPLORING THE KNOWLEDGE BEHIND BIG DATA ENABLED INNOVATION IN BROADCASTING: THE CASE OF NETFLIX AND BBC

Ruiz Navas Santiago ルイズナヴァス サンティアゴ

Miyazaki Kumiko 宮崎久美子

ABSTRACT

Big data is considered to play an increasingly important role in many aspects of our lives as indicated by the OECD, 2013. To find insights in how to exploit the big data technologies in the least time possible we focus on which kind identify the knowledge implemented in successful big data enabled projects.
This paper proposes a method to identify what knowledge that is being used by companies to exploit big data's potential, as target sector broadcasting is selected, from it we analyze 2 companies that were able to innovate at some level by making use of big data technologies, which are BBC R&D and Netflix.
We on the knowledge words that they were diverse and the knowledge they represent has been previously not used together, which call for action on capability building for companies that are interested to exploit the potential value of big data technologies.

KEYWORDS: Big data, Knowledge, keyword analysis, innovation, broadcasting

1. INTRODUCTION

It started as a government supported activities in early 1900, its mission as first described by David Sarnoff in 1922 "broadcasting represents a job of entertaining, informing, and educating the nation". Time passed this activity is now realized by private broadcasters and public broadcasters, in a strategic sense privates focus on content and business models that bring them revenues and public focus on quality. To be able to entertain, inform and educate, broadcasters communicate and publish information with a value to their user in form of programs which are composed by video, audio and additional data or in a more general fashion content.

The environment for the broadcasting sector is under a significant change and one of the main reasons for this change is been given by technology which is having an effect on the media content value chain and is generating changes in different actors, to mention some examples we have Twitter acquiring the live broadcasting app periscope and Netflix successfully entering the original content creation, actors that where not directly related to activities covered by broadcasters now are offering services broadcasters offer these two examples are two companies that live from data and are known for their ability to use it to offer added value. This fact present challenges as the disruption of the current service offering status quo of the broadcasters, but also opportunities for broadcasters to learn of their use of big data, adapt to change by creating, adapting, acquiring capabilities and respond to these challenges.

Big Data have been present in our society since long ago and the challenges for managing it have been of interest for organizations since they use it in their activities, but now why speak about big data? recently the rate of data creation and its accumulation, real time decision making, non-standardized data are presenting challenges for organizations (McAfee, A., and Brynjolfsson, 2010)(OECD, 2013), but due to the evolution of IT technologies (Jee K, Kim G, 2013) and in response to this requirements, many technologies for aggregate, manipulate, analyze, and visualize data are being implemented and integrated in one term "Big Data" (Manjika, J., et.al., 2011). The term is commonly defined by using the Vs framework the more convenient for the authors is the four Vs, volume, velocity, variety, veracity as characteristics of the data (Raghupathi W, Raghupathi V, 2014), the authors consider this because the 4Vs framework link the characteristics of data to the challenges presented by it. The big data technologies, then, are the methods and technologies required to get value from it (de Mauro A, Greco M, Grimaldi M, 2015). A previous study (Ruiz-navas S, Miyazaki K, 2014) previously revealed many other applications biometrics, smart cities, security systems, agriculture and finance among others. From the existence these applications and their wide range we can argue that these technologies are being used to help humans and systems to do their work better by using available data in their different flavors (Vs of big data) and also brings its challenges like data security, privacy, ownership, transactions, veracity, transportability (Katal A, Wazid M, Goudar R H, 2013)

The big data technologies are recently being studied and as IT are living a rapid change and different new applications, services and startups are created, but one resource all these initiatives have in common is the knowledge that is the base to build them, detecting these particular pieces of knowledge that put together the big data enabled services could be a good start for companies to act on these environmental change and start to take action. Taking into account this research focuses its efforts on discovering these pieces of knowledge that are present in projects related to big data enabled services

2. Big Data for the broadcasting sector opportunities

The broadcasting sector itself is produces and uses data, the media content they produce, transmit, manage is composed of sound, text, video and other kinds of data, the decision making, processes and daily activity related data and other public available sources of data open opportunities for broadcasters to innovate. An overview of some applications was presented by NHK R&D and defining data sources for potential services like Social Networks (SNS), access logs of web services, GPS data from mobile phones and IC card usage logs. These could enable applications as discerning popular trends by using viewer's tweets, track people's movements in emergencies to avoid bad consequences and the combination of different data sources with specific broadcaster's data to personalize online services and create effective advertising and PR methods for business (Nakagawa, T, 2014).

3. Media content

Media content can be live/real time: the one that is reported in real time, news, sports, events, concerts, to multiple receivers and not-live: not happening in real time like DVD, MP3, and the likes and not necessarily is showed to multiple receivers. This environment is being disrupted as main examples we cite the live broadcasting apps periscope and Meerkat which allows anyone with a mobile phone with camera and data connection to become a broadcaster (Boucher J, 2015) and for the non-live content Netflix a company that started renting DVDs, changed its platform to streaming content and now is also providing original content and this company is proud to say that its success is primarily given for its capacity to get value from the data of Its users (Palmer D, 2014).

In general there are different opportunities, applications, technologies, knowledge, data sources involved with the use of big data in the broadcasting sector, in fig 1, we show a little summary of these dimensions. As companies are engaging in the development of the big data for media content services, we see that it brings opportunities to broadcasters and also to other actors associated with the media content business

(Netflix and Hulu for example), new opportunities for them are open and this means new challenges for classic broadcasters, it means a time of change and adaptation, where the development, acquisition and adoption of competences becomes important for firms.

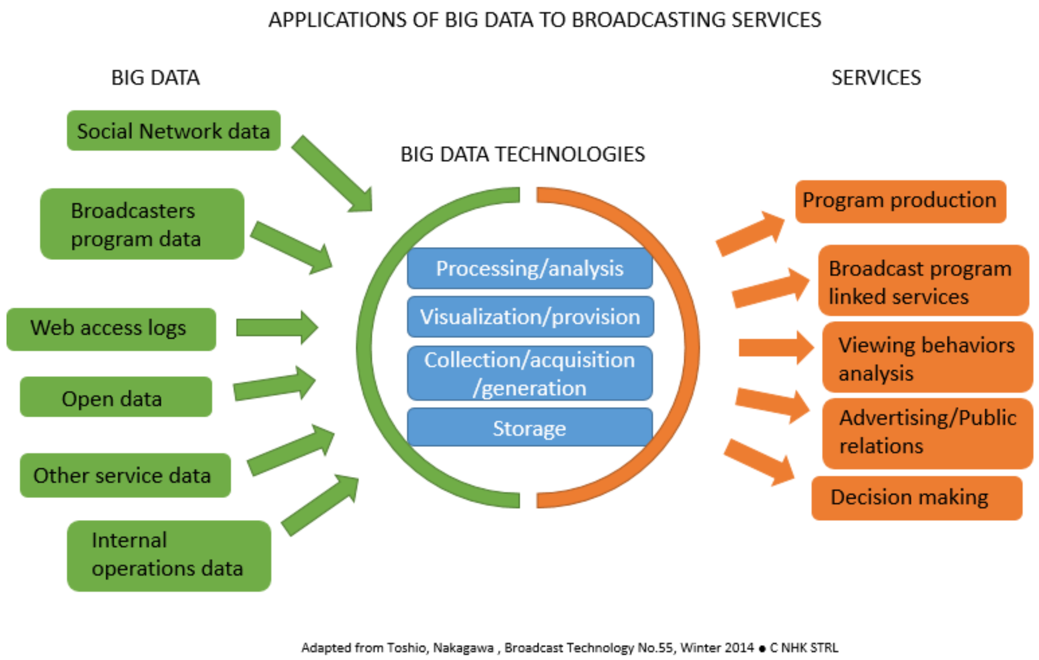

Fig1. Big data and services in the broadcasting sector.

4. THEORETICAL BACKGROUND

4.1. Resource based view of the firm and capabilities

Scholars defined the Resource based view of the firm as a way to explain why some firms were able to arrive to an apparent sustainable competitive position in their markets. The theory found itself on resources as key to competitiveness and these are understood as “all assets, capabilities, organizational processes, firms attributes, information, knowledge, etc. controlled by a firm that enable the firm to conceive of and implement strategies that improve its efficiency and effectiveness” (Barney J, 1991). This theory also signals the importance of managing these resources in times of changes, or Schumpeterian shocks where firm’s strategic positions might change and when resources might change of status and stop being competitive and become regular resources (Barney J, 1991). Other concepts that complement this approach are the definition of the “core competences” to synchronize resources, business units and products (Prahalad CK, Hamel G, 1990), “dynamic capabilities” which are a subset of competence/capabilities that allow the company to adapt to changes in their environment (Teece D, Pisano G, 1994) and the absorptive capabilities that allow the firm to acquire resources from external sources (Cohen W M, Levinthal D A, 1994), to mention some, all which affirm the importance of special types of resources in the firm.

4.2. Knowledge as an Important Resource

Knowledge is part of these key resources for firm competitiveness and adoption to change and even as a branch of the resource based view the knowledge based view of the firm was developed and it proposes that the primary role of the firm is to “integrate the specialist knowledge resident in individuals into goods and knowledge transformation" as the main purpose of the firms (Grant R M, 1996). In his work, Grant, 1996 for practical reasons and to bring the concept to the firm's context, defines two types of knowledge, tacit and explicit. Tacit knowledge refers to know how, cannot be codified and is observed though its application and explicit with knowing about facts and theories and it is revealed by its communication, can be codified. Knowledge also plays and important role on description of complex approaches to explain innovation as in national systems of innovation where the innovative capability of a nation is described by the knowledge exchange interactions between different actors (Patel P, Pavitt K, 1994), as well as in sectoral (Malerba F, 2002) and regional systems of innovation (Cooke P, Uranga M G, Etxebarria G, 1997)

Knowledge is one of the different resources that are important to competitiveness and innovation, also the importance of managing resources in times of change is critical for these two objectives, which leads us to look into the importance of knowledge in the apparent moment of change that the broadcasting sector is facing and how by leveraging from the service and products configuration of other actors, the broadcasters can learn and prepare for their reconfiguration of competences/capabilities in order to achieve competitiveness and innovation.

4.3. Now how to find knowledge?

Explicit knowledge is the codified representation of knowledge, following the stages of the knowledge cycle proposed by Nonoka I and Takeguchi H, 1995, the tacit knowledge after certain stages in the company becomes codified and part of it turns into explicit knowledge, generating more resources and bases for the development of new tacit knowledge, stating a relationship between the two. Normally for research on knowledge scientist usually turn to codified form of knowledge like scientific publications, patent data bases, technical reports and documents, product catalogs, companies data bases, web publications and other different sources or combination of the latter. By using words as a proxy of knowledge some authors have been using these data bases to study knowledge, for example mapping of scientific areas (Avila-Robinson A, Miyazaki K, 2013), science disciplines fusion in nanotechnology (Islam N, Miyazaki K, 2010) by using the words as proxy for to study knowledge also some warnings have been made, as the words can have different meanings according to context (science, jargon, country, application, discipline) which makes the reviewing of great number of words difficult under unsupervised conditions and under supervised conditions to tedious. For knowledge analysis the champion data bases are usually company internal, scientific data bases or patent data bases, yet with the improvement of IT and processing power more complex data sources and analysis techniques are been made available for researchers in different fields (Science, 2011).

Innovation and Management of Technology (MOT) is not an exception, some examples are the use of literature based discovery to link technology and social issues (Ittipanuvat V, et. Al., 2014), using unstructured data for map evolution of technological fields (Jurowetzki R, Hain D S, 2014) and detection of service characteristics and service innovation in the B2B mobile business in Korea (Hong S, Miyazaki K, 2014) all these new potential data sources, techniques and applications are available for the whole community and in order to adequately exploit them we should start explore and appropriate them to our discipline specifics.

Appreciating these studies we agree with these authors and decide that words can be used to represent knowledge, taking care of that the number of words to analyze is workable under supervised conditions, the familiarity with the studied topic and the different meanings that the words might present.

4.4. R&D Projects and Innovation

The competitive resources are the ones that allow companies to develop efficiency or efficacy in their activities, normally in successful R&D projects of companies one of these conditions is present, which means that by analyzing companies successful R&D projects we can identify key resources the companies are implementing in them, so by focusing on knowledge we could find what knowledge are these companies implementing.

5. THE CASE STUDY

We part from the fact that some socio technical phenomena can be modeled by words as explained by Michael Callon in 1990 in his work about techno-economic networks, focusing latter on the importance of knowledge as a key resource of the companies and the changes these resources can have in moments of change, some become less important and others become key resources which the firms start to explore, acquire, develop.

Words also have been used before to analyze knowledge, the new data sources, data analysis techniques that are being made available by open source projects and advancement in the computers processing power. We propose a framework to analyze key knowledge resources implemented by firms in moments of change by analyzing their public information publications.

As a test case we are set to detect the knowledge related to big data implemented in new services in two content media related companies, the BBC international recognized broadcaster and Netflix, emerging media content actor which has been recognized for its strategy based in data analysis, by searching for words that represent big data related knowledge in their public R&D data in this case the R&D blogs for the BBC and Netflix.

6. METHOD

To detect knowledge words related to big data technologies from text published in the BBC&RD we will implement an automatic word extraction process and to verify the results we will do a second analysis by reading the information and summarizing it in a structured way but specifically extracting knowledge words related to big data technologies and finally compare both analysis and make conclusions

6.1. Automatic word extraction: for this analysis the three initial stages of the CRISP model are followed, business understanding, data understanding and Data preparation (Chapman P, et. Al., 2000), after that data analysis. In parallel to this process also the big data related knowledge data base have to be prepared for word analysis as it is used as reference to extract the knowledge words in the data analysis step, all these steps are implemented in R(R Core Team (2015).

Business understanding: is executed as the first part of the research question statement and the background literature review in order to understand the problem that wants to be solved

Data understanding: The desired text data is identified as the blog has different projects and different sections which do not have information related to the project, the not useful text is discarded and the desired project text is identified.

Data preparation: in this step the data is acquired, cleaned and formatted to be properly processed in the next step. The basic steps to prepare a data set for text mining are executed, removing punctuation, stop words, non-necessary white spaces and foreign characters from the words. After applying these steps the words can proceed to the next step were the knowledge words will be extracted.

Data analysis: in this step we do not follow the CRISP framework as we wish to extract knowledge words, in order to do that we defined a dictionary method which compares the text data against a set of target words and selects only the words from the text data that are present in the target set of words, in this case by exact match. As the objective is to detect big data related knowledge, a set of words collected by the authors in a previous work (Ruiz-navas S, Miyazaki K, 2014) is used as dictionary or target knowledge words. After comparing the text extracted from the three projects we obtain the words present in the projects that are present in the dictionary set of words and complete the big data knowledge word extraction.

6.2. Manual word extraction: For the summarization of each project a basic project structure is defined to understand its nature and identify the knowledge words. The basic structure is defined by

- Project name and date: basic information about the creation and finalization of the project and its name
- Objectives: which objectives were or are pursued by the project
- Problem: brief description of the problem the project solves or solved
- How ?: a description of the solution implemented
- Knowledge words: a set of words present in the project's text related to big data knowledge
- Technology words: a set of words present in the project's text related to big data technologies
- Data words: a set of words present in the project's text related to sources or kinds of data

After filling all these fields for this project, the comparison of the automatic words extraction and manual extraction is done, in this comparison a relative measure of effectiveness is implemented, words detected automatically/Words detected manually and conclusions about the whole process are obtained.

7. RESULTS

From the information published in the BBC&RD web site about three projects related to big data enabled service, COMMA, World Radio Service (WRS) and Kiwi project a manually (reading) and an automatic (automatic extraction of words) analysis we present both results for each project one after another and then general characteristics of the analysis of the three projects.

7.1. Knowledge words extraction

KIWI project manual extraction: From the characteristics we decided to analyze we could tell the project was for internal (Service to Within - S2W) use in the BBC and its focus was directed to improve BBCs internal process. As for the knowledge they implemented it is interesting the presence of words like algorithm, tagging, distributing computation, which are closely related to distributed systems and data analysis, which are general knowledge fields related to Data and when applied to relatively Big data sets, 70000 radio programs can be called a big volume of audio data. The general results are presented in table 1

KIWI - 2011 – 2012

Objectives	Problem	How?
Assigning topics to large programme archives in a reasonable time. 70,000 programmes in around two weeks	BBC manually tags recent programmes. Editors draw and assign these tags from open datasets made available within the Linked Data cloud, but this is a time consuming process. Aside from recent programming, which is tagged, the BBC has a very large radio archive that is currently untagged	Automatically identifying topics in speech radio programmes, with topic identifiers being drawn from Linked Open Data sources such as DBpedia
Knowledge	**Technology**	**Data**
Automated Tagging algorithm Distributing computations HUB4 acoustic model language model Tagging algorithm disambiguation SKOS hierarchy	Automatically tag Automatic tagging cloud resources processing infrastructure CMU Sphinx-3 software message-queueing system RabbitMQ cloud infrastructure Github	Linked Open Data speech radio programmes archive content Gigaword corpus Dbpedia Wikipedia

Table 1. Project characteristics summarized after reading the published information and list of words detected from this analysis.

Kiwi project automatic word extraction: From the automatic extraction as we could see that the only word related to knowledge that was detected was the KOS hierarchies and if we compare with the number of words detected by manual extraction we could say that we got 1/7(14%) of efficacy, which should be improved. The results are presented in table2.

WORDS	FREQUENCY
infrastructure	5
terms	4
topics	4
similarity	3
framework	2
Radio	2
Skos	2
Tools	2
annotations	1
classification	1

Table 2. 10 most frequent words automatically extracted from the Kiwi project's main page of the BBC&RD blog.

COMMA project manual extraction: The COMMA project was more directed as a final product or packaging of the result of the previous related projects and its objective is to provide service to third parties (Service to business- S2B), which is reflected in the relative fewer number of knowledge words detected if we compare with the other projects. The results are presented in table 3.

COMMA - 2013 - 2015		
Objectives	**Problem**	**How?**
COMMA stands for 'Cloud Marketplace for Media Analysis Develop a prototype platform for the extraction of metadata from media archives	There are many cultural institutions, commercial archives and content creators who have audio, film, photos and video that they would like to put to new uses. Digitalization is from big physical to big digital, so without easy ways to navigate the data there's no way for your users to get to the bits they want	content owners can upload their audio, video, photos or text and try out a range of different metadata extraction algorithms and researchers or businesses working in the field of automated media analysis can find customers
Knowledge	**Technology**	**Data**
automated media analysis algorithms	metadata extraction algorithms metadata extraction platform speech-to-text face recognition computing platform speaker recognition cloud-based processing system monitoring system payment system	Video Audio photos text

Table 3. Project characteristics summarized after reading the published information and list of words detected from this analysis.

COMMA project automatic word extraction: for this project we could say that the efficiency was ½ 50% but it would not be fair to say it because of the little presence of knowledge related words, the only hit was the word “algorithms” which helped us to realize the importance of the stemming process for word extraction. The results are presented in table 4.

WORDS	FREQUENCY
extraction	5
metadata	5
algorithms	4
media	4
prototype	4
innovation	2
research	2
value	2
design	1
development	1

Table 4. 10 most frequent words automatically extracted from the COMMA project’s main page of the BBC R&D blog.

World Radio Service (WRS) project manual extraction: This project is related to the creation of a new service for the beneficiaries of the BBC (Service to Customers - S2C). It is a project in development and this status is reflected in the medium level of knowledge words we detected, if we compare with the other projects and a very interesting finding is the combination of computer power and crowd sourcing to improve the service, which as explained by the case, helps to improve the technical aspect of tagging algorithms and generates user's data of the interaction with the service which is later used to improve it. A summary is presented in table 5.

WRS - 2011 - present		
Objectives	**Problem**	**How?**
Make available to the people the archive of all radio programms the BBC has. 70,000 radio programmes, or more than three years' worth of continuous audio.	Manually tagging programmes with metadata about them is expensive and time-consuming	We have done this by developing algorithms that listen to the radio programmes and create new descriptive metadata automatically and crowd sourcing to improve the results
Knowledge	**Technology**	**Data**
cloud processing machine-learning techniques advanced algorithms crowd-sourcing	speech-to-text technology semantic tagger Github Amazon Web Services	machine-generated metadata metadata tags massive audio archive

Table 5. Project characteristics summarized after reading the published information and list of words detected from this analysis.

WRS project automatic word extraction: if we are to evaluate the efficacy of the knowledge extraction in this case is 0 as the exact knowledge words were not detected, we can see again the limitation of the single word approach implemented as the word "advanced algorithms" is present but the automatic extraction method only detected "algorithms" as a single word. The extracted words are presented in table 6.

WORDS	**FRECUENCY**
metadata	12
radio	5
prototype	4
algorithms	2
power	1
technology	1
topics	1

Table 6. Words automatically extracted from the WRS project's main page of the BBC&RD blog.

7.2. General analysis

After comparing the automatic and manual extraction results, for the three projects we realized that to be able to detect knowledge as proposed, the automatic extraction method needs to be improved in different at least 4 ways:

a) **Multiword detection**, the majority of the words in the data base describing knowledge are Multiword, which do not correspond to the automatic extraction methods applied in this paper, and explain in part the little effectiveness of the knowledge detected and extracted from the analyzed projects.

b) **Stemming of the words**, the most characteristic case is the word algorithm which is present in all the projects selected and also in the big data knowledge data base but it was not detected because in the data base was saved in plural form “algorithms” rather than algorithm.

c) **Context** to have a reference for the identification of knowledge words as in this paper the researchers implemented their own knowledge of the big data technologies to do the manual extraction. This could be achieved by implementing the word

d) **Complement distance measures:** Improving the word detection, by implementing diverse word distance measures like the Levenshtein distance (Levenshtein V I, 1966) which helps to avoid the problem with not stemmed and multi words distance detection.

8. DISCUSSION

As a points for discussion we start with the research’s limitations, the samples studied are small only three projects and from one company, the data source used is the R&D blog were the information provided is to inform to a general public and the information is not necessary in deep about the projects been realized, which means that much of the information of the project is not published in this site. The automatic extraction set of tools used was very limited and can be improved in several ways in order to obtain better results, also is evident the need more standard measures for the effectiveness of the tools implemented by the authors, also the research is still in process and advancing the Netflix case is still to be completed. Having mention these issues and missing others we share other topics that called our attention while doing this paper.

The implications of using different sets of words in the automatic word extraction, the big data knowledge data base which was constructed from the scientific publications’ author keywords field, these are defined by humans, and in an academic context, which could be different from the words used in companies Blogs and extracted by automatic algorithms. These differences could in the end could be reflected in the final automatic extraction efficacy.

The new availability of tools and data sources for analyzing technology and innovation, for the authors this paper is considered as a small drop of water in the ocean that is being discovered, around new sources and tools for innovation research, as we did our analysis using very simple tools and three projects from 1 R&D blog. Today there is availability of many data sources, like government open data in many countries, social networks data, company internal data, that could expand the reach of the studies in our fields of interest and help us to make better informed decisions if used in a responsible fashion.

9. CONCLUSION

An automatic and manual analysis of information of 3 R&D projects was realized for the BBC, from this we conclude that the in order to implement automatic knowledge detection using text data from web sites public data, it is necessary the implementation of diverse algorithms and their adaptation to the specific needs of the discipline in question and the question this wants to answer, for this paper key knowledge resources that the companies were implementing in their big data enabled service R&D projects.

The original automatic knowledge extraction objective proposed by the paper was not fully achieved, as in the end the automatic extraction needed a better implementation, yet this processes of research is taken as a first step and a road map for the improvement of the method.

From the manual analysis it was possible to observe some of the knowledge being implemented by the BBC in these projects, confirming in some point that big data related knowledge as a resource is taking importance for broadcasters and is being implemented in service. Also was detected from the 3 projects that the big data knowledge has been used three ways, for service to the company (S2W) to improve process efficiency, new service to their customers or service beneficiaries (S2C) and new services to other organizations (S2B).

We found that the knowledge words are related to diverse knowledge fields like distributed systems, artificial intelligence, natural language processing and signal processing, this variety show us that for the application of the big data technologies for innovation, the knowledge range is diverse and not previously being used together, implying that to exploit these technologies some new skills are required which call for action on capability building and a short term exploration for companies that are interested to exploit the potential value of big data technologies

All these preliminary findings will help to improve the final result of this research.

REFERENCES

Ávila-Robinson, A., & Miyazaki, K. (2013). Dynamics of scientific knowledge bases as proxies for discerning technological emergence — The case of MEMS/NEMS technologies. Technological Forecasting and Social Change, 80(6), 1071–1084. doi:10.1016/j.techfore.2012.07.012

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. Journal of Management, 17(1), 99–120. doi:10.1177/014920639101700108

Boucher, J. (2015). Up Periscope: Broadcasting Live Video On Your Smartphone. huffpost tech. Retrieved January 5, 2015, from http://www.huffingtonpost.com/jason-boucher/up-periscope-broadcasting-live-video-on-your-smartphone_b_6988258.html

Callon, M. (1991). Techno-economic Networks and Irreversibility. A Sociology of Monsters: Essays on Power, Technology …, 132–165. doi:10.1007/s10661-009-1100-9

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). Step-by-step data mining guide. SPSS Inc, 78, 1–78. Retrieved from http://www.crisp-dm.org/CRISPWP-0800.pdf

Cohen, W. M., & Levinthal, D. a. (1994). Absorptive Capacity : A New Perspective on and Innovation Learning, 35(1), 128–152.

Cooke, P., Uranga, M. G., & Etxebarria, G. (1997). Regional innovation systems: Institutional and organizational dimensions. Research Policy, 26, 475–491. Retrieved from http://www.sciencedirect.com/science/article/pii/S0048733397000255

De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. In AIP Conference proceedings 1644 (Vol. 97, pp. 97–104). AIP Publishing. doi:10.1063/1.4907823

Grant, R. (1996). Toward a Knowledge Based Theory of the Firm. Strategic Management Journal, 17(1), 109–122.

Hong, S., & Miyazaki, K. (2013). Technological and non-technological innovations in B2B mobile services in Korea. Asian Journal of Technology Innovation, 21(1), 1–20. doi:10.1080/19761597.2013.815480

Islam, N., & Miyazaki, K. (2010). An empirical analysis of nanotechnology research domains. Technovation, 30(4), 229–237. doi:10.1016/j.technovation.2009.10.002

Ittipanuvat, V., Fujita, K., Sakata, I., & Kajikawa, Y. (2014). Finding linkage between technology and social issue: A Literature Based Discovery approach. Journal of Engineering and Technology Management - JET-M, 32, 160–184. doi:10.1016/j.jengtecman.2013.05.006

Jee, K., & Kim, G. H. (2013). Potentiality of big data in the medical sector: Focus on how to reshape the healthcare system. Healthcare Informatics Research, 19(2), 79–85. do:10.4258/hir.2013.19.2.79

Jurowetzki, R., & Hain, D. S. (2014). Mapping the (R-) Evolution of Technological Fields – A Semantic Network Approach. In L.-M. Aiello & D. McFarland (Eds.), Social Informatics (pp. 359–383). Springer International Publishing. doi:10.1007/978-3-319-13734-6_27

Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: Issues, challenges, tools and Good practices. 2013 Sixth International Conference on Contemporary Computing (IC3), 404–409. doi:10.1109/IC3.2013.6612229

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady, 10(8), 707–710.

Malerba, F. (2002). Sectoral systems of innovation and production. Research Policy, 31, 247–264. Retrieved from http://www.sciencedirect.com/science/article/pii/S0048733301001391

Manyika, J., Chui, M., Brown, B., Bughin, J., Richard, D., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Quarterly, (June). Retrieved from http://www.citeulike.org/group/18242/article/9341321

McAfee, Andrew; Brynjolfsson, E. (2012). Big Data: The Management Revolution. Harvard Business Review, 90(10), 60–68.

Nakagawa, T. (2014). Overview of Big Data Uses and Applications. Broadcast Technology, (55), 16–17.

Nonaka, I., & Takeuchi, H. (1995). The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation (1st ed.). New York, New York, USA: Oxford University Press.

OECD (2013), "Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by "Big Data"", OECD Digital Economy Papers, No. 222, OECD Publishing. http://dx.doi.org/10.1787/5k47zw3fcp43-en

Palmer, D. (2014). Big data “crucial” to daily decision making at Netflix. Computing. Retrieved January 1, 2015, from http://www.computing.co.uk/ctg/analysis/2354840/big-data-crucial-to-daily-decision-making-at-netflix#

Prahalad, C., & Hamel, G. (1990). The core competence of the corporation. Boston (MA). Retrieved from http://books.google.com/books?hl=en&lr=&id=W4XODWvUs4oC&oi=fnd&pg=PA41&dq=The+core+competenece+of+the+corporation&ots=yOkLRHor8S&sig=d1tbsc3f6wOIEyloYRyCwWxpfsI

Patel, P., & Pavitt, K. (1994). National Innovation Systems: Why They Are Important, And How They Might Be Measured And Compared. Economics of Innovation and New Technology. doi:10.1080/10438599400000004

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. Health Information Science and Systems, 2, 3. doi:10.1186/2047-2501-2-3

Ruiz-navas, S., & Miyazaki, K. (2014). Unveiling the Knowledge Convergence on “ Big data ” : Analysis of Scientific Keywords . Asialics 2014, 1–23.

Staff S. (2011). Challenges and opportunities. Science, 331, 692–693.doi:10.1126/science.331.6018.692

Teece, D., & Pisano, G. (1994). The dynamic capabilities of firms: an introduction. Industrial and Corporate Change, 3. Retrieved from http://icc.oxfordjournals.org/content/3/3/537.2.short